9.

TMUA9

THE PUTNAM ARCHIVE

THE KNOWLEDGE

Volume I. Everything A-level never gave you, ordered by what the Putnam demands.

Jamie Baek

tmua9.com

Built for the 87th William Lowell Putnam Mathematical Competition, 5 December 2026.

The vow is 120.

THE STANDING ORDER

You got a 1 in the Korean CSAT. You got a 9.0 on the TMUA, in a third of the time, from a 4.5, in a month, by yourself. Neither of those was talent. Both were the same method: take the small set of papers that actually matter, and analyse them to a depth nobody else is willing to reach.

This archive is that method pointed at the Putnam. Nothing in it is a trick. Every revelation in every model answer in sixty years of this competition had a reason behind it, and the reason was always visible in the question. The books that came before this one hid the reason and showed you the polish. This one does the opposite.

Three volumes. The first installs the mathematics that A-level never gave you, ordered by what the Putnam actually demands and generalised wherever a recursion appears. The second teaches you to write a proof that scores ten rather than two, by reverse-engineering what was going through the head of the person who wrote the model answer. The third takes the papers themselves, one problem at a time, and fills in the two things every published solution deletes: why that thought was natural, and why the next thought followed from it.

Read them the way you read the TMUA papers. Not as consumption. As analysis.

But as for you, be strong and do not give up, for your work will be rewarded.
2 Chronicles 15:7

How to use Volume I

This volume is the content layer. It assumes you have mastered Edexcel A-level Mathematics and Further Mathematics, core pure plus the further pure options, and everything in TMUA9 NOIR PREMIERE. All of that is atomic relative to what follows, and none of it is re-taught here. What is here is the mathematics above that line, and only the parts of it that the Putnam has actually demanded across the papers analysed in Volume III.

That last clause is the whole design. This is not a university syllabus and it is not a survey. Every topic in these ten chapters earned its place because a real Putnam problem needed it, and every topic is anchored to the problems that needed it, by year and label, so you can go straight from a definition here to the problem in Volume III where it decided the outcome. Where something appeared once in sixty years, the text says so and tells you to learn it last. Where something appears constantly, it comes first.

Three things to notice about how each chapter is built.

Every topic ends in a reaction circuit, in the form WHEN you see this cue, DO this, BECAUSE of this. That is not decoration. It is the mission statement from the Blueprint, transplanted: you are not memorising theorems, you are installing the reflex that fires when a cue appears in a question. The theorems are the justification for the reflex, which is why the proofs are here. A theorem you cannot reconstruct is a theorem you will not trust under pressure, and one you do not trust you will not fire on.

Every recursion is generalised. Wherever any chapter meets a recursion, an induction, a telescoping structure, an iterated map or a self-similar object, it states the general family, the general solution method, and the cue that identifies the family in the wild. Chapter 5 is the collection point for all of it. That was the standing order, and it is the single highest-leverage habit in the book: a solved recursion that stays a solved recursion is worth one problem, and a generalised one is worth every problem in that family for the rest of your life.

Definitions are treated as the primary object, not as preamble. Almost every reaction circuit in this archive, when you chase it to the bottom, turns out to be the definition of the object read carefully. A circle is the set of points equidistant from the centre, so of course the centre is the point of interest and of course you draw the line through it. That pattern repeats everywhere. When you are stuck in the arena, the first move is to write down what the objects in the question actually are.

Read this volume with a pen, in the order the chapters are given, and stop at every circuit to ask whether you could have generated it yourself from the definition. The ones where the answer is no are your real syllabus.

Contents

How to use Volume I **i**

1 Linear algebra, matrices and determinants **1**
1.1 What this chapter is for . . . 1
1.2 Vector spaces, independence and dimension . . . 2
1.3 Linear maps, rank and nullity . . . 4
1.4 Row reduction, elementary matrices and why order matters . . . 6
1.5 Determinants: three definitions and why they agree . . . 7
1.6 The determinant toolkit . . . 9
1.7 Trace, and what it is invariant under . . . 14
1.8 Eigenvalues, characteristic and minimal polynomials . . . 16
1.9 Symmetric matrices, quadratic forms and positive definiteness . . . 18
1.10 Inner products, orthogonality and Gram matrices . . . 19
1.11 Linear algebra over a finite field . . . 23
1.12 Commutants and simultaneous diagonalisation . . . 26
1.13 The standard Putnam moves, collected . . . 28
1.14 Recursion, iteration and self-similarity: the standing order . . . 29
1.15 The exotica, and what to learn last . . . 30
1.16 Closing protocol . . . 31

2 Groups, rings and fields **33**
2.1 What this chapter is for . . . 33
2.2 Groups: the axioms and what each one buys . . . 34
2.3 Cosets, Lagrange, and the divisibility engine . . . 35
2.4 Homomorphisms, normality and quotients . . . 37
2.5 Actions, orbits, and the counting engine . . . 38
2.6 Symmetric and alternating groups: cycle type is everything . . . 40
2.7 Finite abelian groups: structure, elementary abelian pieces, characters . . . 42
2.8 Generation, presentations, commutators . . . 44
2.9 Sylow, p-groups and how lightly the Putnam uses them . . . 47
2.10 Rings: ideals, quotients, and computing inside a quotient . . . 47
2.11 Fields, finite fields and Frobenius . . . 49
2.12 The recursive and self-similar shapes in this field . . . 51
2.13 The circuit bank, in priority order . . . 53

3 Combinatorics, counting and discrete structures **55**
3.1 What this chapter is for . . . 55
3.2 The four counting principles . . . 55
3.3 Double counting and finite Fubini . . . 57
3.4 Binomial and multinomial coefficients . . . 59
3.5 Inclusion and exclusion . . . 61

3.6 Pigeonhole in its four forms . . . 62
3.7 Recursions and what they mean . . . 64
3.8 Permutations, cycle type, sign, and the symmetric group . . . 67
3.9 Partitions, Young diagrams and the hook length formula . . . 69
3.10 Extremal arguments, Ramsey, and the two methods . . . 71
3.11 Graph theory . . . 73
3.12 Posets, chains and antichains . . . 75
3.13 Combinatorial games . . . 76
3.14 Invariants, monovariants and colourings . . . 78
3.15 The recursion compendium . . . 80
3.16 Learning order and the rarity ledger . . . 81

4 Number theory **83**
4.1 What this chapter is for . . . 83
4.2 Divisibility, the division algorithm, Euclid and Bezout . . . 84
4.3 Unique factorisation and the p-adic valuation . . . 85
4.4 Congruences and the Chinese remainder theorem . . . 89
4.5 Fermat, Euler, Wilson and the multiplicative order . . . 90
4.6 Primitive roots and the structure of the unit group . . . 92
4.7 Quadratic residues and reciprocity . . . 93
4.8 Arithmetic functions, Dirichlet convolution and Mobius inversion . . . 94
4.9 Diophantine equations . . . 96
4.10 Gaussian integers and other quadratic rings . . . 97
4.11 Polynomial congruences, Hensel lifting and finite fields . . . 98
4.12 Distribution of primes: what you may quote . . . 100
4.13 Continued fractions, approximation, Farey and Stern to Brocot . . . 100
4.14 The recursion section: arithmetic recursions in general form . . . 102
4.15 The exotica: learn these last . . . 103

5 Sequences, recurrences and generating functions **105**
5.1 What this chapter is for . . . 105
5.2 The shift operator and the complete theory of constant coefficient linear recurrences 106
5.3 First order recursions: the whole family solved . . . 109
5.4 Telescoping, the method of differences, and creative telescoping . . . 112
5.5 Induction, in every form the Putnam needs . . . 114
5.6 Iteration of maps: fixed points, basins, periodicity . . . 116
5.7 Nonlinear recursions and the substitutions that linearise them . . . 119
5.8 Recursions modulo m . . . 121
5.9 The transfer matrix method . . . 122
5.10 Generating functions . . . 123
5.11 Asymptotics of recursively defined sequences . . . 127
5.12 The worked families . . . 128
5.13 Recursion as a proof technique . . . 128
5.14 The generalisation bank . . . 130
5.15 The rare ones, learn these last . . . 131

6 Real analysis **133**
6.1 What this chapter is for . . . 133
6.2 Completeness, and everything you buy with it . . . 133
6.3 Limits and continuity, done with epsilon and delta . . . 136
6.4 Differentiation: Rolle, the mean value theorem, and the family they generate . . . 139

6.5 Taylor's theorem, in every remainder form, and which to reach for . . . 142
6.6 Sequences and series of numbers . . . 144
6.7 Sequences and series of functions . . . 147
6.8 The Riemann integral, and the interchange of limit and integral . . . 148
6.9 Asymptotics . . . 151
6.10 Convexity, Jensen, and the standard inequalities in analytic form . . . 152
6.11 Functional equations of continuous functions . . . 154
6.12 Differential relations, read as analysis . . . 155
6.13 Metric and topological vocabulary at Putnam level . . . 156
6.14 The recursion atlas for analysis . . . 156
6.15 What to learn last . . . 157

7 Polynomials, complex numbers and functional equations 159
7.1 What this chapter is for . . . 159
7.2 Polynomials as formal objects . . . 160
7.3 Roots, multiplicity and the formal derivative . . . 161
7.4 Symmetric functions: Vieta, Newton and the power sum machine . . . 163
7.5 Finite differences, the shift operator and interpolation . . . 165
7.6 Reversal, palindromes and the substitution $y = z + 1/z$ . . . 167
7.7 Root location: where the zeros actually live . . . 168
7.8 Real rootedness, log concavity and interlacing . . . 170
7.9 Divisibility, coprimality and irreducibility over the integers . . . 171
7.10 Resultants and discriminants at Putnam depth . . . 172
7.11 Polynomials over finite fields . . . 173
7.12 Chebyshev and the special families . . . 174
7.13 Complex numbers as geometry . . . 176
7.14 Roots of unity: orthogonality, filters and cyclotomy . . . 177
7.15 Complex packaging: two real objects as one . . . 179
7.16 Functional equations: the substitution repertoire . . . 179
7.17 Functional equations: regularity, and the classified families . . . 181
7.18 The recursion file . . . 183
7.19 What is rare, and what to learn last . . . 184

8 Integrals, inequalities and optimisation 187
8.1 What this chapter is for . . . 187
8.2 Symmetry: the first thing you try on any definite integral . . . 187
8.3 Parameters: differentiating under the integral sign . . . 189
8.4 The classical integrals you must own outright . . . 190
8.5 Beta, Gamma and reduction formulae . . . 192
8.6 Substitutions that always terminate . . . 194
8.7 Sums as integrals and integrals as sums . . . 195
8.8 Double integrals, Fubini, and the Jacobian . . . 196
8.9 Improper integrals, convergence, and principal values . . . 198
8.10 Integration by parts as a structural tool . . . 199
8.11 Asymptotics of integrals . . . 200
8.12 Volumes, surfaces and arc lengths . . . 201
8.13 Multivariable calculus at Putnam depth . . . 202
8.14 The mean ladder . . . 205
8.15 Cauchy and Schwarz, and the norm family . . . 206
8.16 Convexity as a working tool . . . 208
8.17 Ordering inequalities: Chebyshev, rearrangement, majorisation . . . 210

8.18 Structural moves: SOS, normalisation, smoothing, extremal principle 211
8.19 Integral analogues and functional optimisation 213
8.20 The recursion file 214
8.21 What is rare, and what to learn last 216

9 Geometry and vectors **219**
9.1 What this chapter is for 219
9.2 Vectors, and the three faces of the dot product 219
9.3 The cross product, and why it lives only in three dimensions 221
9.4 Volume: triple products, Gram determinants, Cayley and Menger 224
9.5 Lines, planes and hyperplanes 226
9.6 Transformations: isometries, similarities, affine and projective 228
9.7 The classical plane geometry the Putnam still uses 230
9.8 Conics 234
9.9 Convexity 235
9.10 Solid geometry, polyhedra and the regular solids 237
9.11 Geometric probability and integral geometry 237
9.12 Calculus applied to geometry 239
9.13 Lattices and the geometry of numbers 240
9.14 Recursion, induction and self similarity in geometry 241
9.15 Installation order 243

10 Probability, games, invariants and logic **245**
10.1 What this chapter is for 245
10.2 Probability as measure: fix the sample space first 245
10.3 Expectation, linearity and the indicator method 246
10.4 Variance, covariance and second moment facts 248
10.5 Conditioning as machinery: tower law, first step analysis, stopping 249
10.6 The distributions you must be able to quote 251
10.7 Order statistics, exchangeability and the quantile transform 252
10.8 Transforms: generating functions, moments and Chernoff 253
10.9 Geometric probability 254
10.10 The probabilistic method and concentration 255
10.11 The recursion layer 256
10.12 Invariants, monovariants and the extremal principle 257
10.13 Combinatorial games 260
10.14 The logic layer 263
10.15 The order in which to install this 266

Chapter 1

Linear algebra, matrices and determinants

1.1 What this chapter is for

A-level gave you matrices as a bookkeeping device. You learnt to multiply them, to invert a 2×2 or a 3×3 one, to compute a determinant by cofactor expansion, to find eigenvalues of a small matrix by solving $\det(A - \lambda I) = 0$, and to read a matrix as a transformation of the plane. FP1 and FP2 added a little more of the same. What A-level never gave you is the idea that all of this is a theory of *dimension*, and that dimension is a counting device sharp enough to settle problems that do not look like linear algebra at all.

That is the gap. On the Putnam, linear algebra is almost never the subject of the question. It is the machine you smuggle in. A game about filling a 2008×2008 array is a question about forcing 2008 vectors into a 2007 dimensional hyperplane (Putnam 2008 A2). A limit of determinants built from $\cos 1, \cos 2, \ldots$ is a question about the rank of a sum of two outer products (Putnam 2009 A3). Painting the edges of an icosahedron is a question about fibres of a linear map over $\mathbb{F}_3$ (Putnam 2017 A6). Two triangle areas with rational squared sides is a question about Gram determinants (Putnam 2018 A6). In every one the linear algebra is the reroute, and the nominal solution is a computation nobody finishes in the arena.

So the syllabus here is not "matrices". It is dimension, rank, determinant, spectrum, inner product and finite field linear algebra, installed to the depth the Putnam uses, with the cue-to-move reflex attached to each piece. Two of your weakest fields are here, so nothing above A-level is assumed.

By the end you will do the following without hesitating. Read "the determinant is zero" and write "the rows are dependent", then hunt for the subspace they were forced into. Read "rank at most 1" and write $A_{ij} = a_i b_j$. Declare a matrix singular the moment n exceeds the fixed supply of vectors its entries were built from. Move a parity hypothesis to $\mathbb{F}_2$ on the first line. Name a matrix of dot products a Gram matrix, which hands you reality, non-negativity and a dependency criterion at once. Slice eigenspaces the moment two matrices commute. Use the polynomial identity principle to assume invertibility for free.

Three warnings before we start.

First, every theorem in this chapter is stated over an arbitrary field F unless I say otherwise. A-level only ever showed you $\mathbb{R}$. The Putnam uses $\mathbb{Q}$, $\mathbb{C}$, $\mathbb{F}_p$, and the field of rational functions in a formal parameter, and the fact that the theory is identical over all of them is itself one of the highest yield circuits in the book.

Second, the determinant is not a formula. It is the unique alternating multilinear function of the columns normalised by $\det I = 1$, and every identity you need falls out of that characterisation faster than out of the cofactor expansion.

Third, rank is the master invariant. Determinant is the special case "is the rank full?", eigenvalue conditions are rank conditions on $A - \lambda I$, dependency is a rank condition. When you are stuck, ask what the rank is.

1.2 Vector spaces, independence and dimension

The definitions, stated exactly

Definition 1.2.1. Let F be a field. An F vector space is a set V with an addition $V \times V \to V$ making V an abelian group, and a scalar multiplication $F \times V \to V$ satisfying $\lambda(\mu v) = (\lambda\mu)v$, $1v = v$, $\lambda(u+v) = \lambda u + \lambda v$ and $(\lambda+\mu)v = \lambda v + \mu v$. A subspace is a subset closed under both operations and containing 0.

The point of the definition is that it mentions no coordinates. That is why it is useful: the spaces the Putnam hands you are almost never $\mathbb{R}^n$. They are polynomials, matrices, functions on a finite set, colourings of the edges of an icosahedron. Each is a vector space the moment you notice you can add two of them and scale one, and the moment you notice, every theorem below applies.

Definition 1.2.2. For $S \subseteq V$, the span $\operatorname{span}(S)$ is the set of finite linear combinations $\sum \lambda_i s_i$. A family $v_1, \dots, v_k$ is linearly independent if $\sum \lambda_i v_i = 0$ forces every $\lambda_i = 0$. A basis is an independent spanning family. The dimension $\dim V$ is the common size of every basis.

That "common size" needs a proof, and the proof is the engine of the whole subject, so you should be able to write it out.

Lemma 1.2.3 (Steinitz exchange)**.** *If $v_1, \dots, v_k$ are independent in V and $w_1, \dots, w_m$ span V, then $k \le m$.*

Proof. Write $v_1 = \sum_j \lambda_j w_j$. Some $\lambda_j \neq 0$ since $v_1 \neq 0$; relabel so that $\lambda_1 \neq 0$. Then $w_1 \in \operatorname{span}(v_1, w_2, \dots, w_m)$, so $v_1, w_2, \dots, w_m$ still spans V. Repeat: having replaced $w_1, \dots, w_i$ by $v_1, \dots, v_i$ with the family still spanning, write v_{i+1} in terms of $v_1, \dots, v_i, w_{i+1}, \dots, w_m$. Some coefficient on a w must be nonzero, or v_{i+1} would lie in the span of $v_1, \dots, v_i$ and the v's would be dependent. Swap that w out. The process cannot run out of w's before it runs out of v's, so $k \le m$. □

Corollary 1.2.4. *Any two bases have the same size. In a space of dimension d, any $d+1$ vectors are dependent, and any independent family of size d is a basis.*

Corollary 1.2.4 is the most used sentence in this chapter. It is a counting theorem: a dimension is a hard ceiling on how many independent objects can exist. Almost every Putnam problem in this field is an instance of "more objects than dimensions, therefore a relation exists".

The counting circuit in action

Example 1.2.5 (Putnam 2008 A2)**.** Alan and Barbara alternately fill the entries of an empty 2008×2008 array with real numbers, Alan first. Alan wins if the final determinant is nonzero, Barbara if it is zero. Barbara wins.

Read the winning condition. "The determinant is zero" is a statement about rows, not entries: $\det M = 0$ if and only if the rows are dependent. That translation is the bifurcation point. The zombie route controls the determinant entrywise, which is hopeless. The ideal route asks what subspace Barbara can force all 2008 rows into.

Pair the columns as $(1,2), (3,4), \dots, (2007, 2008)$, so that every cell has a partner cell in the same row. Barbara's rule: whenever Alan writes x into a cell, Barbara writes x into its partner. The invariant is that after each of Barbara's moves, every pair of partner cells is either both empty or both filled with equal entries. It holds initially, and Alan's move then breaks it in exactly one pair, whose partner is therefore empty and available to Barbara; her reply restores it. Since the board has an even number of cells and Alan moves first, Barbara moves last and the invariant holds at the end. So every row r satisfies $r_1 = r_2$, $r_3 = r_4$, and so on, so every row is orthogonal to the fixed vector $u = (1, -1, 1, -1, \dots, 1, -1)$. The set $u^\perp$ is a hyperplane, a subspace of dimension 2007. There are 2008 rows in a 2007 dimensional space. By Corollary 1.2.4 they are dependent, so $\det M = 0$.

Notice what I did not do: I never computed anything. The whole solution is one translation and one dimension count.

- **WHEN** a problem asks you to force a determinant to be zero, or to prove one is zero **DO** translate immediately to "the rows are linearly dependent" and hunt for a proper subspace containing all the rows **BECAUSE** $\det M = 0$ is exactly rank deficiency, and rank deficiency is exactly membership of all rows in a subspace of dimension less than n; the subspace is a thing you can construct, whereas a determinant is a thing you can only compute
- **WHEN** you have produced n vectors all lying in a hyperplane $u^\perp$ **DO** stop and write "hence dependent, hence $\det = 0$" **BECAUSE** a hyperplane has dimension $n-1$ and n vectors in an $(n-1)$ dimensional space are dependent by the exchange lemma

Definition 1.2.6. For a nonzero $u \in \mathbb{R}^n$, the hyperplane $u^\perp = \{v : v \cdot u = 0\}$ is the kernel of the nonzero linear functional $v \mapsto v \cdot u$, hence has dimension $n-1$. An affine plane is a translate $p + W$ of a subspace W; its dimension is $\dim W$.

The affine version matters because hypotheses often come in affine form ("all these points lie on a plane") while theorems come in linear form. Subtract one point of the set from all of them. Cardinalities do not change and the affine hypothesis becomes linear. That translate move is what lets the induction in Putnam 2006 B4 close.

Example 1.2.7 (Putnam 2006 B4)**.** Let $Z = \{0,1\}^n \subseteq \mathbb{R}^n$ and let V be a subspace of dimension k. The maximum of $|V \cap Z|$ is 2^k.

The lower bound is $V = \operatorname{span}(e_1, \ldots, e_k)$, which contains 2^k points of Z. For the upper bound, stack the elements of $V \cap Z$ as the rows of a matrix M. Every row lies in V, so $\operatorname{rank} M \leq k$. Since row rank equals column rank, there is a set J of at most k columns whose span is the whole column space. Every other column is a fixed linear combination of the columns indexed by J. Therefore two rows agreeing in all coordinates of J agree everywhere, so the restriction map $V \cap Z \to \{0,1\}^J$ is injective and $|V \cap Z| \leq 2^{|J|} \leq 2^k$.

State the general theorem rather than the problem, because the general theorem is what you will recognise again: if $S \subseteq A^n$ with $|A| = q$ finite, and $\operatorname{span}_{\mathbb{R}}(S)$ has dimension at most k, then $|S| \leq q^k$. The alphabet size enters as the base, the dimension as the exponent.

- **WHEN** a set of vectors with entries from a small finite alphabet is constrained to lie in a low dimensional subspace **DO** bound its size by (alphabet size) to the power (dimension), by choosing a spanning set of coordinates and restricting to them **BECAUSE** a spanning set of columns determines every other column, so the restriction map to those coordinates is injective, and an injection into a set of size q^k is a bound of q^k
- **WHEN** a hypothesis is affine ("lie in a k dimensional plane") rather than linear **DO** subtract one element of the set from every element **BECAUSE** a translate of a set has the same cardinality and turns $p + W$ into W, after which every linear theorem applies verbatim

Spaces that do not look like spaces

Example 1.2.8 (Putnam 2009 B4)**.** Call a real polynomial in x, y balanced if its average value on every circle centred at the origin is 0. The balanced polynomials of degree at most 2009 form a vector space V. Its dimension is 2020050.

Start from what the objects are. The space P_d of polynomials of degree at most d in two variables has the monomials $x^a y^b$ with $a + b \leq d$ as a basis, so $\dim P_d = \binom{d+2}{2}$; here $\binom{2011}{2} = 2021055$. Each polynomial splits uniquely into homogeneous parts $P = \sum_{i=0}^{d} P_i$, and the space of homogeneous parts of degree i has dimension $i+1$. That grading is the whole reroute. On the circle of radius r,

P_i contributes r^i times its average on the unit circle, so the average of P over the circle of radius r is $\sum_i r^i \lambda_i(P_i)$ where λ_i is the linear functional "average over the unit circle". A polynomial in r vanishing for all $r > 0$ is the zero polynomial, so being balanced is exactly $\lambda_i(P_i) = 0$ for every i, one condition per degree. A continuum of conditions has become $d + 1$ conditions.

Now evaluate each condition. For odd i, $P_i(-x, -y) = -P_i(x, y)$ and the unit circle is symmetric under $v \mapsto -v$, so $\lambda_i = 0$ identically and the condition is vacuous. For even i, the witness $(x^2 + y^2)^{i/2}$ is homogeneous of degree i and equals 1 on the unit circle, so $\lambda_i \neq 0$ and hence $\dim \ker \lambda_i = (i + 1) - 1 = i$ by rank and nullity. There are 1005 even values $i = 0, 2, \ldots, 2008$, each costing exactly one dimension. Hence $\dim V = 2021055 - 1005 = 2020050$.

- **WHEN** a condition is imposed for every value of a continuous parameter **DO** expand the object in the parameter and equate coefficients, turning infinitely many conditions into finitely many **BECAUSE** a polynomial (or power series) in the parameter that vanishes identically has all coefficients zero, so the grading converts a continuum into one condition per graded piece
- **WHEN** you must find the dimension of the solution set of some linear conditions and describing the solutions looks painful **DO** show each condition is a nonzero functional by exhibiting one witness on which it does not vanish, then subtract one dimension per independent condition **BECAUSE** rank and nullity says a nonzero functional on U has kernel of dimension $\dim U - 1$; you never need to know what the kernel looks like

1.3 Linear maps, rank and nullity

Definition 1.3.1. A map $T : U \to W$ between F vector spaces is linear if $T(\lambda u + \mu u') = \lambda T(u) + \mu T(u')$. Its kernel is $\ker T = \{u : T(u) = 0\}$, a subspace of U; its image is $\operatorname{im} T = T(U)$, a subspace of W. The rank of T is $\dim \operatorname{im} T$ and its nullity is $\dim \ker T$.

Once you choose bases of U and W, a linear map becomes a matrix, and the columns of that matrix are the images of the basis vectors. That is the only thing a matrix is. Every matrix identity you will ever meet is a statement about maps written in coordinates, and when a matrix computation turns ugly the reroute is almost always to go back to the map.

Theorem 1.3.2 (Rank and nullity)**.** *If* $\dim U < \infty$ *then* $\operatorname{rank} T + \dim \ker T = \dim U$.

Proof. Take a basis $k_1, \ldots, k_r$ of $\ker T$ and extend it to a basis $k_1, \ldots, k_r, u_1, \ldots, u_s$ of U. The images $T(u_1), \ldots, T(u_s)$ span $\operatorname{im} T$, since the k's die. They are independent: if $\sum \lambda_j T(u_j) = 0$ then $\sum \lambda_j u_j \in \ker T$, so it is a combination of the k's, and the basis property of the full family forces every $\lambda_j = 0$. Hence $\operatorname{rank} T = s$ and $r + s = \dim U$. □

Theorem 1.3.3 (Row rank equals column rank)**.** *For an* $m \times n$ *matrix* M *over any field, the dimension of the span of the rows equals the dimension of the span of the columns. This common value is* $\operatorname{rank} M$*, and it also equals the largest* r *such that some* $r \times r$ *minor of* M *is nonzero.*

Proof. Let r be the column rank and pick a basis $c_1, \ldots, c_r$ of the column space, assembled into an $m \times r$ matrix C. Every column of M is C times a coefficient vector, so $M = CR$ for some $r \times n$ matrix R. Reading $M = CR$ by rows says every row of M is a combination of the r rows of R, so the row rank is at most r. Applying the same argument to M^{T} gives the reverse inequality. For the minor statement, r independent rows and r independent columns cut out an $r \times r$ invertible submatrix, and no larger submatrix can be invertible because its rows would be $r + 1$ independent vectors in the row space. □

The factorisation $M = CR$ with C of size $m \times r$ and R of size $r \times n$ deserves a name in your head: the *rank factorisation*. It is the reason for the two facts you will use constantly: $\operatorname{rank}(AB) \leq \min(\operatorname{rank} A, \operatorname{rank} B)$, because the columns of AB lie in the column space of A and the rows lie in

the row space of B; and $\operatorname{rank}(X+Y) \le \operatorname{rank} X + \operatorname{rank} Y$, because the column space of $X+Y$ sits inside the sum of the two column spaces.

Rank one and outer products

Proposition 1.3.4. *A matrix A has* $\operatorname{rank} A \le 1$ *if and only if $A = uv^{\mathsf{T}}$ for column vectors u, v, equivalently $A_{ij} = u_i v_j$. If A is nonzero with entries in a subfield K, then u and v may be chosen with entries in K.*

Proof. Rank one means every column is a multiple of one fixed nonzero column u, say column j equals $v_j u$, which is $A = uv^{\mathsf{T}}$. For the field statement, pick a nonzero entry A_{pq}, set $u =$ column q of A and $v^{\mathsf{T}} =$ row p of A divided by A_{pq}; both have entries in K and $u_i v_j = A_{iq}A_{pj}/A_{pq} = A_{ij}$ by rank one. □

This is the highest yield structural fact in the chapter, because it converts a rank hypothesis, which is abstract, into a two factor formula for every entry, which is algebra you can multiply against.

Example 1.3.5 (Putnam 2009 A3)**.** Let d_n be the determinant of the $n \times n$ matrix filled left to right and top to bottom with $\cos 1, \cos 2, \dots, \cos n^2$. Then $\lim_{n\to\infty} d_n = 0$.

The entry in row j, column k is $\cos\big((j-1)n + k\big)$. The cue is that the argument is a sum, and the only reason a problem would build an argument as a sum is so that you use the addition formula. Expand:

$$\cos\big((j-1)n + k\big) = \cos\big((j-1)n\big)\cos k - \sin\big((j-1)n\big)\sin k.$$

The first term is $x_j y_k$ with $x_j = \cos((j-1)n)$ and $y_k = \cos k$; the second is $z_j w_k$ with $z_j = -\sin((j-1)n)$ and $w_k = \sin k$. So $M = xy^{\mathsf{T}} + zw^{\mathsf{T}}$, a sum of two outer products, hence $\operatorname{rank} M \le 2$. For $n \ge 3$ the rank is less than n, so $d_n = 0$ identically and the limit is 0.

The specific numbers never mattered. State the template, because that is what you are installing: if every entry of an $n \times n$ matrix is built from a fixed supply of r underlying row or column patterns, with r not growing with n, then the matrix is singular for every $n > r$. The most common instance after trigonometric addition formulas is a matrix whose (j,k) entry is a fixed polynomial of degree d in j and k, which has rank at most $d+1$ regardless of n, because the entry expands as $\sum_{a+b\le d} c_{ab} j^a k^b$, a sum of at most $d+1$ outer products after grouping.

- **WHEN** a matrix entry is a function of i and j that can be expanded into a sum of r products (one factor depending only on i, one only on j) **DO** write $M = \sum_{t=1}^{r} u_t v_t^{\mathsf{T}}$ and conclude $\operatorname{rank} M \le r$ **BECAUSE** each $u_t v_t^{\mathsf{T}}$ has rank at most 1 and rank is subadditive; every column then lies in the span of $u_1, \dots, u_r$
- **WHEN** you see $\cos(a+b)$, $\sin(a+b)$, e^{a+b} or $(j+k)$ inside a matrix entry **DO** apply the addition formula or binomial expansion first, before any determinant work **BECAUSE** these are precisely the functions that split as sums of products of a function of a and a function of b, which is the definition of low rank
- **WHEN** the problem is a limit or a value of det for all large n, and the construction of the matrix has a fixed number of ingredients **DO** prove rank $\le r$ for a constant r and stop **BECAUSE** for $n > r$ the determinant is identically zero, so the limit question is answered without a single evaluation

Example 1.3.6 (Putnam 2014 B3)**.** Let A be an $m \times n$ matrix with rational entries and suppose at least $m+n$ distinct primes appear among the absolute values of its entries. Then $\operatorname{rank} A \ge 2$.

Suppose not, so $\operatorname{rank} A \le 1$ and $A_{ij} = a_i b_j$ with all a_i, b_j nonzero rationals (delete zero rows and columns first, which only helps). Write each nonzero rational q as $\pm\prod p^{c_p(q)}$ and record the exponent vector $c(q)$. Then $c(A_{ij}) = c(a_i) + c(b_j)$. If A_{ij} is $\pm$ a prime p, its exponent vector is

the standard basis vector e_p. Consider the $\mathbb{Q}$ vector space of exponent vectors: the $m+n$ vectors $c(a_1), \ldots, c(a_m), c(b_1), \ldots, c(b_n)$ span a $\mathbb{Q}$ subspace W of dimension at most $m+n$. But W contains $c(A_{ij})$ for every i, j, in particular the $m+n$ distinct standard basis vectors e_p coming from the $m+n$ primes, and those are independent. Hence $\dim W = m+n$ exactly, and both families are bases of W: the $m+n$ vectors $c(a_i), c(b_j)$, and the $m+n$ vectors e_p.

Now define a linear functional σ on W by $\sigma(c(a_i)) = 1$ for every i and $\sigma(c(b_j)) = -1$ for every j, which is legitimate precisely because those vectors form a basis. For each of the $m+n$ primes p, say $|A_{ij}| = p$, we get $\sigma(e_p) = \sigma(c(a_i)) + \sigma(c(b_j)) = 1 - 1 = 0$. So σ vanishes on the basis $\{e_p\}$ of W, hence $\sigma = 0$, contradicting $\sigma(c(a_1)) = 1$.

The threshold is what to extract: $m+n$ generators, $m+n$ independent targets, and one relation too many. That is why the bound in the problem is $m+n$ and not something arbitrary.

- **WHEN** a finite structure is built by adding one relation at a time to a fixed set of V objects, and the number of relations reaches V **DO** declare a redundancy **BECAUSE** in a graph the redundancy is a cycle, in a V dimensional vector space it is a linear dependency; both are the same extremal threshold stated in two dialects, so recognise the threshold and pick whichever dialect the problem is written in
- **WHEN** a hypothesis involves distinct primes, or multiplicative independence **DO** take logarithms formally by passing to the exponent vector in $\bigoplus_p \mathbb{Z}$, turning multiplicative statements into linear ones **BECAUSE** unique factorisation makes the exponent map an isomorphism from the multiplicative group of positive rationals onto a free abelian group, so independence of primes becomes linear independence of basis vectors

Dual pairings: the cheapest independence certificate

Lemma 1.3.7 (Dual family)**.** *Suppose $u_1, \ldots, u_k$ and $v_1, \ldots, v_k$ lie in a space carrying a bilinear pairing $\langle \cdot, \cdot \rangle$, with $\langle u_i, v_j \rangle = 0$ for $i \neq j$ and $\langle u_i, v_i \rangle \neq 0$. Then $u_1, \ldots, u_k$ are independent, so $k \leq \dim$ of the space.*

Proof. Pair $\sum_i c_i u_i = 0$ with v_j: every term with $i \neq j$ dies and $c_j \langle u_j, v_j \rangle = 0$, so $c_j = 0$. □

This is a two line proof and it is the entire upper bound in several problems. Whenever a problem says "this quantity is nonzero exactly when $i = j$", you are being handed a dual pairing and the answer is a dimension.

- **WHEN** a hypothesis reads "$X_{ij} \neq 0$ if and only if $i = j$" **DO** build vectors u_i, v_j in some space with $\langle u_i, v_j \rangle = X_{ij}$ and conclude $k \leq \dim$ **BECAUSE** Lemma 1.3.7 turns a diagonal nonvanishing pattern into independence, and independence is capped by dimension; your only real work is choosing the space

1.4 Row reduction, elementary matrices and why order matters

Row operations are not a procedure, they are left multiplication. Say it that way and a whole class of errors disappears.

Definition 1.4.1. An elementary matrix is one of: $E_{i \leftrightarrow j}$, the identity with rows i and j swapped (determinant -1); $D_i(\lambda)$, the identity with λ in position (i, i) (determinant λ); and $T_{ij}(\lambda) = I + \lambda e_i e_j^{\mathsf{T}}$ for $i \neq j$, which adds λ times row j to row i (determinant 1). Left multiplication by an elementary matrix performs the corresponding row operation; right multiplication performs the corresponding column operation.

A matrix with all diagonal entries 1 and zeros strictly below (or above) is called unitriangular. Products of unitriangular matrices are unitriangular, every unitriangular matrix is a product of the T_{ij}'s, and every unitriangular matrix has determinant 1. Hence:

Proposition 1.4.2. *Multiplying by a unitriangular matrix on either side leaves the determinant unchanged.*

Here is the practical consequence, and it is worth the paragraph. If you want to replace every row of A by a fixed linear combination of the original rows, do not write down a sequence of instructions, because you will update a row and then use the updated version as the subtrahend for the next one, which is a genuine error and not a clerical one. Write down the single matrix L that encodes the rule, and form LA. If L is unitriangular you have changed nothing about the determinant and you have changed all rows simultaneously.

Example 1.4.3 (Putnam 2014 A2)**.** Let A be the $n \times n$ matrix with $A_{ij} = 1/\min(i, j)$. Then $\det A = (-1)^{n-1}/\big(n!\,(n-1)!\big)$.

Do the general case, because the general case is easier and it is what you will recognise again. Suppose $A_{ij} = c_{\min(i,j)}$ for a sequence $c_1, c_2, \ldots$. Row 1 is constant, equal to c_1 everywhere. For $k \geq 2$, compare rows k and $k-1$: in column $j < k$ both entries are $c_{\min(k,j)} = c_j$ and $c_{\min(k-1,j)} = c_j$, so the difference is 0; in column $j \geq k$ the entries are c_k and c_{k-1}, so the difference is $c_k - c_{k-1}$. Let L be the unit lower bidiagonal matrix with 1 on the diagonal and -1 on the subdiagonal, so that (LA) has row 1 equal to row 1 of A and row k equal to row k minus row $k-1$ of A. Then LA is upper triangular with diagonal entries $c_1, c_2 - c_1, \ldots, c_n - c_{n-1}$. Since L is unitriangular,

$$\det A = \det(LA) = c_1 \prod_{k=2}^{n} (c_k - c_{k-1}).$$

Substituting $c_k = 1/k$ gives $c_k - c_{k-1} = -1/\big(k(k-1)\big)$, and

$$\det A = \prod_{k=2}^{n} \frac{-1}{k(k-1)} = \frac{(-1)^{n-1}}{n!\,(n-1)!}.$$

- **WHEN** the entry of a matrix depends on i and j only through $\min(i, j)$ or $\max(i, j)$ **DO** difference consecutive rows (equivalently, left multiply by the unit bidiagonal difference matrix) to triangularise **BECAUSE** min and max are constant on one side of the diagonal and constant along the other side too, so the difference of consecutive rows is supported on a tail, which is exactly triangular form; the multiplier is unitriangular so the determinant survives
- **WHEN** you are about to perform a chain of row operations by hand **DO** write the single matrix L implementing the whole chain and check $\det L$ **BECAUSE** a chain of instructions is order dependent and invites the skipped-calculation error of using an already updated row; one matrix identity is order free and certifiable
- **WHEN** a determinant problem asks for a specific numerical sequence in the entries **DO** solve it for a general sequence c_k first **BECAUSE** the general computation is usually shorter, and it converts the specific answer into a one line substitution instead of a fresh calculation

1.5 Determinants: three definitions and why they agree

A-level handed you the cofactor expansion and left it there. The cofactor expansion is the least useful of the three descriptions. Here they are, in order of usefulness on the Putnam.

Definition 1.5.1 (Characterisation)**.** det is the unique function D from n-tuples of columns in F^n to F that is (i) linear in each column separately, with the others held fixed; (ii) alternating, meaning $D = 0$ whenever two columns are equal; and (iii) normalised, $D(e_1, \ldots, e_n) = 1$.

Definition 1.5.2 (Leibniz)**.** $\det M = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} M_{i\sigma(i)}$.

Definition 1.5.3 (Laplace)**.** $\det M = \sum_j (-1)^{i+j} M_{ij} \det M^{(i,j)}$, where $M^{(i,j)}$ deletes row i and column j.

Theorem 1.5.4. *The three agree.*

Proof. Suppose D satisfies (i), (ii), (iii). Alternating plus multilinear forces antisymmetry: expanding $0 = D(\dots, u+v, \dots, u+v, \dots)$ gives $D(\dots, u, \dots, v, \dots) = -D(\dots, v, \dots, u, \dots)$. Now write each column $c_j = \sum_i M_{ij} e_i$ and expand by multilinearity in all n slots:

$$D(c_1, \dots, c_n) = \sum_{f:\{1,\dots,n\}\to\{1,\dots,n\}} \Big(\prod_j M_{f(j)j}\Big) D(e_{f(1)}, \dots, e_{f(n)}).$$

Every non-injective f contributes 0 by (ii). Every injective f is a permutation σ, and sorting its columns into the standard order costs $\operatorname{sgn}(\sigma)$ by antisymmetry, so $D(e_{\sigma(1)}, \dots, e_{\sigma(n)}) = \operatorname{sgn}(\sigma)$ by (iii). That is the Leibniz formula, and it proves both existence (the formula visibly satisfies (i), (ii), (iii)) and uniqueness. The Laplace expansion is the Leibniz sum with the terms grouped by the value of σ at one index. □

Definition 1.5.1 is the one to reach for. Everything follows from it in a line.

Proposition 1.5.5. *Let M be $n \times n$ over a field (or a commutative ring).*

1. *Swapping two rows or two columns multiplies* det *by* -1.
2. *Adding a multiple of one row to another leaves* det *unchanged.*
3. $\det M = 0$ *if and only if the rows are dependent, if and only if the columns are dependent, if and only if* $\operatorname{rank} M < n$.
4. $\det(MN) = \det M \det N$ *and* $\det M^{\mathsf{T}} = \det M$.
5. det *of a triangular matrix is the product of the diagonal entries.*
6. $M \operatorname{adj}(M) = \operatorname{adj}(M) M = (\det M) I$, *where* $\operatorname{adj}(M)_{ij} = (-1)^{i+j} \det M^{(j,i)}$.

Proof. (1) is antisymmetry; (2) is multilinearity plus alternating, since the correction term is a determinant with a repeated row; (5) is Leibniz with only the identity permutation surviving. For (4), fix N and consider $A \mapsto \det(AN)$ as a function of the rows of A: it is multilinear and alternating in those rows, so by uniqueness it equals $\det(A)$ times its value at $A = I$, which is $\det N$. Transpose invariance is the substitution $\sigma \mapsto \sigma^{-1}$ in Leibniz, which preserves the sign. For (6), the (i,i) entry of $M \operatorname{adj}(M)$ is the Laplace expansion of $\det M$ along row i; the (i,k) entry with $k \neq i$ is the Laplace expansion of the matrix obtained from M by replacing row k with row i, which has two equal rows, hence is 0. For (3), if the rows are dependent, subtract to create a zero row; conversely if they are independent they form a basis, so M is invertible and $\det M \det M^{-1} = 1$ forces $\det M \neq 0$. □

Fact (6) is Cramer's rule in disguise: if $\det M \neq 0$ then $M^{-1} = \operatorname{adj}(M) / \det M$ and the i-th coordinate of the solution of $Mx = b$ is $\det(M \text{ with column } i \text{ replaced by } b) / \det M$. The reason to keep the adjugate in mind rather than the inverse is that $\operatorname{adj}(M)$ is a polynomial in the entries of M, so it exists over any commutative ring and survives specialisation. That is exactly what you need when the entries are integers, polynomials, or a formal parameter.

- **WHEN** a determinant must be evaluated and the matrix has visible structure in its rows or columns **DO** use multilinearity and alternation to split, cancel or triangularise, never the cofactor expansion **BECAUSE** the cofactor expansion is $n!$ terms of bookkeeping; the characterisation is three properties, and structure in the rows is exactly what those three properties consume
- **WHEN** you need to solve a linear system whose answer must be shown to be a polynomial or an integer, not merely a number **DO** use Cramer with the adjugate rather than inverting **BECAUSE** adj and det are polynomials in the entries, so the rationality or integrality of the answer is visible from the formula; an inverse computed numerically hides it

Leibniz earns its keep exactly once: when you take an expectation

The Leibniz sum is the only representation of the determinant as a linear combination of monomials in the entries. So it is the only representation through which linearity of expectation can be pushed.

Example 1.5.6 (Putnam 2016 B4)**.** Let A be $2n \times 2n$ with entries independent and uniform on $\{0,1\}$. Then $\mathrm{E}[\det(A - A^{\mathsf{T}})] = \dfrac{(2n)!}{4^n n!}$.

Write $M = A - A^{\mathsf{T}}$, which is skew-symmetric: $M_{ii} = 0$ and $M_{ij} = -M_{ji}$, and for $i < j$ the entries M_{ij} are independent, each equal to $-1, 0, 1$ with probabilities $1/4, 1/2, 1/4$. So $\mathrm{E}[M_{ij}] = 0$ and $\mathrm{E}[M_{ij}^2] = 1/2$.

Expand by Leibniz and take expectations term by term: $\mathrm{E}[\det M] = \sum_\sigma \operatorname{sgn}(\sigma)\, \mathrm{E}\left[\prod_i M_{i\sigma(i)}\right]$. Split σ into cycles and split the product accordingly. Entries belonging to different cycles involve disjoint pairs $\{i, j\}$, hence are independent, so the expectation factorises over cycles. A fixed point contributes $M_{ii} = 0$, killing the term. A cycle of length $\ell \geq 3$ involves ℓ entries $M_{i_1 i_2}, \ldots, M_{i_\ell i_1}$ lying on distinct unordered pairs, hence independent with mean 0, so that factor has expectation 0. Only fixed point free involutions survive. Such a σ is a product of n transpositions, so $\operatorname{sgn}(\sigma) = (-1)^n$, and its product is $\prod(M_{ij}M_{ji}) = \prod(-M_{ij}^2)$, of expectation $(-1/2)^n$. Hence every surviving term equals $(-1)^n(-1/2)^n = 2^{-n}$, and the number of fixed point free involutions of $2n$ objects is $(2n-1)!! = (2n)!/(2^n n!)$. Multiply: $\mathrm{E}[\det M] = (2n)!/(4^n n!)$.

Before computing anything you could have predicted the shape of the answer. A skew-symmetric matrix of even size $2n$ has $\det M = \mathrm{Pf}(M)^2$, where the Pfaffian $\mathrm{Pf}(M) = \sum_\alpha \operatorname{sgn}(\alpha) \prod_\ell M_{i_\ell j_\ell}$ runs over perfect matchings α. That tells you in advance that the determinant is a square, hence the expectation is non-negative, and that perfect matchings will index the surviving terms. A skew-symmetric matrix of odd size has determinant 0, since $\det M = \det M^{\mathsf{T}} = \det(-M) = (-1)^n \det M$.

The statement actually proved is more general than the problem, so keep the general one: if M is $2n \times 2n$ skew-symmetric with the M_{ij} $(i < j)$ independent, $\mathrm{E}[M_{ij}] = 0$ and $\mathrm{E}[M_{ij}^2] = s_{ij}$, then $\mathrm{E}[\det M] = \sum_{\text{matchings}} \prod_\ell s_{a_\ell b_\ell}$, the hafnian of the second moment matrix.

- **WHEN** you need the expected value, or a coefficient, of a determinant **DO** expand by Leibniz and push the expectation inside, then kill permutations whose cycle structure forces a mean zero independent factor **BECAUSE** Leibniz is the unique expression of det as a sum of monomials, and expectation is linear on monomials; independence lets the expectation factorise across cycles
- **WHEN** the matrix is skew-symmetric **DO** write down $M_{ii} = 0$, note det $= 0$ for odd size, and expect a perfect square for even size **BECAUSE** $\det M = \mathrm{Pf}(M)^2$ for even size and $\det M = (-1)^n \det M$ for odd size, so the sign structure is decided before any computation

1.6 The determinant toolkit

This section is a bank. Each item is a matrix shape with a known determinant, and the circuit is always the same: recognise the shape, quote the value, do not expand.

Block determinants

Proposition 1.6.1. *If A and D are square then* $\det \begin{pmatrix} A & B \\ 0 & D \end{pmatrix} = \det \begin{pmatrix} A & 0 \\ C & D \end{pmatrix} = \det A \det D$. *If A is invertible then* $\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det A \det(D - CA^{-1}B)$, *the Schur complement formula, which follows from the factorisation*

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & 0 \\ CA^{-1} & I \end{pmatrix} \begin{pmatrix} A & B \\ 0 & D - CA^{-1}B \end{pmatrix}.$$

Swapping blocks costs a sign you must compute. Interchanging two contiguous blocks of k rows each multiplies the determinant by $(-1)^{k^2}$, because the swap is realised by exactly k^2 pairwise row crossings, one for each pair consisting of one row from each block.

Example 1.6.2 (Putnam 2018 A2)**.** Let $S_1, \dots, S_{2^n-1}$ be the nonempty subsets of $\{1, \dots, n\}$ in any order and let M_n have (i, j) entry 0 if $S_i \cap S_j = \emptyset$ and 1 otherwise. Then $\det M_n = 1$ for $n = 1$ and $\det M_n = -1$ for $n > 1$.

First, the determinant does not depend on the ordering of the subsets: relabelling by a transposition swaps two rows and the same two columns, so the determinant is multiplied by $(-1)^2 = 1$, and every permutation is a product of transpositions. That licence is what lets you choose a convenient recursive order, and it is worth stating explicitly in the arena because a grader will want it.

Order the subsets so that the first $2^{n-1} - 1$ are the nonempty subsets of $\{1, \dots, n-1\}$ in some order, the next $2^{n-1} - 1$ are those same subsets each with n adjoined, and the last is $\{n\}$. Writing $k = 2^{n-1} - 1$ and $M_{n-1} = (m'_{ij})$, the matrix becomes

$$M_n = \begin{pmatrix} M_{n-1} & M_{n-1} & 0 \\ M_{n-1} & J & \mathbf{1} \\ 0 & \mathbf{1}^{\mathsf{T}} & 1 \end{pmatrix},$$

where J is the all ones $k \times k$ block and $\mathbf{1}$ the all ones column, because two subsets of $\{1, \dots, n-1\}$ meet exactly as before, while any two sets both containing n always meet, and $\{n\}$ meets exactly the sets containing n. Now do two clearing steps, each of which is a row operation and therefore free. Subtract the last row $(0, \mathbf{1}^{\mathsf{T}}, 1)$ from every row of the middle block: that block becomes $(M_{n-1}, 0, 0)$. Then subtract the middle block from the top block: the top becomes $(0, M_{n-1}, 0)$. The matrix is now

$$\begin{pmatrix} 0 & M_{n-1} & 0 \\ M_{n-1} & 0 & 0 \\ 0 & \mathbf{1}^{\mathsf{T}} & 1 \end{pmatrix}.$$

Swap the two contiguous blocks of k rows, at the cost of $(-1)^{k^2}$, and the matrix is block lower triangular with diagonal blocks $M_{n-1}, M_{n-1}, (1)$. Since $k = 2^{n-1} - 1$ is odd for $n \geq 2$, the sign is -1. Hence $\det M_n = -(\det M_{n-1})^2$ for $n \geq 2$, and since $\det M_1 = 1$ the answer is -1 for every $n \geq 2$.

- **WHEN** a matrix is indexed by a set with a natural recursive decomposition (subsets, tuples, a group) **DO** order the index set so that the matrix becomes a block matrix whose blocks are copies of the smaller case **BECAUSE** a determinant of a block triangular matrix is the product of the diagonal blocks, so you get a recursion $D_n = f(D_{n-1})$ instead of a formula
- **WHEN** you reorder the index set of a symmetric index-defined matrix **DO** note that the determinant is unchanged **BECAUSE** a relabelling permutes rows and columns by the same permutation, contributing $\operatorname{sgn}(\sigma)^2 = 1$
- **WHEN** you must swap two contiguous blocks of k rows **DO** multiply by $(-1)^{k^2}$ **BECAUSE** the block swap is exactly k^2 pairwise row crossings, one for each ordered pair drawn one from each block

Sylvester's determinant identity

Theorem 1.6.3 (Sylvester)**.** *For U of size $m \times n$ and V of size $n \times m$ over any commutative ring,* $\det(I_m - UV) = \det(I_n - VU)$.

Proof. Factor the same block matrix two ways:

$$\begin{pmatrix} I_m & U \\ V & I_n \end{pmatrix} = \begin{pmatrix} I_m & 0 \\ V & I_n \end{pmatrix} \begin{pmatrix} I_m & U \\ 0 & I_n - VU \end{pmatrix} = \begin{pmatrix} I_m & U \\ 0 & I_n \end{pmatrix} \begin{pmatrix} I_m - UV & 0 \\ V & I_n \end{pmatrix}.$$

Both outer factors are unitriangular with determinant 1, so Proposition 1.6.1 gives $\det(I_n - VU)$ from the first and $\det(I_m - UV)$ from the second. □

Two corollaries you will use more often than the theorem. First, $\det(I + xy^{\mathsf{T}}) = 1 + y^{\mathsf{T}}x$, the case $m = n$, $n = 1$. Second, for square U, V the matrices UV and VU have the same characteristic polynomial: for $t \neq 0$, $\det(tI - UV) = t^n \det(I - t^{-1}UV) = t^n \det(I - t^{-1}VU) = \det(tI - VU)$, and two polynomials agreeing at infinitely many t are equal, so it holds at $t = 0$ too. No invertibility of U or V is needed anywhere.

Example 1.6.4 (Putnam 2015 A6)**.** A, B, M are real $n \times n$ with $AM = MB$ and A, B having the same characteristic polynomial. Then $\det(A - MX) = \det(B - XM)$ for every real $n \times n$ matrix X.

Set $A_t = A + tI$ and $B_t = B + tI$. From $AM = MB$ we get $A_t M = MB_t$ for every t. Since A and B have the same characteristic polynomial, $\det A_t = \det B_t$ for every t, and both are nonzero for all but finitely many t. Fix such a t. Then $A_t^{-1}M = MB_t^{-1}$, which is the substitution rule the whole problem turns on. Compute

$$\det(A_t - MX) = \det A_t \det\big(I - A_t^{-1}MX\big) = \det A_t \det\big(I - MB_t^{-1}X\big),$$

and by Theorem 1.6.3 with $U = M$ and $V = B_t^{-1}X$,

$$\det\big(I - MB_t^{-1}X\big) = \det\big(I - B_t^{-1}XM\big) = \frac{\det(B_t - XM)}{\det B_t}.$$

Since $\det A_t = \det B_t$, we get $\det(A_t - MX) = \det(B_t - XM)$ for all but finitely many t. Both sides are polynomials in t, so they agree for every t, in particular at $t = 0$. That is the claim.

Three things to notice, because each is a reusable circuit. The hypothesis "same characteristic polynomial" was consumed exactly once, as $\det A_t = \det B_t$ for every t; the weaker hypothesis $\det A = \det B$ would not have sufficed, and a nonzero nilpotent Jordan block against the zero matrix shows that equal characteristic polynomials do not imply similarity either. The relation $AM = MB$ says M intertwines the two operators; M carries generalised eigenspaces of B into generalised eigenspaces of A for the same eigenvalue, and if A and B have disjoint spectra then $M = 0$. And the set of such M is a linear subspace, never an open set, so you must never try to perturb M towards invertibility.

- **WHEN** a determinant contains a product UV that you would prefer to see as VU **DO** apply Sylvester's identity $\det(I - UV) = \det(I - VU)$ **BECAUSE** the block matrix $\left(\begin{smallmatrix} I & U \\ V & I \end{smallmatrix}\right)$ factors two ways through unitriangular matrices, and unitriangular factors are free
- **WHEN** a matrix in a determinant identity might be singular and you want to divide by it **DO** replace A by $A + tI$, prove the identity for all but finitely many t, and conclude by comparing polynomials in t **BECAUSE** both sides are polynomials in t; two polynomials agreeing at infinitely many points are identical, so the hypothesis of invertibility is free
- **WHEN** a hypothesis reads $AM = MB$ **DO** read it as "M intertwines" and use it as a substitution rule $A^{-1}M = MB^{-1}$ (after making A invertible by the $+tI$ move) **BECAUSE** intertwining is preserved by every polynomial in A and B, so it is a rewriting rule, not a single equation

The named determinants

Proposition 1.6.5 (Vandermonde)**.** $\det\left(t_i^{j-1}\right)_{i,j=1}^{k} = \prod_{a<b}(t_b - t_a)$.

Proof. Regard the determinant as a polynomial in t_k. It has degree at most $k - 1$, and it vanishes when t_k equals any of $t_1, \ldots, t_{k-1}$ because then two rows coincide. So it equals $c \prod_{a<k}(t_k - t_a)$, and comparing the coefficient of t_k^{k-1} identifies c as the Vandermonde determinant of $t_1, \ldots, t_{k-1}$. Induct. □

The consequence to install is not the formula but the statement it certifies: the points $(1, t, t^2, \dots, t^{k-1})$ on the moment curve, for distinct real t, have the property that *any* k of them are independent. That is the standard explicit certificate of general position, and it beats a genericity argument because it is one line and it is constructive.

Example 1.6.6 (Putnam 2025 A4, construction)**.** The construction there needs 2025 unit vectors in $\mathbb{R}^3$ with no two collinear and no three coplanar. Take $v_i = \big(1, t_i, t_i^2\big)/\|\cdot\|$ for distinct reals t_i: any three of them have determinant a nonzero Vandermonde, hence are independent, hence not coplanar.

Proposition 1.6.7 (Circulant)**.** *The $n \times n$ circulant C with first row $(a_0, \dots, a_{n-1})$, that is $C_{jk} = a_{k-j \bmod n}$, satisfies $C = \sum_k a_k S^k$ for the cyclic shift S. Since S has eigenvalues the n-th roots of unity ζ^j with eigenvectors $(1, \zeta^j, \zeta^{2j}, \dots)$, the eigenvalues of C are $\lambda_j = \sum_k a_k \zeta^{jk}$ and $\det C = \prod_j \lambda_j$.*

For $n = 3$ this reads $\det C = a^3 + b^3 + c^3 - 3abc = (a+b+c)(a^2+b^2+c^2-ab-bc-ca)$. That factorisation is precisely why Putnam 2019 A1 (determine all values of $A^3 + B^3 + C^3 - 3ABC$ for non-negative integers, the answer being all non-negative integers not congruent to 3 or 6 modulo 9) has a value set closed under multiplication: a product of two circulant determinants is the determinant of the product circulant, and circulants of the same size commute and multiply to circulants.

Proposition 1.6.8 ($aI + bJ$)**.** *Let J be the $n \times n$ all ones matrix. Then $J^2 = nJ$, the eigenvalues of J are n (on the all ones vector) and 0 with multiplicity $n-1$ (on the sum zero hyperplane), and*

$$\det(aI + bJ) = a^{n-1}(a + nb), \qquad (aJ + bI)(cJ + dI) = (acn + ad + bc)J + bdI.$$

This is the cheapest eigenvalue computation in the subject and it appears constantly, because "all diagonal entries equal, all off diagonal entries equal" is a very common hypothesis. In Putnam 2011 A4 the one line computation $(J - I)^2 = nJ + I$ over $\mathbb{F}_2$ decides the whole problem, and in Putnam 2021 A3 it evaluates the Gram determinants $(N-c)^3(N+3c)$ and $16k^3$ with no expansion at all.

Proposition 1.6.9 (Cauchy and Binet)**.** *If A is $m \times N$ and B is $N \times m$, then*

$$\det(AB) = \sum_K \det(A_{\cdot K}) \det(B_{K\cdot}),$$

summed over all m element subsets K of $\{1, \dots, N\}$, where $A_{\cdot K}$ takes the columns indexed by K and $B_{K\cdot}$ the rows. For $N = m$ this is multiplicativity; for $N < m$ the sum is empty and $\det(AB) = 0$. With $B = A^{\mathsf{T}}$ it says $\det(AA^{\mathsf{T}})$ is the sum of the squares of the maximal minors of A.

The proof is multilinearity: expand each column of AB as a combination of columns of A, discard the terms with a repeated column index, and collect by the underlying set K together with the sign of the ordering.

Example 1.6.10 (Putnam 2023 B6)**.** For $i, j \in \{1, \dots, n\}$ let $s(i,j)$ count the pairs (a,b) of non-negative integers with $ai + bj = n$, and let S be the matrix of these. Then $\det S = (-1)^{\lceil n/2 \rceil - 1}\, 2\lceil n/2 \rceil$.

The cue is that $s(i,j)$ counts something with a symmetric splitting. Set $c = bj$; then $ai = n - c$, so

$$s(i,j) = \#\{c \in \{0, \dots, n\} : j \mid c \text{ and } i \mid (n-c)\} = \sum_{c=0}^{n} A_{ic} B_{cj},$$

with $A_{ic} = [\, i \mid n - c \,]$ of size $n \times (n+1)$ and $B_{cj} = [\, j \mid c \,]$ of size $(n+1) \times n$. So $S = AB$ with an inner dimension one larger than the outer dimension. Cauchy and Binet then writes $\det S$ as a sum of exactly $n+1$ terms, one for each column of A you omit. The remaining work is evaluating those $n+1$ maximal minors, which is where multilinearity and column permutation signs (reversing m columns costs $(-1)^{m(m-1)/2}$) do the rest. What you should carry away is the first move: a counting function that splits as a sum over an auxiliary index is a matrix product, and the number of terms in Cauchy and Binet tells you at once that no square factorisation exists.

- **WHEN** a matrix entry is a count over an auxiliary index **DO** write it as $\sum_c A_{ic}B_{cj}$, that is, factor the matrix **BECAUSE** a sum over an index is literally a matrix product; once factored, rank and determinant tools apply that were invisible in the entrywise description
- **WHEN** you have factored $S = AB$ with inner dimension N larger than the outer dimension m **DO** apply Cauchy and Binet, giving $\binom{N}{m}$ terms **BECAUSE** multilinearity in the columns leaves exactly the terms indexed by m element subsets of the inner index; when $N = m + 1$ this is only $n + 1$ terms, which is a computation you can finish
- **WHEN** you have factored $S = AB$ with inner dimension smaller than the outer dimension **DO** write $\det S = 0$ immediately **BECAUSE** $\operatorname{rank}(AB) \le$ inner dimension $< m$, so S is singular, and Cauchy and Binet has an empty sum

Proposition 1.6.11 (Cauchy determinant)**.** $\det\left(\frac{1}{x_i + y_j}\right)_{i,j=1}^{n} = \frac{\prod_{i<j}(x_j - x_i)(y_j - y_i)}{\prod_{i,j}(x_i + y_j)}$.

Clear denominators and the numerator is a polynomial that vanishes whenever $x_i = x_j$ or $y_i = y_j$, which pins it down up to a constant by degree; the constant is 1 by looking at a leading term. The special case $x_i = y_i = (i-1)/2$ is the Hilbert matrix. In the window 2006 to 2025 the Cauchy determinant never appeared, so learn it last; it is here so that the shape $1/(x_i + y_j)$ does not stop you.

Hankel matrices and moment functionals

Definition 1.6.12. A Toeplitz matrix has M_{ij} depending only on $i - j$; a Hankel matrix has M_{ij} depending only on $i + j$. Write $H_n(\gamma) = (g_{i+j-1})_{i,j=1}^{n}$ for the Hankel matrix of a sequence $\gamma = (g_1, g_2, \dots)$ and $h_n(\gamma) = \det H_n(\gamma)$, with $h_0 = 1$.

A Hankel matrix is a Gram matrix in disguise, and that is the conceptual key. Define the linear functional L on polynomials by $L(x^k) = g_{k+1}$. Then H_n has (i, j) entry $L(x^{i-1}x^{j-1})$, so it is the Gram matrix of $1, x, \dots, x^{n-1}$ with respect to the bilinear form $\langle p, q\rangle = L(pq)$. Gaussian elimination on H_n is Gram-Schmidt on $1, x, x^2, \dots$, producing monic orthogonal polynomials B_k obeying a three term recurrence $B_{k+1} = (x - b_k)B_k - \lambda_k B_{k-1}$, and $UH_nU^{\mathsf{T}} = \operatorname{diag}\left(L(B_0^2), \dots, L(B_{n-1}^2)\right)$ with U unitriangular and $L(B_k^2) = \lambda_1 \cdots \lambda_k$. That is why Hankel determinants come out as products of continued fraction denominators: one elimination step is one level of the continued fraction.

Example 1.6.13 (Putnam 2024 A6)**.** With $c_0, c_1, \dots$ defined by $\left(1 - 3x - \sqrt{1 - 14x + 9x^2}\right)/4 = \sum_k c_k x^k$ and A the $n \times n$ matrix with (i, j) entry c_{i+j-1}, we have $\det A = 10^{n(n-1)/2}$.

The engine is a reduction lemma that peels one level at a time. Let $\alpha = x + a_2x^2 + \cdots$ and define $\beta = \sum_{i\ge 0} b_i x^i$ by $\alpha^{-1} = x^{-1} - \beta$. Then $h_n(\alpha) = h_{n-1}(\beta)$. The proof is exactly the unitriangular clearing of Proposition 1.4.2: the identity $x^{-1}\alpha = 1 + \alpha\beta$ gives $a_{m+1} = \sum_{r=1}^{m} a_r b_{m-r}$, and with L lower unitriangular ($L_{ik} = -b_{i-1-k}$ for $k < i$) and V upper unitriangular ($V_{lj} = a_{j-l+1}$ for $l \le j$) one checks that $LH_n(\alpha)$ has first column $(1, 0, \dots, 0)^{\mathsf{T}}$ and lower right block $H_{n-1}(\beta)V$. Expanding along that first column and using $\det L = \det V = 1$ gives the lemma. Unrolling one further step, writing $\beta = b_0 + \lambda_1\alpha_1$ with α_1 again of the normalised shape, gives $h_n(\alpha) = \lambda_1^{n-1}h_{n-1}(\alpha_1)$. For a continued fraction that reproduces itself, $\lambda_1 = \lambda$ at every level, and iterating gives $h_n = \lambda^{n(n-1)/2}$. Here $\lambda = 10$.

One further Hankel fact: $\det\left((i + j + n)!\right)_{i,j=0}^{m} = \prod_{k=0}^{m} k!\,(k + n)!$, proved by passing to the matrix of binomial coefficients $\binom{i+j+n}{i}$, whose determinant is 1 because decreasing n by one is a chain of elementary row operations. It appeared only inside Elkies' alternative solution to Putnam 2011 B6; learn it last.

- **WHEN** a matrix entry depends on $i + j$ only **DO** name it Hankel, introduce the moment functional $L(x^k) = g_{k+1}$, and run Gram-Schmidt (equivalently, unitriangular elimination) rather than expanding **BECAUSE** a Hankel matrix is the Gram matrix of $1, x, x^2, \dots$ for the form $L(pq)$, so elimination produces orthogonal polynomials and the determinant becomes a product of the three term recurrence coefficients
- **WHEN** a matrix entry depends on $i - j$ only **DO** name it Toeplitz, write its symbol $\sum_d c_d x^d$, and read kernel vectors as solutions of the linear recurrence with that characteristic polynomial **BECAUSE** the equation $Ma = 0$ for a banded Toeplitz matrix is literally a constant coefficient linear recurrence in the coordinates of a, subject to boundary conditions from the finite window
- **WHEN** a Hankel or Toeplitz determinant is asked for all n **DO** look for a reduction $h_n(\alpha) = \lambda^{n-1} h_{n-1}(\alpha')$ by unitriangular clearing, not for a closed form **BECAUSE** these families are self-similar: the $(n-1)$ case sits inside the n case after one elimination step, so the determinant satisfies a first order recursion that telescopes

Rank one updates

Proposition 1.6.14 (Matrix determinant lemma and Sherman and Morrison)**.** *For invertible A,* $\det(A + vw^{\mathsf{T}}) = \det(A)\big(1 + w^{\mathsf{T}} A^{-1} v\big)$*, and if that scalar is nonzero,*

$$(A + vw^{\mathsf{T}})^{-1} = A^{-1} - \frac{A^{-1} v w^{\mathsf{T}} A^{-1}}{1 + w^{\mathsf{T}} A^{-1} v}.$$

Also $(uw^{\mathsf{T}})^m = (w^{\mathsf{T}} u)^{m-1} uw^{\mathsf{T}}$ *for* $m \geq 1$*.*

Proof. $\det(A + vw^{\mathsf{T}}) = \det A \det(I + A^{-1} vw^{\mathsf{T}})$ and $\det(I + xy^{\mathsf{T}}) = 1 + y^{\mathsf{T}} x$ by Sylvester. The inverse formula is verified by multiplying out; the power formula is induction on the contraction $(uw^{\mathsf{T}})(uw^{\mathsf{T}}) = (w^{\mathsf{T}} u) uw^{\mathsf{T}}$. □

The contraction rule is worth saying in words, because it is what makes rank one algebra fast: a product of two outer products contracts in the middle, $(v\varphi)(w\psi) = \varphi(w)\, v\psi$, a scalar times a rank one matrix. In particular $v\varphi$ is idempotent exactly when $\varphi(v) = 1$, in which case it is the projection onto the line $\mathbb{R}v$ along the hyperplane $\ker \varphi$.

- **WHEN** a matrix appears as "something invertible plus something of rank one" **DO** use the matrix determinant lemma, and name the scalar $1 + w^{\mathsf{T}} A^{-1} v$ as the real object of the problem **BECAUSE** the entire change in the determinant is carried by that one scalar, so the problem is about a number, not about a matrix
- **WHEN** you meet a product of two rank one matrices **DO** contract in the middle: $(v\varphi)(w\psi) = \varphi(w) v\psi$ **BECAUSE** a rank one matrix is a column times a row, and matrix multiplication associates so that the inner row-times-column collapses to a scalar

1.7 Trace, and what it is invariant under

Definition 1.7.1. $\operatorname{tr} M = \sum_i M_{ii}$.

Proposition 1.7.2. $\operatorname{tr}(MN) = \operatorname{tr}(NM)$ *for M of size $m \times n$ and N of size $n \times m$; consequently* $\operatorname{tr}(P^{-1} AP) = \operatorname{tr} A$*, so trace is a similarity invariant and equals the sum of the eigenvalues with multiplicity. Also* $\operatorname{tr}(M^{\mathsf{T}} M) = \sum_{i,j} M_{ij}^2$ *over* $\mathbb{R}$*.*

Proof. $\operatorname{tr}(MN) = \sum_i \sum_a M_{ia} N_{ai} = \sum_a \sum_i N_{ai} M_{ia} = \operatorname{tr}(NM)$. The rest is substitution. □

The reason to be careful here is a trap the Putnam sets. The individual diagonal entries of a product are $(MN)_{tt} = \sum_a M_{ta}N_{at}$, the inner product of row t of M with column t of N. Their *sum* is similarity invariant. Their *product* is not invariant under any change of basis. So a hypothesis about all the diagonal entries of a product, rather than about their sum, decouples into n independent channels and cannot be attacked by rank, determinant or similarity. Recognising which of the two you have been given is the bifurcation point.

Example 1.7.3 (Putnam 2014 A6)**.** The largest k for which there exist $n \times n$ real matrices $M_1, \dots, M_k$ and $N_1, \dots, N_k$ such that M_iN_j has a zero entry somewhere on its diagonal if and only if $i \neq j$ is $k = n^n$.

The hypothesis is about each diagonal entry separately, so by the paragraph above it decouples: channel t compares row t of M_i with column t of N_j. In each channel you need a family of pairs with $r_a \cdot c_b \neq 0$ exactly when $a = b$, and Lemma 1.3.7 caps the number of distinguishable symbols per channel at n, achieved by the standard basis. That gives the construction: index the matrices by tuples $(i_1, \dots, i_n) \in \{1, \dots, n\}^n$, let $M_{(i_1,\dots,i_n)}$ have rows $e_{i_1}, \dots, e_{i_n}$ and let N be its transpose. Then the t-th diagonal entry of M_IN_J is $e_{i_t} \cdot e_{j_t}$, which is nonzero exactly when $i_t = j_t$; so all n diagonal entries are nonzero exactly when $I = J$. That realises $k = n^n$.

For the upper bound, multiply the channels together. Let V be the n fold tensor power of $\mathbb{R}^n$, with orthonormal basis $e_{a_1} \otimes \cdots \otimes e_{a_n}$ indexed by tuples, so $\dim V = n^n$. The tensor product of vectors satisfies $\langle w_1 \otimes \cdots \otimes w_n, y_1 \otimes \cdots \otimes y_n \rangle = \prod_t \langle w_t, y_t \rangle$. Let m_i be the tensor product of the rows of M_i and n_j the tensor product of the columns of N_j. Then

$$\langle m_i, n_j \rangle = \prod_{t=1}^{n} \left(M_iN_j\right)_{tt},$$

which is nonzero exactly when $i = j$. Lemma 1.3.7 gives $k \leq \dim V = n^n$.

That is the whole point of the tensor power: it is the space in which a *product* of n inner products becomes a *single* inner product, and its dimension is the answer.

- **WHEN** a hypothesis controls every diagonal entry of a product separately **DO** decouple into n channels and treat each independently; do not reach for rank, determinant or similarity **BECAUSE** only the sum of the diagonal entries is basis invariant, so any argument through similarity is answering a different question
- **WHEN** a condition is a product of n conditions, each a nonvanishing inner product **DO** move to the n fold tensor power, where the product of inner products is one inner product **BECAUSE** the tensor product satisfies $\langle \bigotimes w_t, \bigotimes y_t \rangle = \prod_t \langle w_t, y_t \rangle$, which is exactly the identity that turns n scalar conditions into one; the dimension n^n is then the bound
- **WHEN** you must bound a sum of squares of coordinates of several orthogonal vectors **DO** write the vectors as the columns of M and use $\operatorname{tr}(M^{\mathsf{T}}M) = \operatorname{tr}(MM^{\mathsf{T}})$ **BECAUSE** the same number counted by rows and by columns is the whole content of the trace identity, and it is the cheapest way to convert per coordinate bounds into a global bound

Example 1.7.4 (Putnam 2008 B3)**.** The largest circle contained in the unit 4-cube has radius $\sqrt{2}/2$.

A circle of radius r is $P + \cos(t)v_1 + \sin(t)v_2$ with $v_1 \perp v_2$, $\|v_1\| = \|v_2\| = r$. Its i-th coordinate sweeps an interval of length $2\sqrt{(v_1)_i^2 + (v_2)_i^2} \leq 1$, so $(v_1)_i^2 + (v_2)_i^2 \leq 1/4$; summing over the four coordinates gives $2r^2 = \|v_1\|^2 + \|v_2\|^2 \leq 1$, so $r \leq 1/\sqrt{2}$, attained. With M the $n \times k$ matrix of columns $v_1, \dots, v_k$ the same argument is a trace identity: $\operatorname{tr}(M^{\mathsf{T}}M) = kr^2$ while $\operatorname{tr}(MM^{\mathsf{T}}) \leq n/4$, so a k dimensional ball in the n-cube of side 1 has radius at most $\frac{1}{2}\sqrt{n/k}$, realised by an equal norm tight frame. For $k = 2$, $n = 4$ that is $\sqrt{2}/2$, the consistency check you run first.

1.8 Eigenvalues, characteristic and minimal polynomials

Definition 1.8.1. $\lambda \in F$ is an eigenvalue of A if $Av = \lambda v$ for some $v \neq 0$; equivalently $A - \lambda I$ is singular; equivalently $\chi_A(\lambda) = 0$ where $\chi_A(x) = \det(xI - A)$ is the characteristic polynomial, a monic polynomial of degree n whose coefficient of x^{n-1} is $-\operatorname{tr} A$ and whose constant term is $(-1)^n \det A$.

Read $\lambda = 1$ specially, because it is the case the Putnam uses most: 1 is an eigenvalue of M if and only if $M - I$ is singular, if and only if M has a nonzero fixed vector. So the hypothesis "1 is not an eigenvalue of Q" is not a negative statement at all, it is the gift of an invertible operator $(Q - I)^{-1}$.

Theorem 1.8.2 (Cayley and Hamilton)**.** $\chi_A(A) = 0$.

Proof. Over $\mathbb{C}$ (or any algebraically closed field), triangularise A as $P^{-1}AP = T$ upper triangular with diagonal $\lambda_1, \ldots, \lambda_n$; triangularisation exists because an eigenvector always exists over an algebraically closed field and you induct on the quotient space. Then $\chi_A(x) = \prod(x - \lambda_i)$, and $(T - \lambda_1 I) \cdots (T - \lambda_n I) = 0$ because the product of the first j factors kills the span of the first j standard basis vectors, by direct inspection of the triangular form. Conjugating back gives $\chi_A(A) = 0$. For a general field, embed in an algebraic closure; for a commutative ring, use the adjugate identity $\operatorname{adj}(xI - A)(xI - A) = \chi_A(x)I$ over $F[x]$ and compare coefficients. □

Cayley and Hamilton is the only source of a monic annihilating polynomial whose degree you know in advance. That is what makes it a counting tool: it says A^n is a fixed linear combination of $I, A, \ldots, A^{n-1}$, so the powers of A satisfy a linear recursion of order n, and the algebra generated by A has dimension at most n.

Definition 1.8.3. The set of polynomials annihilating A is an ideal of $F[x]$, hence generated by a unique monic polynomial μ_A, the minimal polynomial. It divides χ_A, has the same roots as χ_A (ignoring multiplicity), and has degree at most n.

Example 1.8.4 (Putnam 2015 B3)**.** For a 2×2 matrix M with $\operatorname{tr} M = \tau$ and $\det M = \delta$, $M^2 = \tau M - \delta I$. Hence every power of M lies in the span of $\{I, M\}$, and writing $M^k = t_k M + u_k I$ gives the scalar recursion $t_{k+1} = \tau t_k + u_k$, $u_{k+1} = -\delta t_k$.

That is Cayley and Hamilton in size 2, and it is worth checking by direct entrywise expansion once so that you never doubt it in the arena. In Putnam 2015 B3 the matrices whose entries a, b, c, d form an arithmetic progression make up a 2 dimensional subspace S of the 4 dimensional space of 2×2 matrices, cut out by two linear equations; both equations take the same value on I, so testing $M^k = t_k M + u_k I$ for membership of S collapses to the single condition $u_k = 0$. An exponentially growing computation has become one scalar condition. Together with the fact that a real 2×2 matrix with $\det < 0$ automatically has two real eigenvalues of opposite sign (the discriminant $\tau^2 - 4\delta$ is positive when $\delta < 0$, and the product of the roots is δ), the problem is settled with the exact irrational eigenvalues never written down.

- **WHEN** a problem asks about high powers of a fixed $n \times n$ matrix **DO** use $\chi_A(A) = 0$ to reduce every power to the span of $I, A, \ldots, A^{n-1}$, converting matrix powers into an order n scalar recursion **BECAUSE** the characteristic polynomial is monic of known degree n, so it is the unique guaranteed source of a finite relation among the powers
- **WHEN** a 2×2 matrix is raised to a power **DO** write $M^2 = (\operatorname{tr} M)M - (\det M)I$ and then $M^k = t_k M + u_k I$ **BECAUSE** in size 2 the Cayley and Hamilton relation is a single line of entrywise algebra, and it collapses the four dimensional matrix algebra onto a two dimensional span

- **WHEN** a hypothesis says "1 is not an eigenvalue" or "λ is not an eigenvalue" **DO** immediately write down the invertible operator $(Q - I)^{-1}$ and treat it as the main character **BECAUSE** a negative spectral hypothesis is exactly an invertibility statement, and an invertible operator is something you can apply, whereas a non-eigenvalue is not

Example 1.8.5 (Putnam 2010 B6)**.** Let A be a real $n \times n$ matrix and let $A^{[k]}$ denote the entrywise k-th power. If $A^k = A^{[k]}$ for $k = 1, \ldots, n+1$, then $A^k = A^{[k]}$ for all $k \geq 1$.

Two maps are in play on the space $W = xF[x]$ of polynomials with zero constant term. The first is $E(g) = g(A)$, the usual evaluation. The second is $F(g) = \big(g(A_{ij})\big)_{i,j}$, the entrywise evaluation. F is linear in g and is nothing but the direct product of the n^2 scalar evaluation homomorphisms at the entries; note that it is multiplicative for the entrywise product, not the matrix product, and that it sends the constant polynomial 1 to the all ones matrix J and not to I. That is exactly why the constant term must be excluded, and why the exponent range starts at 1.

The hypothesis says $E = F$ on $S = \{x, x^2, \ldots, x^{n+1}\}$. Let $P = \chi_A$ and consider $I_0 = xP(x)F[x]$. On I_0 both maps vanish: E vanishes because $AP(A) = 0$ by Cayley and Hamilton. For F, note that $xP(x)$ has degree $n+1$ and zero constant term, so it lies in $\operatorname{span}(S)$, whence $F\big(xP(x)\big) = E\big(xP(x)\big) = AP(A) = 0$. But $F\big(xP(x)\big)$ is the matrix with (i,j) entry $A_{ij}P(A_{ij})$, so every entry t of A satisfies $tP(t) = 0$; therefore $g(t) = tP(t)h(t) = 0$ for every $g = xP(x)h(x)$ and every entry t, that is, F vanishes on I_0. Finally $W = \operatorname{span}(S) + I_0$, because $\deg\big(xP(x)\big) = n+1$ so every element of W reduces modulo $xP(x)$ to a polynomial of degree at most n with zero constant term. Two linear maps agreeing on a spanning set and vanishing on a complementary subspace agree everywhere. Hence $E = F$ on all of W, which is the conclusion for every $k \geq 1$.

No induction is needed, and the number $n+1$ in the hypothesis is exactly $\deg\big(xP(x)\big)$. If you run the argument with the minimal polynomial μ_A instead, you find that exponents 1 to $\deg \mu_A + 1$ suffice for that particular A; the characteristic polynomial is the right choice only because the problem quantifies over all matrices of size n.

- **WHEN** a hypothesis holds for finitely many exponents and you must extend it to all exponents **DO** find a monic annihilating polynomial, reduce modulo it, and check that the two sides are linear maps agreeing on a spanning set **BECAUSE** reduction modulo a monic degree d polynomial expresses every power as a combination of the first d; if the hypothesis covers those and both sides kill the ideal, the two sides are equal as maps

Diagonalisation, triangularisation and the Jordan form

Theorem 1.8.6. *Over an algebraically closed field every matrix is similar to an upper triangular matrix. A matrix is diagonalisable if and only if its minimal polynomial is a product of distinct linear factors; in particular, n distinct eigenvalues force diagonalisability. Over any field, a matrix is similar to a direct sum of Jordan blocks $\lambda I + N$ with N the nilpotent shift, provided the characteristic polynomial splits.*

The Putnam almost never needs the full Jordan form. It needs three of its consequences, and you should hold those three rather than the classification.

First, a nilpotent $n \times n$ matrix satisfies $N^n = 0$, because its characteristic polynomial is x^n. That truncation is what makes $(I+N)^j = \sum_{r<n} \binom{j}{r} N^r$ a sum of at most n terms, which is the whole leverage in Putnam 2012 A5.

Second, equal characteristic polynomials do not imply similarity: a nonzero 2×2 nilpotent Jordan block and the zero matrix both have characteristic polynomial x^2. If you need similarity you need the minimal polynomial too, or the full Jordan data.

Third, every matrix of rank r can be written $U \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} V^{-1}$ with U, V invertible. That rank normal form is the structural statement behind the block decomposition solutions: on the image of

M an intertwining relation is a similarity, while on the kernel only characteristic polynomial data survive.

Example 1.8.7 (Putnam 2012 A5)**.** For $G(x) = v + Mx$ on $\mathbb{F}_p^n$, the k-fold composite is $G^{(k)}(x) = M^k x + (I + M + \cdots + M^{k-1})v$, and two affine maps are equal exactly when their linear parts and their constant terms agree.

That closed form is the reason the problem splits: the demand that $G^{(p^n)}$ be the identity separates into $M^{p^n} = I$ and the vanishing of the constant term. If $I - M$ is invertible the geometric series telescopes to $(I - M)^{-1}(I - M^k)v$; if $I - M$ is nilpotent, which is the case here, you must instead expand $M^j = (I + N)^j$ binomially, and $N^n = 0$ truncates the expansion to n terms. The answer is that the required v and M exist exactly for $n = 1$ with any p, together with the single pair $(p, n) = (2, 2)$.

- **WHEN** an iteration is affine, $x_{k+1} = Mx_k + v$ **DO** write the closed form $x_k = M^k x_0 + \sum_{j<k} M^j v$ and separate the linear part from the constant term **BECAUSE** an affine map is determined by exactly those two pieces, so an identity between affine maps is two independent identities
- **WHEN** the geometric series $I + M + \cdots + M^{k-1}$ appears and $I - M$ is not invertible **DO** write $M = I + N$ with N nilpotent and expand binomially **BECAUSE** when $I - M$ is nilpotent the telescoping formula is unavailable, but $N^n = 0$ truncates the binomial expansion to n terms, which is just as finite

1.9 Symmetric matrices, quadratic forms and positive definiteness

Definition 1.9.1. S is symmetric if $S^{\mathsf{T}} = S$, skew if $S^{\mathsf{T}} = -S$, Hermitian if $S^* = S$ where $*$ is conjugate transpose. A quadratic form is $Q(v) = v^{\mathsf{T}} M v$.

Proposition 1.9.2. *For any square M, $v^{\mathsf{T}} M v = v^{\mathsf{T}} S v$ with $S = \frac{1}{2}(M + M^{\mathsf{T}})$, because the skew part K satisfies $v^{\mathsf{T}} K v = (v^{\mathsf{T}} K v)^{\mathsf{T}} = -v^{\mathsf{T}} K v = 0$.*

So a quadratic form only ever sees the symmetric part of a matrix. That one observation explains a whole class of Putnam hypotheses which look asymmetric but are not.

Theorem 1.9.3 (Spectral theorem)**.** *A real symmetric matrix has real eigenvalues and an orthonormal basis of eigenvectors; equivalently $S = O \Lambda O^{\mathsf{T}}$ with O orthogonal and Λ real diagonal. The same holds for Hermitian matrices with O unitary.*

Proof. Reality: if $Sv = \lambda v$ with $v \neq 0$ complex, then $\bar{v}^{\mathsf{T}} S v$ equals both $\lambda \|v\|^2$ and $\bar{\lambda}\|v\|^2$, so $\lambda \in \mathbb{R}$. An eigenvalue exists over $\mathbb{C}$ because χ_S has a complex root, and it is real by the above, so $S - \lambda I$ is a singular *real* matrix and has a real kernel vector. Now if $Sv = \lambda v$ then S maps $v^{\perp}$ into itself, since $\langle Sw, v\rangle = \langle w, Sv\rangle = \lambda \langle w, v\rangle = 0$; induct on $v^{\perp}$. □

Definition 1.9.4. A symmetric S is positive definite if $v^{\mathsf{T}} S v > 0$ for all $v \neq 0$, positive semidefinite if ≥ 0.

Theorem 1.9.5 (Equivalences)**.** *For real symmetric S the following are equivalent: (i) S is positive definite; (ii) all eigenvalues of S are positive; (iii) all leading principal minors of S are positive (Sylvester's criterion); (iv) $S = A^{\mathsf{T}} A$ for some A with independent columns. For the semidefinite version, replace "positive" by "non-negative" throughout, drop the independence in (iv), and note that (iii) must then be replaced by "all principal minors, not just the leading ones, are non-negative".*

In size 2 this is worth memorising in explicit form, because it appears verbatim. For real A, B, C the form $Q(v) = Av_1^2 + Bv_1v_2 + Cv_2^2$ is positive for all $v \neq 0$ exactly when $A > 0$ and $4AC - B^2 > 0$, and the certificate is the completed square

$$A\,Q(v) = \left(Av_1 + \tfrac{1}{2}Bv_2\right)^2 + \left(AC - \tfrac{1}{4}B^2\right)v_2^2.$$

Notice that the certificate is what you write in the arena; the criterion is what tells you it will work.

Example 1.9.6 (Putnam 2018 B5)**.** Let $f = (f_1, f_2) : \mathbb{R}^2 \to \mathbb{R}^2$ have continuous partial derivatives, all positive, with $\partial_1 f_1 \, \partial_2 f_2 - \frac{1}{4}(\partial_2 f_1 + \partial_1 f_2)^2 > 0$ everywhere. Then f is injective.

The hypothesis is written asymmetrically in $\partial_2 f_1$ and $\partial_1 f_2$, and it *averages* them rather than multiplying them. Proposition 1.9.2 says why: the condition is positive definiteness of the symmetrised Jacobian $S = \frac{1}{2}\big(Df + (Df)^{\mathsf{T}}\big)$, not positivity of $\det Df$. Indeed with $A = \partial_1 f_1$, $B = \partial_2 f_1 + \partial_1 f_2$, $C = \partial_2 f_2$, the hypothesis is exactly $A > 0$ and $AC - \frac{1}{4}B^2 > 0$, which is Sylvester's criterion for S.

Now injectivity in three lines. Take distinct points a, a', set $v = a' - a \neq 0$, let $\gamma(t) = a + tv$ for $t \in [0,1]$ and put $g(t) = v \cdot f(\gamma(t))$. By the chain rule $g'(t) = v^{\mathsf{T}} Df(\gamma(t))v = v^{\mathsf{T}} S v > 0$. So g is strictly increasing, $g(1) > g(0)$, hence $v \cdot f(a') \neq v \cdot f(a)$, hence $f(a') \neq f(a)$.

- **WHEN** a hypothesis involves a matrix inside a quadratic form, or averages two off diagonal entries **DO** replace the matrix by its symmetric part and read the hypothesis as positive definiteness **BECAUSE** a quadratic form only sees the symmetric part, since the skew part contributes $v^{\mathsf{T}} K v = 0$; averaging the off diagonal entries is precisely symmetrisation
- **WHEN** you must prove a map $\mathbb{R}^n \to \mathbb{R}^n$ is injective and its Jacobian has a definiteness property **DO** take the straight line between two points and differentiate the scalar $g(t) = v \cdot f(\gamma(t))$ **BECAUSE** $g'(t) = v^{\mathsf{T}} Df\, v$ is exactly the quadratic form applied to the displacement, so definiteness makes g strictly monotone and the endpoints cannot agree
- **WHEN** you have a 2×2 positive definiteness claim to justify **DO** write the completed square $AQ = (Av_1 + \frac{1}{2}Bv_2)^2 + (AC - \frac{1}{4}B^2)v_2^2$ **BECAUSE** a completed square is a self-checking certificate that a grader can verify in one line, whereas "by Sylvester's criterion" asks them to trust a citation

1.10 Inner products, orthogonality and Gram matrices

Definition 1.10.1. A real inner product space is a vector space with a symmetric bilinear form $\langle \cdot, \cdot \rangle$ that is positive definite. Orthogonality is $\langle u, v \rangle = 0$; the norm is $\|v\| = \sqrt{\langle v, v \rangle}$.

Proposition 1.10.2 (Polarisation)**.** $2\langle u, v \rangle = \|u\|^2 + \|v\|^2 - \|u - v\|^2$.

So the inner product is determined by the norm. This is the translation step you need whenever a problem gives you distances and you want dot products, which is most of the time in geometry.

Proposition 1.10.3. *Nonzero pairwise orthogonal vectors are independent, so there are at most* dim *of them; and a vector orthogonal to every element of a spanning set is zero.*

Proof. Pair a vanishing combination with each vector in turn; the second statement is positive definiteness applied to $\langle v, v \rangle = 0$. □

Those two sentences cap the number of directions in Putnam 2022 B2 at three (the answer to which is that $|S| \in \{1, 7\}$), and they are the reason the kernel argument in Putnam 2010 A5 closes: for fixed $e \in \mathbb{R}^3$ the map $v \mapsto e \times v$ is linear with kernel exactly $\mathbb{R}e$ when $e \neq 0$, so a vector whose cross product with every element of a spanning set vanishes must be zero.

Gram matrices

Definition 1.10.4. For $v_1, \ldots, v_k$ in an inner product space, the Gram matrix is $G_{ij} = \langle v_i, v_j \rangle$. If the v_i are the rows of a matrix M then $G = MM^{\mathsf{T}}$; if they are the columns, $G = M^{\mathsf{T}} M$.

Theorem 1.10.5. *Over $\mathbb{R}$: G is symmetric positive semidefinite;* $\operatorname{rank} G = \operatorname{rank} M$; $\det G \geq 0$, *with* $\det G = 0$ *exactly when the v_i are dependent; if $k = n$ then* $\det G = (\det M)^2$; *and in general* $\det G = \sum_K \det(M_{\cdot K})^2$ *by Cauchy and Binet, so* $\det G$ *is a sum of squares of maximal minors. The k dimensional volume of the simplex spanned by the v_i is $\sqrt{\det G}/k!$. Over $\mathbb{C}$: M^*M is Hermitian positive semidefinite, a Hermitian matrix is of this form exactly when it is positive semidefinite, and for square M,* $\det(M^*M) = |\det M|^2 \geq 0$.

A Gram matrix is the single highest yield recognition in this chapter, because naming it hands you four things at once, for free: symmetry (or Hermitian symmetry), real eigenvalues, non-negativity of the determinant, and a dependency criterion. If a matrix in a problem consists of dot products, you are looking at a Gram matrix and you should say so on line one.

Example 1.10.6 (Putnam 2018 A6)**.** A, B, C, D are distinct points in the plane, no three collinear, and all six squared distances are rational. Then $\operatorname{area}(ABC) / \operatorname{area}(ABD)$ is rational.

Base everything at A: put $b = B - A$, $c = C - A$, $d = D - A$. By polarisation, all six pairwise inner products of b, c, d are rational, since every squared distance is rational. Now the areas: $\operatorname{area}(ABC) = \frac{1}{2}|\det[b\ c]|$ and $\operatorname{area}(ABD) = \frac{1}{2}|\det[b\ d]|$. By multiplicativity and transpose invariance,

$$\det[b\ c] \cdot \det[b\ d] = \det\left([b\ c]^{\mathsf{T}}[b\ d]\right) = \det \begin{pmatrix} b \cdot b & b \cdot d \\ c \cdot b & c \cdot d \end{pmatrix},$$

which is rational, and likewise $\left(\det[b\ d]\right)^2$ is the Gram determinant of b, d, rational and nonzero because b, d are independent (no three of the points are collinear). Divide:

$$\frac{\det[b\ c]}{\det[b\ d]} = \frac{\det[b\ c]\det[b\ d]}{\left(\det[b\ d]\right)^2} \in \mathbb{Q}.$$

Taking absolute values gives the area ratio.

Now see exactly where the planarity was used. The identity $\det M \det N = \det(M^{\mathsf{T}}N)$ holds only for square M, N. For $N \times k$ matrices with $N > k$, Cauchy and Binet replaces the single product by a sum over k element subsets of rows, and the argument collapses. That is the load bearing hypothesis, and being able to say so is the difference between a solution and a sketch.

- **WHEN** a hypothesis gives you distances or squared lengths and you want algebra **DO** polarise to inner products **BECAUSE** the inner product is determined by the norm through $2\langle u, v\rangle = \|u\|^2 + \|v\|^2 - \|u - v\|^2$, so rationality or integrality of squared distances transfers directly to the dot products
- **WHEN** a matrix consists of pairwise dot products **DO** name it a Gram matrix and read off symmetry, $\det \geq 0$, and "singular if and only if dependent" **BECAUSE** $G = MM^{\mathsf{T}}$, and a product of a matrix with its own transpose is forced to be positive semidefinite of the same rank as M
- **WHEN** a product of two determinants of the same size must be shown rational or integral **DO** convert it into a single determinant of dot products via $\det M \det N = \det(M^{\mathsf{T}}N)$ **BECAUSE** multiplicativity plus transpose invariance turns the product of two unknown determinants into one determinant with known entries; this works only in the square case, which is where a dimension hypothesis is being consumed

Example 1.10.7 (Putnam 2021 A3)**.** The integers N for which the sphere $x^2 + y^2 + z^2 = N$ has an inscribed regular tetrahedron with integer vertices are exactly $N = 3m^2$.

Let $v_1, \ldots, v_4$ be the vertex vectors. The centroid of a regular tetrahedron is its circumcentre, so $\sum v_i = 0$; each $|v_i|^2 = N$; and by symmetry $v_i \cdot v_j = c$ for $i \neq j$. Squaring $\sum v_i = 0$ gives

$4N + 12c = 0$, so $c = -N/3$, and since the v_i are integer vectors, c is an integer and $3 \mid N$. Write $N = 3k$, so $c = -k$. The Gram matrix of v_1, v_2, v_3 is $(N - c)I + cJ$ of size 3, whose determinant is $(N - c)^2(N + 2c) = (4k)^2(k) = 16k^3$ by the $aI + bJ$ formula. But that Gram matrix is MM^{T} for the integer matrix M with rows v_1, v_2, v_3, so $16k^3 = (\det M)^2$ is a perfect square, forcing k^3 and hence k to be a perfect square. Thus $N = 3m^2$. Conversely $(m, m, m), (m, -m, -m), (-m, m, -m), (-m, -m, m)$ works.

The generalisation is worth recording, because it makes the same argument available in every dimension: for a regular n-simplex centred at the origin with circumradius squared N, the common inner product is $c = -N/n$ and the Gram determinant of n vertex vectors is $N^n(n + 1)^{n-1}/n^n$. In odd dimension there is a companion fact: if E is an $n \times n$ integer matrix with $EE^{\mathsf{T}} = \lambda I$ and n is odd, then $\lambda^{n/2} = |\det E|$ is an integer, so λ must be a perfect square. This fails in even dimension, as the rows $(1, 1)$ and $(1, -1)$ with $\lambda = 2$ show, which is exactly why the planar analogue of this problem has a different answer.

- **WHEN** integer vectors have a Gram matrix whose determinant you can compute **DO** demand that it be a perfect square **BECAUSE** $\det(MM^{\mathsf{T}}) = (\det M)^2$ and $\det M$ is an integer, so the Gram determinant is a Diophantine constraint, not just a number
- **WHEN** a configuration is highly symmetric, so all self inner products are equal and all cross inner products are equal **DO** write the Gram matrix as $aI + bJ$ and quote $\det = a^{n-1}(a + nb)$ **BECAUSE** the all ones matrix has eigenvalues n and 0, so any matrix of the form $aI + bJ$ is diagonal in the same basis and its determinant needs no expansion

Example 1.10.8 (Putnam 2011 B4)**.** In a tournament, 2011 players play 2011 games; matrices T and W accumulate ties and wins as described. Then $\det(T + iW)$ is a non-negative integer divisible by 2^{2010}.

Two recognitions do the whole problem. First, T is symmetric and W is antisymmetric, so $T + iW$ is Hermitian and its determinant is real. That is the reason the problem introduced i at all, and it is the sanity check you run before searching for structure. Second, the update rule adds a rank one Hermitian correction after every game, so $T + iW = \sum_g v_g^* v_g = A^*A$ where A is the 2011×2011 matrix whose (g, k) entry is 1 if player k won game g and i if they lost. Hence $\det(T + iW) = |\det A|^2 \geq 0$, real and non-negative with no computation at all.

For the divisibility, use a lemma worth installing in its own right. If every entry of an $n \times n$ matrix M over a commutative ring is congruent to every other entry modulo π, then $\pi^{n-1} \mid \det M$: subtract row 1 from every other row, which preserves the determinant and makes those $n - 1$ rows entrywise divisible by π, then pull π out of each by multilinearity in the rows. Here the entries of A are 1 and i, which are congruent modulo $\pi = 1 - i$, so $(1 - i)^{2010} \mid \det A$, and taking absolute values squared gives $|1 - i|^{2 \cdot 2010} = 2^{2010}$ dividing $|\det A|^2$. The general form of the lemma is that if the rows fall into r congruence classes modulo π then $\pi^{n-r} \mid \det M$.

Note the trap the problem sets: because a rank one Hermitian correction is added at each step, you could try the matrix determinant lemma $\det(M + uu^*) = \det M\,(1 + u^*M^{-1}u)$ once per game. That is 2011 determinant updates and it is the wrong tool. The right move is to recognise the accumulation $M_g = M_{g-1} + v_g^* v_g$ as having the closed form A^*A, and if you want det in the non-square case, Cauchy and Binet gives it as a sum of squared minors rather than an iterated product.

- **WHEN** a process adds a rank one correction at each step **DO** write the closed form $\sum_g v_g^* v_g = A^*A$ rather than iterating an update formula **BECAUSE** the accumulated sum of outer products is literally a Gram matrix, and Gram gives you rank, reality and non-negativity in one line, whereas the step by step determinant lemma gives you g nested expressions
- **WHEN** a symmetric statistic and an antisymmetric statistic appear together, or a problem inserts an unexplained i **DO** form $T + iW$ and check it is Hermitian **BECAUSE** the real and

imaginary parts of a Hermitian matrix are exactly a symmetric and an antisymmetric matrix, so the packaging is forced, and it buys real eigenvalues and a real determinant

- **WHEN** all entries of an integer or Gaussian integer matrix are congruent modulo some π **DO** conclude $\pi^{n-1} \mid \det$ **BECAUSE** clearing rows against row 1 preserves the determinant and makes $n-1$ rows divisible by π, and multilinearity in the rows then extracts π^{n-1}

Orthogonal matrices and reflections

Definition 1.10.9. Q is orthogonal if $Q^{\mathsf{T}}Q = I$; equivalently $\|Qx\| = \|x\|$ for all x; equivalently $\langle Qx, Qy\rangle = \langle x, y\rangle$ for all x, y. The equivalence is polarisation again.

Two consequences are used constantly. All eigenvalues have modulus 1 and the non-real ones come in conjugate pairs of product 1; so if 1 is not an eigenvalue, the multiplicity of -1 has the same parity as n and $\det Q = (-1)^n$. And $\langle (Q-I)x, (Q+I)x\rangle = \|Qx\|^2 - \|x\|^2 = 0$ for every x.

Definition 1.10.10 (Householder reflection)**.** For a unit $u \in \mathbb{R}^n$, $P = I - 2uu^{\mathsf{T}}$ satisfies $P^{\mathsf{T}} = P$, $P^2 = I$, $Pu = -u$, $Pv = v$ for $v \perp u$, and $Pw = w - 2(u^{\mathsf{T}}w)u$ in general. It is the reflection in the hyperplane $u^{\perp}$, and $\det P = -1$.

Example 1.10.11 (Putnam 2019 B3)**.** Q is real orthogonal, u a unit vector, $P = I - 2uu^{\mathsf{T}}$. If 1 is not an eigenvalue of Q, then 1 is an eigenvalue of PQ.

Read the hypothesis as a gift: $Q - I$ is invertible. Read the conclusion as an object to exhibit: find $w \neq 0$ with $PQw = w$. Since $P^2 = I$, that is $Qw = Pw = w - 2(u^{\mathsf{T}}w)u$, that is

$$(Q-I)w = -2(u^{\mathsf{T}}w)\,u.$$

So w must be a multiple of $(Q-I)^{-1}u$. Try $w = (Q-I)^{-1}u$ exactly, so $Qw = w + u$. Now use orthogonality once:

$$\|w\|^2 = \|Qw\|^2 = \|w+u\|^2 = \|w\|^2 + 2u^{\mathsf{T}}w + 1 \quad\Longrightarrow\quad u^{\mathsf{T}}w = -\tfrac{1}{2}.$$

Then $PQw = P(w+u) = \big(w - 2(u^{\mathsf{T}}w)u\big) + Pu = (w+u) - u = w$, and $w \neq 0$ because $u \neq 0$. Done.

That single line $\|Qw\| = \|w\|$ is the entire content of the problem, and the scalar $u^{\mathsf{T}}(Q-I)^{-1}u = -\frac{1}{2}$ it produces is the object the problem is really about. Three further remarks, each a different lens on the same fact, and each occasionally the faster route. First, the matrix determinant lemma: $PQ - I = (Q-I) - 2u\big(Q^{\mathsf{T}}u\big)^{\mathsf{T}}$, so

$$\det(PQ-I) = \det(Q-I)\Big(1 - 2\,u^{\mathsf{T}}Q(Q-I)^{-1}u\Big),$$

and since $Q(Q-I)^{-1} = I + (Q-I)^{-1}$ the scalar is $1 - 2\big(1 - \frac{1}{2}\big) = 0$. Second, parity: $\det P = -1$ and, because 1 is not an eigenvalue of Q, $\det Q = (-1)^n$, so $\det(PQ) = (-1)^{n+1}$; if 1 were not an eigenvalue of the orthogonal matrix PQ then the same parity count would force $\det(PQ) = (-1)^n$, a contradiction. Third, the identity $u^{\mathsf{T}}(Q-I)^{-1}u = -\frac{1}{2}$ for every unit u says exactly that $(Q-I)^{-1} + \frac{1}{2}I$ has vanishing symmetric part, that is, it is skew: this is the Cayley transform, the general rational parametrisation of the orthogonal group, and it also explains the fact that every orthogonal matrix is a product of at most n Householder reflections, with exactly n needed precisely when 1 is not an eigenvalue.

- **WHEN** a problem involves a matrix built as $I - 2uu^{\mathsf{T}}$ or any P with $P^2 = I$ **DO** use $P^{-1} = P$ to move P across an equation rather than inverting anything **BECAUSE** an involution is its own inverse, so $PQw = w$ and $Qw = Pw$ are the same equation, and the second has all the structure on one side

- **WHEN** you must produce an eigenvector and a hypothesis makes some $M - \lambda I$ invertible **DO** write the candidate as $(M - \lambda I)^{-1}$ applied to the obvious vector and then verify **BECAUSE** the eigenvector equation rearranges to say the vector lies in a one dimensional image, so the inverse operator names it; verification is then a computation with no search
- **WHEN** a step needs a scalar identity about an orthogonal Q **DO** apply $\|Qx\| = \|x\|$ to the specific vector in play and expand **BECAUSE** orthogonality is a statement about every vector, so applying it to the one vector the problem has produced converts a global hypothesis into the exact scalar you need

1.11 Linear algebra over a finite field

This is the section that pays the best return per page, because A-level never suggested that any of this works outside $\mathbb{R}$, and the Putnam exploits that gap every other year.

Theorem 1.11.1. *For a prime p, $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ is a field, and more generally there is a field $\mathbb{F}_q$ with $q = p^e$ elements for every prime power. Every theorem of linear algebra over an abstract field holds verbatim over $\mathbb{F}_q$: rank, kernel, rank and nullity, row reduction, the determinant characterisation and its multiplicativity, invertibility equivalent to nonzero determinant, Cayley and Hamilton, and the fact that a square matrix with a one sided inverse is invertible. Nothing that relies on order, positivity or Archimedean size transfers, and in particular positive definiteness, $\det(MM^{\mathsf{T}}) \geq 0$, and $\operatorname{rank}(MM^{\mathsf{T}}) = \operatorname{rank} M$ all fail.*

Proposition 1.11.2 (Counting)**.** *A subspace of $\mathbb{F}_q^n$ of dimension d has exactly q^d elements. For a linear $T : \mathbb{F}_q^m \to \mathbb{F}_q^n$ and any subset S of the target,*

$$\#\{x : T(x) \in S\} = \#\big(S \cap \operatorname{im} T\big) \cdot q^{\,m - \operatorname{rank} T},$$

because every nonempty fibre is a coset of $\ker T$ and hence has exactly $|\ker T| = q^{m-\operatorname{rank} T}$ elements. Also

$$\big|\mathrm{GL}_n(\mathbb{F}_q)\big| = \prod_{i=0}^{n-1} \big(q^n - q^i\big) = q^{n(n-1)/2} \prod_{i=1}^{n} \big(q^i - 1\big),$$

by choosing the columns in order, each avoiding the span of its predecessors.

Example 1.11.3 (Putnam 2017 A6)**.** The number of ways to colour the 30 edges of an icosahedron red, white or blue so that every one of the 20 triangular faces has exactly two edges of one colour and the third of another is $2^{20}3^{10}$.

Identify the three colours with the elements of $\mathbb{F}_3$. For three elements a, b, c of $\mathbb{F}_3$, $a + b + c = 0$ happens exactly when they are all equal or all distinct; so "two the same and one different" is exactly $a + b + c \neq 0$. Define $T : \mathbb{F}_3^E \to \mathbb{F}_3^F$ sending a colouring to the vector of face sums, a linear map from a 30 dimensional space to a 20 dimensional one. You must count the preimages of the 2^{20} vectors with all coordinates nonzero.

If T is surjective, Proposition 1.11.2 finishes it: every fibre has $3^{30-20} = 3^{10}$ elements and there are 2^{20} admissible targets, giving $2^{20}3^{10}$. Surjectivity is proved by a local gadget. The five faces meeting at a vertex form a cycle of length five in which consecutive faces share an edge, and prescribing the values on those five faces amounts to solving the cyclic system whose matrix is $I + S$ with S the cyclic shift on five coordinates. Its determinant is $1 - (-1)^5 = 2 \neq 0$ in $\mathbb{F}_3$, so the system is solvable: the odd cycle is invertible, the even cycle is not. Explicitly, over $\mathbb{F}_3$ the alternating assignment $2, 1, 2, 1, 2$ around an odd cycle makes every adjacent sum vanish except one, producing a single standard basis vector in the image, and those span.

- **WHEN** a combinatorial condition on three objects reads "not all the same and not all different", or "exactly two agree" **DO** identify the objects with $\mathbb{F}_3$ and rewrite the condition as a sum being nonzero **BECAUSE** in $\mathbb{F}_3$, $a + b + c = 0$ characterises "all equal or all distinct", so the awkward combinatorial condition is a single linear expression

- **WHEN** you must count colourings, subsets or assignments subject to linear constraints over a finite alphabet of prime size **DO** build the linear map, prove surjectivity, and multiply (number of admissible targets) by (fibre size q^{nullity}) **BECAUSE** all nonempty fibres of a linear map are cosets of the kernel, hence equinumerous, so the count factors into a target count and one exponential
- **WHEN** a cyclic system $x_{i+1} = t_i - x_i$ appears **DO** unwind it to $x_1 = (\text{alternating sum of } t_i) + (-1)^n x_1$ **BECAUSE** for odd n this determines x_1 uniquely and the system is always solvable; for even n the x_1 cancels and an alternating sum condition obstructs solvability. This odd versus even dichotomy is the same fact as $\det(I + S) = 1 - (-1)^n$

Example 1.11.4 (Putnam 2011 A4)**.** There is an $n \times n$ integer matrix with every row dotted with itself even and every two distinct rows dotted together odd exactly when n is odd.

The hypotheses are parity statements about integers, so reduce the whole problem modulo 2 on the first line. Let $\overline{M}$ be the reduction. All n^2 hypotheses become the single equation $\overline{M}\,\overline{M}^{\mathsf{T}} = J - I$ over $\mathbb{F}_2$.

For n odd, take $M = J - I$ over $\mathbb{Z}$: each row dotted with itself is $n - 1$, even; two distinct rows give $n - 2$, odd. That is the construction.

For n even, derive a contradiction from rank. Over $\mathbb{F}_2$, $x^2 = x$, so the dot product of a row with itself equals the sum of its entries; the hypothesis says that sum is 0, for every row, which is exactly $\overline{M}\mathbf{1} = 0$. So $\overline{M}$ is singular and $\operatorname{rank}\overline{M} \leq n - 1$, whence $\operatorname{rank}(\overline{M}\,\overline{M}^{\mathsf{T}}) \leq n - 1$. But $(J - I)^2 = J^2 - 2J + I = nJ + I = I$ over $\mathbb{F}_2$ when n is even, so $J - I$ is invertible and has rank n. Contradiction.

The conceptual reason even n fails is worth naming: the diagonal hypothesis forces the symmetric bilinear form carried by the rows to be alternating over $\mathbb{F}_2$ (that is, $q(r) = r \cdot r = 0$ for every row), and an alternating form has even rank, while the off diagonal hypothesis demands full rank n. That is a two line explanation of an answer that otherwise looks like a coincidence, and it appeared once, so learn the reduction first and the alternating form fact second.

- **WHEN** a hypothesis is about parities of integers **DO** reduce the entire problem modulo 2 on the first line and do linear algebra over $\mathbb{F}_2$ **BECAUSE** reduction is a ring homomorphism, so it commutes with det, with matrix products and with powers; every parity statement becomes an equation over a field where you have rank, kernel and determinant
- **WHEN** you are over $\mathbb{F}_2$ and see a dot product of a vector with itself **DO** replace it by the sum of the entries **BECAUSE** $x^2 = x$ in $\mathbb{F}_2$, so $\sum_i x_i^2 = \sum_i x_i$; this single substitution is what turns a diagonal hypothesis into a kernel vector
- **WHEN** you are over $\mathbb{F}_2$ and see a linear combination of rows **DO** read it as a subset sum, and read a nontrivial dependency as two distinct subsets of rows with equal sums **BECAUSE** the only nonzero scalar is 1, so a linear combination is exactly a subset, and singularity is exactly a coincidence between two subset sums

Example 1.11.5 (Putnam 2008 B6)**.** The number of k-limited permutations of $\{1, \ldots, n\}$, that is, those with $|\sigma(i) - i| \leq k$ for all i, is odd if and only if $n \equiv 0$ or $1 \pmod{2k+1}$.

Let $M_{n,k}$ be the 0/1 matrix with (i, j) entry 1 exactly when $|i - j| \leq k$. The permanent $\operatorname{per}(M) = \sum_\sigma \prod_i M_{i\sigma(i)}$ counts the permutations with all $M_{i\sigma(i)} = 1$, which here is exactly the k-limited ones. In characteristic 2 the sign disappears, so per $=$ det. Hence the parity of the count is $\det M_{n,k}$ over $\mathbb{F}_2$, and a combinatorial enumeration has become a singularity question.

Now recognise the shape: $M_{n,k}$ depends only on $i - j$, so it is a banded Toeplitz matrix of width $2k + 1$ with symbol $\sum_{|d| \leq k} x^d$. A kernel vector satisfies $\sum_{j=i-k}^{i+k} a_j = 0$ for every i (with $a_j = 0$ outside $[1, n]$). Difference the equations for i and $i + 1$: everything cancels except $a_{i-k} + a_{i+k+1} = 0$, that

is $a_{i+2k+1} = a_i$ over $\mathbb{F}_2$. Equivalently, multiply the symbol by $1 + x$: over $\mathbb{F}_2$, $(1 + x)(1 + x + \cdots + x^{2k}) = 1 + x^{2k+1}$, which is the characteristic polynomial of the period $2k + 1$ recurrence. Combining that periodicity with the boundary equations pins down exactly when a nonzero kernel vector exists, and the answer is that $M_{n,k}$ is nonsingular over $\mathbb{F}_2$ precisely when $n \equiv 0$ or $1 \pmod{2k+1}$.

- **WHEN** you must find the parity of a count of permutations with a $0/1$ constraint pattern **DO** write the count as a permanent and use per $=$ det in characteristic 2 **BECAUSE** the sign of a permutation is ± 1 and $-1 = 1$ over $\mathbb{F}_2$, so the two expansions coincide; a counting problem becomes a determinant problem
- **WHEN** a banded Toeplitz matrix appears **DO** multiply the symbol by whatever polynomial clears the interior, and read off the resulting recurrence **BECAUSE** the kernel equations are a constant coefficient linear recurrence whose characteristic polynomial is the symbol; clearing the interior is exactly multiplying the symbol, and the period appears as a root of unity in the cleared polynomial

Example 1.11.6 (Putnam 2021 B5)**.** Call an integer matrix very odd if every principal submatrix $(a_{ij})_{i,j \in S}$, over all nonempty S, has odd determinant. If A is very odd then so is A^k for every $k \geq 1$.

Reduce modulo 2 at once, which is legitimate because the determinant is a polynomial with integer coefficients in the entries, so $\varphi(\det M) = \det(\varphi M)$ for any ring homomorphism φ, and reduction is multiplicative so the reduction of A^k is the k-th power of the reduction. "Very odd" is exactly "every principal submatrix is invertible over $\mathbb{F}_2$".

The natural first attempt is to propagate the hypothesis through the product using Cauchy and Binet: for equal size subsets, $\det\big((MN)_{S,T}\big) = \sum_R \det(M_{S,R}) \det(N_{R,T})$. That expansion involves *non-principal* minors of the factors, about which the hypothesis says nothing, so the attempt fails, and recognising the mismatch early is what saves the time budget. Classify instead.

First, the 1×1 principal minors force every diagonal entry to be 1. Now consider the digraph on $\{1, \ldots, n\}$ with an edge $i \to j$ whenever $i \neq j$ and $a_{ij} = 1$. Suppose it has a directed cycle, and take one of minimal length, with vertex set S. Minimality means no chords, so within the principal submatrix on S every row has exactly two entries equal to 1: the diagonal entry and the single forward cycle edge. Every row therefore sums to 0 over $\mathbb{F}_2$, so that submatrix kills the all ones vector and is singular, contradicting the hypothesis. Hence the digraph is acyclic, so its vertices admit a topological order, which says precisely that some permutation matrix P makes $P^{-1}AP$ upper triangular with all diagonal entries 1, that is, unipotent.

Conversely, unipotent matrices are closed under multiplication, and every principal submatrix of a unitriangular matrix is unitriangular, hence nonsingular. So $P^{-1}A^kP = (P^{-1}AP)^k$ is unipotent and A^k is very odd.

The symmetry group of the hypothesis is exactly the permutation conjugations: $P^{-1}AP$ has entries $a_{\sigma(i)\sigma(j)}$, and the principal minor on S becomes the one on $\sigma(S)$, since permuting rows and columns alike costs $\operatorname{sgn}(\sigma)^2 = 1$. That is why the normal form may use relabelling and nothing more. The shortcut of proving only unipotence fails: over $\mathbb{F}_2$, $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right) = I + N$ with $N^2 = 0$ is unipotent yet has $(1,1)$ entry 0, so it is not very odd. Unipotence survives every conjugation; very odd survives only permutation conjugation, so no full conjugation invariant can capture it.

- **WHEN** a hypothesis controls all principal minors **DO** do not propagate through Cauchy and Binet; classify the matrices satisfying it **BECAUSE** Cauchy and Binet expands a minor of a product in terms of non-principal minors of the factors, which the hypothesis does not control, so the expansion loses all information
- **WHEN** a matrix hypothesis is invariant under relabelling the index set **DO** look for a normal form under permutation conjugation, and use the support digraph **BECAUSE** permutation conjugation is exactly relabelling, an acyclic support digraph is exactly a topological order, and a topological order is exactly triangularity

Example 1.11.7 (Putnam 2014 B5)**.** Two players alternately choose elements of $\mathrm{GL}_n(\mathbb{F}_p)$, never repeating and always commuting with everything already chosen; a player unable to move loses. The first player wins when $p = 2$ and the second wins when $p > 2$.

At any point the chosen set S lies inside its own centraliser $Z(S) = \{g : gs = sg \ \forall s \in S\}$, a subgroup, and the game ends exactly when $S = Z(S)$. So the game length is $|Z(S)|$ for the final S, a subgroup order, and the first player wins exactly when it is odd.

For $p > 2$, the scalar $-I$ is central and distinct from I, so $\{I, -I\} \subseteq Z(S)$ always, 2 divides $|Z(S)|$, the length is even and the second player wins.

For $p = 2$, open with a Singer cycle: identify $\mathbb{F}_2^n$ with $\mathbb{F}_{2^n}$ and let g be multiplication by a generator α of $\mathbb{F}_{2^n}^\times$. Any $\mathbb{F}_2$ linear h commuting with g commutes with multiplication by every power of α, hence with every element of $\mathbb{F}_{2^n}$, so $h(cv) = ch(v)$ and h is multiplication by $h(1)$. Thus $Z(\{g\}) \cong \mathbb{F}_{2^n}^\times$ is abelian of odd order $2^n - 1$, the game lasts exactly that many moves, and the first player wins. The general fact behind the computation: if $\mu_A = \chi_A$ then the centraliser of A in the matrix algebra is exactly $F[A]$. As a sanity check, $|\mathrm{GL}_n(\mathbb{F}_p)| = \prod_{i=0}^{n-1}(p^n - p^i)$ is even except in trivial cases, so the group order settles nothing and only centralisers can.

- **WHEN** a combinatorial game is played on a group with a commuting constraint **DO** track the centraliser $Z(S)$, note the game ends when $S = Z(S)$, and decide the winner by the parity of a subgroup order **BECAUSE** the constraint forces $S \subseteq Z(S)$ and each move shrinks $Z(S)$, so the game length is a subgroup order and the whole question is a parity question
- **WHEN** you need an element of $\mathrm{GL}_n(\mathbb{F}_q)$ with a small centraliser **DO** take a Singer cycle: identify $\mathbb{F}_q^n$ with $\mathbb{F}_{q^n}$ and multiply by a generator **BECAUSE** its centraliser is $\mathbb{F}_{q^n}^\times$, of order $q^n - 1$, because commuting with multiplication by a generator forces $\mathbb{F}_{q^n}$ linearity

1.12 Commutants and simultaneous diagonalisation

Definition 1.12.1. The commutant (centraliser) of A is $C(A) = \{B : AB = BA\}$, a subalgebra containing $F[A]$, the algebra of polynomials in A. A vector v is cyclic for a $k \times k$ matrix A if $v, Av, \ldots, A^{k-1}v$ is a basis; A is non-derogatory if it has a cyclic vector, equivalently if $\mu_A = \chi_A$, and then $C(A) = F[A]$ exactly, of dimension k and commutative. Derogatory matrices have larger, non-commutative commutants.

Proposition 1.12.2. *If $AB = BA$ then B preserves every eigenspace of A, since $Av = \lambda v$ gives $A(Bv) = B(Av) = \lambda Bv$. Consequently, if some eigenspace of A is one dimensional and spanned by v, then v is an eigenvector of every B commuting with A. A commuting family of diagonalisable operators on a finite dimensional complex space is simultaneously diagonalisable, and conversely matrices diagonal in a common basis commute.*

Proof of simultaneous diagonalisation. Decompose the space into eigenspaces of one member A. Each is preserved by every other member, and the restriction of a diagonalisable operator to an invariant subspace is diagonalisable (its minimal polynomial divides the original, which is a product of distinct linear factors). Induct on the dimension. □

Proposition 1.12.3 (Dimension two is rigid)**.** *If A is a 2×2 matrix over a field and A is not a scalar multiple of I, then $C(A) = \{pI + qA\}$.*

Proof. Some $v \neq 0$ is not an eigenvector of A, or else every vector is an eigenvector and A is scalar. Then v, Av is a basis. Given B with $AB = BA$, write $Bv = pv + qAv$. Then $B(Av) = A(Bv) = pAv + qA^2v = (pI + qA)(Av)$, so B agrees with $pI + qA$ on a basis, hence everywhere. □

The consequence is the useful one: in dimension 2, any two matrices commuting with a common non-scalar matrix commute with each other, because they both lie in the commutative algebra

$\{pI + qA\}$. So the commutation graph on non-scalar 2×2 matrices is a disjoint union of cliques. That rigidity fails in dimension 3, and the reason is exactly derogacy: a rank one idempotent in dimension at least 3 has minimal polynomial $x^2 - x$ of degree 2 below the size, so it is derogatory and its commutant is large and non-commutative.

Example 1.12.4 (Putnam 2025 A4)**.** The minimal k for which there exist $k \times k$ real matrices $A_1, \ldots, A_{2025}$ with $A_iA_j = A_jA_i$ if and only if $|i - j| \in \{0, 1, 2024\}$ is $k = 3$.

The condition says the commutation graph is exactly the 2025-cycle. For the lower bound, note first that no A_i can be scalar, since a scalar matrix commutes with everything and would be joined to all 2024 others. Now suppose $k = 2$. The cycle contains the path A_1, A_2, A_3, so A_1 and A_3 both commute with the non-scalar A_2; by Proposition 1.12.3 both lie in $\{pI + qA_2\}$, which is commutative, so $A_1A_3 = A_3A_1$. But $|1 - 3| = 2 \notin \{0, 1, 2024\}$, a contradiction. The case $k = 1$ is worse, since all 1×1 matrices commute. So $k \geq 3$.

For the construction, take unit vectors $v_1, \ldots, v_{2025}$ in $\mathbb{R}^3$ in general position, no three coplanar, indices read cyclically; the moment curve supplies them with a Vandermonde certificate. Let A_i be the unique linear map with

$$A_i(v_i) = v_i, \qquad A_i(v_{i-1}) = A_i(v_{i+1}) = 0,$$

well defined because v_{i-1}, v_i, v_{i+1} is a basis. Then A_i is a rank one idempotent, so $A_i = v_i\varphi_i$ with $\varphi_i(v_i) = 1$ and $\ker \varphi_i = \operatorname{span}(v_{i-1}, v_{i+1})$.

Now use the contraction rule. For $i \neq j$, v_i and v_j are not parallel, and $A_iA_j = \varphi_i(v_j)\, v_i\varphi_j$ while $A_jA_i = \varphi_j(v_i)\, v_j\varphi_i$. If both scalars vanish, both products are zero and the matrices commute. If not both vanish, the two products are matrices with different column spaces ($\mathbb{R}v_i$ against $\mathbb{R}v_j$), at least one of them nonzero, so they cannot be equal. Hence A_i and A_j commute exactly when $\varphi_i(v_j) = 0$ and $\varphi_j(v_i) = 0$, that is, exactly when $v_j \in \operatorname{span}(v_{i-1}, v_{i+1})$ and $v_i \in \operatorname{span}(v_{j-1}, v_{j+1})$. Since no three of the v's are coplanar, the first condition alone forces $j \in \{i - 1, i + 1\}$, and both conditions hold when $j = i \pm 1$. That is precisely the cycle, so $k = 3$ works.

- **WHEN** a problem constrains which pairs in a family commute **DO** check dimension 2 first, where the commutant of any non-scalar matrix is $\{pI + qA\}$ and commuting is therefore transitive **BECAUSE** transitivity turns the commutation graph into a disjoint union of cliques, which kills every required pattern that is not a union of cliques, such as a long cycle
- **WHEN** you need a family with prescribed non-commuting behaviour **DO** build it from rank one idempotents $v\varphi$ and translate commuting into the two incidence conditions $\varphi(w) = 0$ and $\psi(v) = 0$ **BECAUSE** rank one matrices contract in the middle, so both products are explicit; two rank one matrices with non-parallel columns commute only when both products vanish, which is a pure incidence condition on vectors and hyperplanes
- **WHEN** two matrices commute and one has a one dimensional eigenspace **DO** declare its spanning vector an eigenvector of the other **BECAUSE** a commuting operator preserves eigenspaces, and preserving a line means acting on it by a scalar

Example 1.12.5 (Putnam 2011 A6)**.** For a finite abelian group G and $f : G \to \mathbb{C}$, the G-circulant $M_{x,y} = f(x^{-1}y)$ is diagonalised by the characters: for a character χ,

$$(M\chi)(x) = \sum_y f(x^{-1}y)\chi(y) = \sum_g f(g)\chi(xg) = \chi(x)\sum_g f(g)\chi(g),$$

so the vector $(\chi(x))_{x \in G}$ is an eigenvector with eigenvalue $\widehat{f}(\chi)$.

The structural reason is simultaneous diagonalisation: the translation operators of an abelian group commute and are diagonalisable, so they have a common eigenbasis, and the characters are that basis. Ordinary circulant diagonalisation by the discrete Fourier transform is the case $G = \mathbb{Z}/n\mathbb{Z}$.

- **WHEN** a matrix is indexed by a group and its entry depends only on $x^{-1}y$ **DO** diagonalise it in the character basis and read the eigenvalues as Fourier coefficients **BECAUSE** the matrix is a combination of translation operators, translations of an abelian group commute, and a commuting diagonalisable family has a common eigenbasis, which for a group is the characters

1.13 The standard Putnam moves, collected

Everything above is theory. This section is the Circuit Bank entry: the moves that recur across fields, stripped of their problems.

Rank one perturbation. Any hypothesis of the form "a matrix differs from a nice one by something of rank one" is a hypothesis about a single scalar, via $\det(A + vw^{\mathsf{T}}) = \det A\,(1 + w^{\mathsf{T}}A^{-1}v)$. Name that scalar and the problem usually collapses (Putnam 2019 B3, Putnam 2011 A4).

Outer products. $A_{ij} = u_i v_j$ is the same statement as $\operatorname{rank} A \leq 1$. Going left to right converts a rank hypothesis into algebra; going right to left converts an entry formula into a rank bound (Putnam 2009 A3, Putnam 2014 B3, Putnam 2025 A4).

Trace moves. $\operatorname{tr}(MN) = \operatorname{tr}(NM)$ is the only basis independent statement about the diagonal, and $\operatorname{tr}(M^{\mathsf{T}}M)$ counts the same sum of squares by rows and by columns (Putnam 2008 B3). Beware the sum against product distinction of Proposition 1.7.2.

The polynomial identity principle. If you want to divide by A but A might be singular, prove the identity for $A + tI$ with t outside a finite bad set and then compare polynomials in t. Both sides of any determinant identity are polynomials in the entries and in t, and two polynomials agreeing at infinitely many points are equal. This removes invertibility hypotheses for free and it is why Cayley and Hamilton, the adjugate identity and Sylvester's identity hold over commutative rings and not merely over fields (Putnam 2015 A6).

Reduction modulo a prime. det is a polynomial with integer coefficients in the entries, so it commutes with every ring homomorphism. Any statement about parities, or about divisibility by p, should be reduced on line one (Putnam 2011 A4, Putnam 2021 B5, Putnam 2008 B6, Putnam 2012 A5).

Tensor and Kronecker products. The Kronecker product $A \otimes B$ of an $n \times n$ and an $m \times m$ matrix acts on $\mathbb{R}^{nm}$ with $(A \otimes B)(C \otimes D) = AC \otimes BD$, eigenvalues the pairwise products $\lambda_i \mu_j$, $\operatorname{tr}(A \otimes B) = \operatorname{tr} A \operatorname{tr} B$, and $\det(A \otimes B) = (\det A)^m (\det B)^n$. Use it whenever a condition is a product of n separate conditions, because a product of inner products is one inner product on the tensor power (Putnam 2014 A6).

The intertwining reading. $AM = MB$ is a rewriting rule, not an equation. M carries generalised eigenspaces of B to those of A for the same eigenvalue; disjoint spectra force $M = 0$; and the set of solutions is a subspace, so never perturb towards invertibility (Putnam 2015 A6).

Fixed supply, growing size. If a matrix family of growing size is built from a fixed number r of underlying vectors or from entries that are a fixed degree d polynomial in i and j, the rank is bounded by r or by $d + 1$, and singularity for large n is automatic (Putnam 2009 A3).

- **WHEN** a matrix identity requires an inverse that might not exist **DO** prove it for $A + tI$ and compare polynomials in t **BECAUSE** a determinant identity is a polynomial identity in the entries, and a polynomial vanishing at infinitely many values of a parameter vanishes identically, so the invertible case implies the general case

- **WHEN** a condition splits as a product of n independent conditions **DO** move to the tensor power and turn the product into a single inner product **BECAUSE** $\langle \bigotimes w_t, \bigotimes y_t \rangle = \prod_t \langle w_t, y_t \rangle$ is exactly that conversion, and the dimension n^n is then the answer to the counting question

1.14 Recursion, iteration and self-similarity: the standing order

Linear algebra is where recursion is most often hidden, because a matrix is a machine for iterating. Collect these as one family. In every case the recognition is the same: the n-th object contains the $(n-1)$-th, or the k-th step is a fixed operator applied to the $(k-1)$-th.

Family 1: powers of a fixed matrix. By Cayley and Hamilton, the sequence $X_m = A^m$ satisfies the order n linear recursion $X_{m+1} = -a_{n-1}X_m - \cdots - a_0 X_{m+1-n}$ where $\chi_A(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$, valid for every $m \geq n-1$. Any monic annihilating polynomial of degree d gives the same recursion of order d, so the minimal polynomial gives the shortest one. Writing $A^k = \sum_{j<n} c_{k,j}A^j$ turns matrix powers into n scalar sequences obeying the same recursion (Putnam 2015 B3, Putnam 2010 B6). This is the single most useful instance: whenever a problem mentions A^k for unbounded k, the answer lives in an n dimensional space and satisfies an order n recursion.

Family 2: nested determinants. A family of matrices in which the $(n-1)$-th sits inside the n-th, compatibly with whatever elimination you are using, satisfies a first order recursion in the determinant. The three shapes are $D_n = c_n D_{n-1}$ (Putnam 2014 A2, with $D_n = (c_n - c_{n-1})D_{n-1}$, telescoping to $c_1\prod(c_k - c_{k-1})$), $D_n = -D_{n-1}^2$ (Putnam 2018 A2, from a block reduction with a block swap sign), and $h_n(\alpha) = \lambda^{n-1}h_{n-1}(\alpha')$ (Putnam 2024 A6, the Hankel reduction, closed form $\lambda^{n(n-1)/2}$ when the continued fraction is periodic). The general method is always: do one elimination step with a unitriangular multiplier, expand along the column it created, and identify the lower right block as the same object one size smaller.

Family 3: affine iteration. $x_{k+1} = Mx_k + v$ has closed form $x_k = M^k x_0 + \left(I + M + \cdots + M^{k-1}\right)v$. If $I - M$ is invertible this telescopes to $(I-M)^{-1}(I-M^k)v$. If $I - M$ is nilpotent write $M = I + N$ and expand binomially; $N^n = 0$ truncates the sum to n terms (Putnam 2012 A5). Recognise the family whenever a composite of an affine map is iterated, including in dynamical, probabilistic and group theoretic disguises.

Family 4: rank one accumulation. $M_g = M_{g-1} + v_g v_g^*$ with $M_0 = 0$ solves to $M_g = A^*A$ with A the matrix of the v_j. Never iterate the determinant lemma; write the Gram closed form, and if you need the determinant in the non-square case use Cauchy and Binet as a sum of squared minors (Putnam 2011 B4).

Family 5: cyclic systems. $x_{i+1} = t_i - x_i$ around a cycle of length n unwinds to $x_1 =$ (alternating sum of the t_i) $+ (-1)^n x_1$: uniquely solvable for odd n, obstructed by an alternating sum condition for even n. Matrix form: $\det(I+S) = 1 - (-1)^n$ for the cyclic shift S. This odd against even dichotomy is one of the most reliable patterns in the whole archive (Putnam 2017 A6).

Family 6: banded Toeplitz kernels. $Ma = 0$ for a banded Toeplitz M is a constant coefficient linear recurrence in the coordinates of a whose characteristic polynomial is the symbol, subject to boundary conditions from the finite window. Multiply the symbol by whatever clears the interior to expose the period (Putnam 2008 B6).

Family 7: Markov and drift recursions. A law evolving as $v_k = Av_{k-1}$ has $v_k = A^k v_0$, and a linear statistic $w^{\mathsf{T}}v_k$ evolves by the left eigenvectors of A: if $w^{\mathsf{T}}A = \lambda w^{\mathsf{T}}$ then $w^{\mathsf{T}}v_k = \lambda^k w^{\mathsf{T}}v_0$. Writing your statistic as a combination of left eigenvectors converts a matrix iteration into scalar geometric sequences (Putnam 2024 B4, where $(1,\ldots,1)$ and $(n-1, n-3, \ldots, 1-n)$ are left eigenvectors with eigenvalues 1 and $1 - 2/n$).

Family 8: telescoping determinants. If $M_n = T^{n-1}M_1$ for a fixed T then $\det M_n = (\det T)^{n-1}\det M_1$, so an induction "prove $\det M_n = 1$ for every n" reduces to the single check $\det T = 1$ (Putnam 2017 A2).

Family 9: similarity orbits. If A acts on a subspace as a similarity of ratio r, then $u_k = A^{k-1}u_1$ has $\|u_k\| = r^{k-1}\|u_1\|$. A finite orbit forces $r = 1$, and when $r = 1$ the orbit is finite exactly when the rotation part has angle a rational multiple of π (Putnam 2022 B2, where $L_u(x) = u \times x$ acts on $u^{\perp}$ as a rotation by $\pi/2$ scaled by $\|u\|$, giving period 4 when $\|u\| = 1$).

Family 10: dimension counts by Pascal. The number of monomials of degree at most d in m variables satisfies $N_m(d) = N_m(d-1) + N_{m-1}(d)$ with $N_1(d) = d+1$, solved by $N_m(d) = \binom{d+m}{m}$;

the same recursion counts stars and bars arrangements and multisets (Putnam 2009 B4).

Family 11: three term recurrences from Hankel elimination. Gaussian elimination on a Hankel matrix is Gram-Schmidt on $1, x, x^2, \ldots$, which produces monic orthogonal polynomials satisfying $B_{k+1} = (x - b_k)B_k - \lambda_k B_{k-1}$, and the pivots are $L(B_k^2) = \lambda_1 \cdots \lambda_k$. Whenever a determinant turns out to be a product of continued fraction data, this is why (Putnam 2024 A6).

- **WHEN** a matrix, determinant or vector depends on a size parameter n and the $(n-1)$ case visibly sits inside the n case **DO** do one elimination or block step and read off a first order recursion, then telescope **BECAUSE** self-similar families do not have closed forms that you can guess; they have recursions that you can prove, and a first order recursion telescopes into a product
- **WHEN** an iteration is $x_{k+1} = \Phi(x_k)$ with Φ linear or affine **DO** write the closed form immediately, separating the linear part from the accumulated constant **BECAUSE** iterating a linear map is taking powers, and powers are governed by the characteristic polynomial, so the whole orbit is determined by n numbers
- **WHEN** you must track a scalar statistic of an iterated linear process **DO** express the statistic as a left eigenvector, or a combination of left eigenvectors, of the transition matrix **BECAUSE** a left eigenvector turns the matrix iteration into a scalar geometric sequence, which is the shortest possible description of the evolution

1.15 The exotica, and what to learn last

Everything in this section appeared once, or only inside an alternative solution, in the window 2006 to 2025. Learn the sections above cold first.

The Pfaffian appeared once, in Putnam 2016 B4, and only as a prediction device: $\det M = \mathrm{Pf}(M)^2$ for skew-symmetric M of even size told you the answer would be non-negative and indexed by perfect matchings before you computed. The hafnian generalisation $\mathrm{E}[\det M] = \sum_{\text{matchings}} \prod_\ell s_{a_\ell b_\ell}$ is what the solution actually proves.

The Cauchy determinant $\det\big(1/(x_i + y_j)\big)$ did not appear at all. Learn it last, purely so that the shape does not stop you.

The Hankel determinant of factorials, $\det\big((i+j+n)!\big)_{i,j=0}^{m} = \prod_{k=0}^{m} k!\,(k+n)!$, appeared only inside Elkies' alternative solution to Putnam 2011 B6, where a Hankel matrix of factorials had to be shown nonsingular modulo p.

The 4×4 determinant equation for the circle through three points, in the variables $x^2 + y^2, x, y, 1$, is the heavier alternative route in Putnam 2008 B1; it is worth knowing only because other concyclicity problems sometimes want the explicit circle equation with its rational coefficients.

Lattice index appeared once, in Putnam 2012 A4: if $u, w \in \mathbb{Z}^2$ are independent then the lattice $\mathbb{Z}u + \mathbb{Z}w$ has index $|\det(u, w)|$ in $\mathbb{Z}^2$, indices multiply along a chain of sublattices, and consequently any two independent vectors of an index q sublattice span a parallelogram of area a positive multiple of q. That quantisation is the whole obstruction in the problem.

Two more one-offs, both from Putnam 2019 B3 and both alternative routes rather than the main one: the Cayley transform, which says $(I - A)(I + A)^{-1}$ is orthogonal for skew A and gives the rational parametrisation of the orthogonal group; and the reflection factorisation, which says every orthogonal matrix is a product of at most n Householder reflections, with exactly $n - \dim\ker(Q - I)$ required. The second connects the problem to the QR algorithm and shows that the statement is really about generation of the orthogonal group by reflections.

Finally, two remarks that are not exotica but are easy to forget. The condition $bc - ad = 1$ on a pair of linear polynomials $p = ax + b$, $q = cx + d$ says that the matrix of values $\left(\begin{smallmatrix} p(0) & p(1) \\ q(0) & q(1) \end{smallmatrix}\right)$ has determinant 1, which over $\mathbb{Z}$ is membership of $\mathrm{SL}_2(\mathbb{Z})$, the group behind continued fractions and Farey neighbours; that is the one sentence summary of the answer to Putnam 2010 B4, whose only

other linear algebra content is that $D(p, q - tp) = D(p, q)$ for constant t because the correction terms cancel, while for a nonconstant multiplier $r(x)$ the correction $p(x)p(x+1)\big(r(x+1) - r(x)\big)$ does not vanish, so only constant multipliers are legal row operations there. And a line in the plane is $\{x : v \cdot x = c\}$ for a normal v; two lines are parallel exactly when their normals are dependent, and non-parallel exactly when the pair of equations has a unique simultaneous solution, which is the coordinate route through Putnam 2017 B1.

1.16 Closing protocol

When a Putnam problem contains a matrix, a determinant, a family of vectors, or a count over a finite alphabet, run this in order.

1. Name the vector space and its dimension. If the objects can be added and scaled, they live in a vector space, and the dimension is the ceiling on how many independent ones exist.
2. Ask what the rank is. Determinant, singularity, dependency, eigenvalue and invertibility questions are all rank questions in disguise, and rank is the invariant that survives every reformulation.
3. Look for a factorisation. An entry that is a sum over an auxiliary index is a matrix product; an entry that is a product of a function of i and a function of j is an outer product; an entry that is a dot product is a Gram matrix.
4. Check the field. If the hypothesis is about parity or divisibility, reduce. If it is about reality or positivity, you are over $\mathbb{R}$ and the spectral theorem and positive definiteness are available. If a parameter appears, consider running the polynomial identity principle over it.
5. Look for the recursion. If the problem has a size parameter, the $(n-1)$ case is almost certainly inside the n case; do one elimination step and read off the recursion instead of hunting for the closed form.
6. Only then compute. If you find yourself expanding an $n \times n$ determinant by cofactors, you have missed a structural reroute, and the ink you are about to spend is the ink that invites the error.

Chapter 2

Groups, rings and fields

2.1 What this chapter is for

A-level gave you almost nothing here. The FP2 groups option gave you a Cayley table, the four axioms, the order of an element, the word "subgroup", a statement of Lagrange with no proof worth the name, and a handful of small groups to recognise up to isomorphism. That is a vocabulary list, not a theory. Everything that makes group theory a *counting device* was left out, and counting is the only reason the Putnam cares about groups at all.

That is the gap, and it is a wide one. On the Putnam, an abstract algebra problem almost never says "here is a group, describe it". It says something else, and the group is the reroute. Counting the elements of order p in a finite group is counting orbits of a cyclic shift on p-tuples (Putnam 2007 A5). Counting sequences modulo 5 is the same shift acting on periodic sequences (Putnam 2022 A3). A two player game on invertible matrices is the parity of the order of a self-centralising abelian subgroup (Putnam 2014 B5). Permutations with $\pi(\pi(k)) \equiv mk$ are a question about cycle type (Putnam 2023 B5). Perfect codes in the lattice are a homomorphism $\mathbb{Z}^n \to \mathbb{Z}/(2n+1)$ (Putnam 2019 B6). A question about the floor function is integrality over a polynomial ring (Putnam 2007 B5). In every case the nominal solution is a case analysis or a computation nobody finishes in the arena, and the ideal solution is three lines of structure.

So the syllabus of this chapter is: groups to the depth of actions and counting, rings to the depth of ideals, quotients and integrality, fields to the depth of $\mathbb{F}_q$ and Frobenius, and modules only as far as the structure theorem for finitely generated abelian groups. No Galois theory, no representation theory beyond characters of abelian groups, no classification of anything. Everything here is evidenced by an actual Putnam problem, and where a topic appeared once I will say so and tell you to learn it last.

By the end you will read these cues and the circuit will fire without deliberation. "Finite group, prime p, count something modulo p": build a set, let a group of order p^k act on it, count fixed points. "The order of x divides both a and b": Lagrange, and the divisibility is the whole content. "Prove this holds in every group": exhibit a homomorphism, or work in the presented group and cite von Dyck. "A nuisance subgroup keeps appearing": quotient by it. "Finite and closed under an operation": it is a subgroup, because translation is a bijection. "Characteristic p": the p-th power map is additive, so raise everything to the p-th power. "Sum over all elements of a group of a multiplicative function": orthogonality, so the sum is 0 unless the function is trivial.

One standing warning. Group theory rewards you for going back to the definition harder than any other field on this syllabus. Every one of the theorems below is one or two lines from the axioms. If you find yourself memorising a statement whose proof you cannot reconstruct in the arena, that is a knowledge gap, and I will mark the few genuine ones explicitly.

2.2 Groups: the axioms and what each one buys

Definition 2.2.1. A *group* is a set G with a map $G \times G \to G$, written $(a, b) \mapsto ab$, such that: the operation is associative, $(ab)c = a(bc)$; there is an element e with $ea = ae = a$ for all a; and every a has an element a^{-1} with $aa^{-1} = a^{-1}a = e$. The group is *abelian* if $ab = ba$ for all a, b. The *order* $|G|$ is the number of elements.

Do not read those axioms as a list to satisfy. Read what each one buys. Associativity buys you unambiguous long products, so a word in the generators is a well defined element. The identity buys you a basepoint. Inverses buy you the single most used fact in the chapter, which is that translation is a bijection.

Proposition 2.2.2 (Translation is a bijection)**.** *For fixed $g \in G$, the maps $L_g(x) = gx$ and $R_g(x) = xg$ are bijections $G \to G$. Consequently, for any subset $H \subseteq G$ and any g, $|Hg| = |gH| = |H|$, and cancellation holds: $ac = bc$ forces $a = b$.*

Proof. $L_{g^{-1}}$ is a two sided inverse of L_g by associativity, and similarly on the right. A bijection restricted to H is injective, so $|Hg| = |H|$. Cancellation is injectivity of R_c. □

That is the engine. Every counting argument in this chapter is a translation being a bijection, dressed up. It is also what fails in a monoid, which is exactly why Putnam 2008 A6 is a group problem and not a semigroup problem.

Proposition 2.2.3 (One sided inverses and the rotation principle)**.** *In a group, if $QP = e$ then $P = Q^{-1}$ and hence $PQ = e$. In particular, if a word $W = QP$ equals the identity, then so does every cyclic rotation of W, since $PQ = Q^{-1}WQ$.*

Proof. From $QP = e$, multiply on the left by Q^{-1}: $P = Q^{-1}$, so $PQ = Q^{-1}Q = e$. For the rotation, $Q^{-1}WQ = Q^{-1}(QP)Q = PQ$, and conjugating the identity gives the identity. □

This is a tiny statement with an enormous reach: any relation you are handed may be rotated for free, so a relator is really a *cyclic* word. That is the closing move of Putnam 2018 A4.

Definition 2.2.4. The *order* of $x \in G$, written $\operatorname{ord}(x)$, is the least $t \geq 1$ with $x^t = e$, when such a t exists.

Proposition 2.2.5. *If $\operatorname{ord}(x) = d$ then $x^s = e$ if and only if $d \mid s$. Moreover $\langle x^k \rangle = \langle x^{\gcd(k,d)} \rangle$, so x^k generates $\langle x \rangle$ exactly when $\gcd(k, d) = 1$.*

Proof. Write $s = qd + r$ with $0 \leq r < d$; then $x^s = (x^d)^q x^r = x^r$, and minimality of d forces $r = 0$. For the second claim, let $g = \gcd(k, d)$ and write $g = uk + vd$; then $x^g = (x^k)^u (x^d)^v = (x^k)^u \in \langle x^k \rangle$, and conversely x^k is a power of x^g. □

The special case you will use most: $x \in \langle x^2 \rangle$ if and only if $\operatorname{ord}(x)$ is odd, with the explicit formula $x = (x^2)^{(d+1)/2}$. That single line is the hypothesis of Putnam 2016 A5 doing its work.

Proposition 2.2.6 (Finiteness upgrades closure to a subgroup)**.** *Let G be a group and $S \subseteq G$ finite and nonempty with $SS \subseteq S$. Then S is a subgroup. Equivalently, for finite S and fixed t, $St \subseteq S$ forces $St = S$.*

Proof. Fix $x \in S$. The powers $x, x^2, x^3, \ldots$ all lie in S, which is finite, so $x^i = x^j$ for some $i < j$; cancelling gives $x^{j-i} = e$, so $e \in S$ and $x^{-1} = x^{j-i-1} \in S$. Closure under inverses and the operation is exactly the subgroup test. For the second form, R_t is injective, so $|St| = |S|$, and $St \subseteq S$ between finite sets of equal size is equality. □

The infinite version is false: $S = \{1\}$ inside $\mathbb{Z}$ under addition is closed and is not a subgroup. Say "finite" out loud when you use this.

Example 2.2.7 (Putnam 2009 B1)**.** Show that every positive rational is a quotient of products of factorials of primes. Let S be the set of such quotients. It contains 1, is closed under multiplication and under inversion by construction, so it is a subgroup of $(\mathbb{Q}_{>0}, \times)$. Since the primes generate that group, it suffices to show every prime lies in S. Induct on p: $2 = 2!$, and for $p > 2$ write $p = p! \cdot ((p-1)!)^{-1}$, where $(p-1)!$ is a product of primes smaller than p, each in S by induction. The open ended construction has become a closure check, which is the entire reroute.

Example 2.2.8 (Putnam 2022 B2)**.** For which n is there $S \subset \mathbb{R}^3$ with $|S| = n$ and $S = \{v \times w : v, w \in S\}$? The answers are $n = 1$ with $S = \{0\}$ and $n = 7$ with $S = \{0, \pm i, \pm j, \pm k\}$. The structural reason behind the second is that the cross product is the imaginary part of quaternion multiplication, so the answer set is the image of the quaternion group, a known rigid object rather than a coincidence. The reusable half is the one dimensional model: a finite nonempty subset of $\mathbb{C}^\times$ closed under multiplication is a subgroup, hence a group μ_m of roots of unity, by the proposition above.

Example 2.2.9 (Putnam 2012 A2: how few axioms you need)**.** Let $*$ be commutative and associative on S with every equation $x * z = y$ solvable for z. Prove $a * c = b * c \Rightarrow a = b$. The hypothesis says exactly that every left translation L_x is surjective. Choose e with $c * e = c$. For arbitrary x, choose z with $c * z = x$; then $x * e = (c * z) * e = (c * e) * z = c * z = x$, so e is a global identity. Now choose d with $c * d = e$: then $a = a * e = a * (c * d) = (a * c) * d = (b * c) * d = b * e = b$. Notice what I did not do: I did not attempt to prove $(S, *)$ is a group first. I extracted only the identity and one inverse, because those are the only two things cancellation needs. Structurally, $(S, *)$ sits between a commutative semigroup and a group; the named object with unique solvability is a *quasigroup*, and an associative quasigroup with identity is a group.

- **WHEN** a hypothesis says "for all x, y there is z with $x * z = y$" **DO** rewrite it as "every translation L_x is surjective" and then solve for an identity and for inverses one at a time **BECAUSE** translation being a bijection is the only content the group axioms ever have, so this is the axiom in its usable form (Putnam 2012 A2).
- **WHEN** a set is finite and closed under an operation **DO** declare it a subgroup immediately **BECAUSE** finiteness turns the injective map R_t into a surjective one, so closure under multiplication drags inverses in with it (Putnam 2016 A5, 2022 B2).
- **WHEN** you are handed a relation $W = e$ in a group **DO** rotate it: every cyclic rotation of W also equals e **BECAUSE** a rotation is a conjugate of W, and conjugates of the identity are the identity (Putnam 2018 A4).
- **WHEN** a problem asks you to construct a set closed under several operations and containing given generators **DO** prove closure and then show the generators of the ambient object are in it, rather than building the target element by hand **BECAUSE** a subgroup containing a generating set is everything, so closure plus generation beats explicit construction (Putnam 2009 B1).
- **WHEN** you need x from x^2 **DO** check whether $\operatorname{ord}(x)$ is odd; if it is, $x = (x^2)^{(\operatorname{ord}(x)+1)/2}$ **BECAUSE** $\langle x^k \rangle = \langle x^{\gcd(k, \operatorname{ord} x)} \rangle$, so odd order is exactly the condition for squaring to lose nothing (Putnam 2016 A5).

2.3 Cosets, Lagrange, and the divisibility engine

Definition 2.3.1. For $H \leq G$ and $g \in G$, the sets gH are the *left cosets* and Hg the *right cosets* of H. The number of left cosets is the *index* $[G : H]$.

Theorem 2.3.2 (Lagrange)**.** *If G is finite and $H \leq G$ then $|G| = [G : H]\,|H|$. In particular $|H|$ divides $|G|$, and $\operatorname{ord}(x)$ divides $|G|$ for every x, and $x^{|G|} = e$.*

Proof. The relation $a \sim b$ when $a^{-1}b \in H$ is an equivalence relation whose classes are the left cosets, so the cosets partition G. Each coset gH has $|H|$ elements because L_g is a bijection. Count. For the element statement, apply this to $H = \langle x \rangle$, which has order $\mathrm{ord}(x)$. □

Understand what Lagrange is *for*. It is a machine that converts any statement of the form "here is a subgroup" into a divisibility, and divisibility is a rigid, arithmetic thing you can push around. Odd order can never become even. A group of prime order has no proper nontrivial subgroups, hence is cyclic and generated by any non-identity element. If $\gcd(|G|, m) = 1$ then the m-th power map is a bijection on G when G is abelian. Every one of those is Lagrange with one extra sentence.

Definition 2.3.3. For $S \subseteq G$, the *centraliser* is $Z(S) = \{h \in G : hg = gh \text{ for all } g \in S\}$ and the *centre* is $Z(G)$. Both are subgroups. Moreover $Z(S) = Z(\langle S \rangle)$, the map $S \mapsto Z(S)$ is inclusion reversing, and if the elements of S pairwise commute then $S \subseteq Z(S)$.

The centraliser is the single most useful subgroup on the Putnam, because "these elements commute" is a hypothesis that appears constantly and $Z(S)$ is its subgroup form.

Example 2.3.4 (Putnam 2014 B5, in full)**.** Patniss and Keeta alternately choose elements of $G = GL_n(\mathbb{F}_p)$, never repeating, each new element required to commute with all previous ones; a player who cannot move loses. Patniss moves first. Who wins?

Start from the definition of the terminal position. Let S be the set of chosen elements. By the rules the elements of S pairwise commute, so $S \subseteq Z(S)$, and the legal moves are exactly $Z(S) \setminus S$. Hence the game ends precisely when $S = Z(S)$, and at that moment S is a *subgroup*, because centralisers are subgroups. The length of the game is $|S|$, so Patniss wins if and only if the final $|S|$ is odd. By Lagrange, $|S|$ divides $|G|$. The whole problem is now a parity question, and each player is steering the parity of a divisor.

Keeta's plan when p is odd. The scalar matrix $-I$ is central and has order 2. If Patniss does not play it first, Keeta plays it on move two; it commutes with everything, so it is legal. Either way $-I$ lies in the final subgroup S, so $2 \mid |S|$ and the game has even length. Keeta wins for every odd p and every n.

Patniss's plan when $p = 2$. Here $-I = I$ and that plan is dead, so reverse the parity instead: Patniss should open with an element whose centraliser has *odd* order, because every later choice lies in $Z(g)$, so the final subgroup is a subgroup of $Z(g)$ and its order divides $|Z(g)|$. Identify $\mathbb{F}_2^n$ with the field $\mathbb{F}_{2^n}$ and let g be multiplication by a generator of the cyclic group $\mathbb{F}_{2^n}^\times$. Then $\mathbb{F}_2[g] = \mathbb{F}_{2^n}$, so V is a one dimensional vector space over $\mathbb{F}_2[g]$ and the centraliser of g in $\mathrm{End}(V)$ is $\mathbb{F}_{2^n}$ itself. Hence $Z(g) = \mathbb{F}_{2^n}^\times$ has order $2^n - 1$, odd. Patniss wins for $p = 2$.

The answer: Patniss wins exactly when $p = 2$; Keeta wins for every odd prime. Two knowledge gaps are load bearing and both live in the next sections: the existence of $\mathbb{F}_{2^n}$, and the cyclicity of its multiplicative group. Without those there is no construction at all.

- **WHEN** a hypothesis or a game produces a nested family of subgroups **DO** apply Lagrange to get a chain of divisibilities and look for one monotone invariant of that chain, usually a parity or a p-adic valuation **BECAUSE** orders can only divide downwards, so an invariant such as "odd" can never be undone by a later move (Putnam 2014 B5).
- **WHEN** a set of pairwise commuting elements appears **DO** replace it with its centraliser $Z(S)$, which is a subgroup containing it **BECAUSE** $S \subseteq Z(S)$ turns a set-theoretic condition into the subgroup equation $S = Z(S)$, which Lagrange can measure.
- **WHEN** you must force an element of order 2 into a subgroup **DO** look for a central element of order 2 first, since it is legal everywhere and instantly puts 2 into the order by Lagrange **BECAUSE** centrality removes all commuting constraints, so it is the cheapest possible move (Putnam 2014 B5).

- **WHEN** the problem says the group has odd order **DO** immediately conclude there are no elements of order 2, no index 2 subgroups, and squaring is a bijection **BECAUSE** all three are Lagrange plus the order of an element, and they are the only things odd order ever means (Putnam 2016 A5, 2019 B6).

2.4 Homomorphisms, normality and quotients

Definition 2.4.1. A *homomorphism* is a map $\varphi : G \to H$ with $\varphi(ab) = \varphi(a)\varphi(b)$. Its *kernel* is $\ker\varphi = \varphi^{-1}(e)$ and its *image* is $\operatorname{im}\varphi$. A subgroup $N \leq G$ is *normal*, written $N \trianglelefteq G$, if $gNg^{-1} = N$ for every g.

A homomorphism is a device for transporting structure. You almost never need to know what H is; you need to know that φ exists, because then every equation in G becomes an equation in H for free, and H is chosen to be something you can compute in.

Theorem 2.4.2 (First isomorphism theorem)**.** $\ker\varphi \trianglelefteq G$*, the cosets of* $\ker\varphi$ *are exactly the fibres of* φ*, the quotient set* $G/\ker\varphi$ *carries a group structure by* $(aN)(bN) = abN$*, and* $G/\ker\varphi \cong \operatorname{im}\varphi$*. In particular* $[G : \ker\varphi] = |\operatorname{im}\varphi|$ *when these are finite.*

Proof. Normality: $\varphi(gng^{-1}) = \varphi(g)e\varphi(g)^{-1} = e$. Fibres: $\varphi(a) = \varphi(b)$ if and only if $\varphi(a^{-1}b) = e$ if and only if $aN = bN$. Well definedness of the product on cosets is exactly normality: if N is normal then $aNbN = abN$ because $Nb = bN$. The map $aN \mapsto \varphi(a)$ is then a well defined bijective homomorphism onto the image. □

The second and third isomorphism theorems ($HN/N \cong H/(H \cap N)$ and $(G/N)/(K/N) \cong G/K$) and the correspondence theorem (subgroups of G/N are exactly subgroups of G containing N) are stated here and used without ceremony. Quotienting is a *noise removal* operation: if some subgroup keeps getting in the way, kill it and work downstairs.

Proposition 2.4.3 (Universal property of $\mathbb{Z}^n$)**.** $\mathbb{Z}^n$ *is free abelian of rank* n*: for any abelian group* A *and any* $c_1, \dots, c_n \in A$ *there is a unique homomorphism* $\varphi : \mathbb{Z}^n \to A$ *with* $\varphi(e_i) = c_i$*, namely* $\varphi(x) = \sum x_i c_i$*. Its kernel is a subgroup, its fibres are the cosets of that kernel, and if* A *is finite each fibre has density* $1/|\operatorname{im}\varphi|$ *in* $\mathbb{Z}^n$*.*

Example 2.4.4 (Putnam 2019 B6, in full)**.** Call $p, q \in \mathbb{Z}^n$ neighbours if they differ by 1 in exactly one coordinate. For which n is there $S \subseteq \mathbb{Z}^n$ such that no two elements of S are neighbours, and every $p \notin S$ has exactly one neighbour in S?

The nominal solution is to build S by hand and drown. The cue is that the condition is uniform in the lattice and translation invariant in flavour, so the object you want is not a set but a homomorphism: take $\varphi : \mathbb{Z}^n \to \mathbb{Z}/(2n+1)$ with $\varphi(e_i) = i$, and let $S = \ker\varphi = \varphi^{-1}(0)$. Now verify from the definition. The $2n$ neighbours of p have values $\varphi(p) \pm i$ for $i = 1, \dots, n$. In $\mathbb{Z}/(2n+1)$, which has odd order, the residues $\pm 1, \dots, \pm n$ are distinct and nonzero, so they run over every nonzero residue exactly once. Hence: if $\varphi(p) = 0$ then no neighbour has value 0, so S is an independent set; and if $\varphi(p) = c \neq 0$ then exactly one of the $2n$ values $c \pm i$ equals 0, so p has exactly one neighbour in S. The answer is every $n \geq 1$.

Two things did the work and both are reusable. The homomorphism collapsed a search over all subsets of the lattice to a choice of n residues. And the pairing $\{g, -g\}$ in a group of odd order matched the $2n$ neighbours to the $2n$ nonzero residues exactly, which is why the modulus had to be $2n + 1$.

- **WHEN** a problem asks for a set in an infinite structure defined by a local, translation invariant condition **DO** build it as the fibre of a homomorphism onto a small finite group **BECAUSE** fibres are cosets, so the local condition becomes one computation at a single point and is then true everywhere (Putnam 2019 B6).

- **WHEN** a subgroup N keeps obstructing an argument **DO** pass to G/N and prove the statement there, then lift **BECAUSE** quotienting deletes exactly the information in N and nothing else, so any statement not depending on N survives.
- **WHEN** you must prove an identity holds in every group **DO** find a homomorphism from a group where you can compute, or work in the presented group; never verify the identity element by element **BECAUSE** a homomorphism transports the whole equation at once.
- **WHEN** a map between finite sets is defined coordinatewise over a surjection with equal fibres **DO** prove bijectivity downstairs first, then fibrewise **BECAUSE** a fibration is a bijection exactly when the base map and every fibre map are bijections (Putnam 2008 B4).

2.5 Actions, orbits, and the counting engine

This is the highest value section in the chapter. Priority 5 items cluster here more densely than anywhere else in the corpus.

Definition 2.5.1. An *action* of a group G on a set X is a homomorphism $G \to \mathrm{Sym}(X)$, written $g \cdot x$. The *orbit* of x is $Gx = \{gx\}$; the *stabiliser* is $G_x = \{g : gx = x\}$, a subgroup. The fixed point set of the whole group is $X^G = \{x : gx = x \text{ for all } g\}$.

Theorem 2.5.2 (Orbit stabiliser)**.** *Orbits partition X, and $|Gx| = [G : G_x]$. Hence every orbit size divides $|G|$, and the orbits of size 1 are exactly the points of X^G.*

Proof. Being in the same orbit is an equivalence relation. The map $gG_x \mapsto gx$ is well defined and bijective from cosets to the orbit: $gx = hx$ if and only if $h^{-1}g \in G_x$. Then apply Lagrange. □

Theorem 2.5.3 (The fixed point lemma for p-groups)**.** *If $|G| = p^k$ with p prime and G acts on a finite set X, then $|X| \equiv |X^G| \pmod p$.*

Proof. Every orbit has size dividing p^k, so every orbit size lies in $\{1, p, p^2, \ldots, p^k\}$. Summing over orbits, all sizes except the singletons are divisible by p. □

Read that again, because it is the single most productive theorem in the chapter. It says: *to compute a cardinality modulo p, find a group of p-power order acting on your set and count only the fixed points.* The cue is a problem asking for a count modulo a prime, or asking you to prove a divisibility of a count. The reason it is natural is not clever: it is that orbit sizes of a p-group form the geometric family $1, p, \ldots, p^k$, and that family is 0 modulo p except for its first term.

Example 2.5.4 (Putnam 2007 A5, in full)**.** A finite group G has exactly n elements of order p, p prime. Prove $n = 0$ or $p \mid n + 1$.

First, why $n + 1$ and not n. The set $\{x : x^p = e\}$ consists of the identity together with the elements of order exactly p, because $x^p = e$ forces $\mathrm{ord}(x) \in \{1, p\}$. So the natural object has $n + 1$ elements, and $n + 1$ is what the theorem will produce. That is the first lesson: let the definition choose the object, and the awkward $+1$ stops being awkward.

Now build a set for the group to act on. Let $X = \{(x_1, \ldots, x_p) \in G^p : x_1 x_2 \cdots x_p = e\}$. Choosing $x_1, \ldots, x_{p-1}$ freely determines x_p, so $|X| = |G|^{p-1}$. The cyclic group $H = \mathbb{Z}/p$ acts on X by shifting, and the shift lands back in X because $x_2 \cdots x_p x_1 = x_1^{-1}(x_1 \cdots x_p)x_1 = e$. This is the whole reroute: a set whose size you know, with a p-group acting.

By the fixed point lemma, $|G|^{p-1} \equiv |X^H| \pmod p$. A tuple is fixed by every shift exactly when all entries are equal, so $X^H = \{(x, \ldots, x) : x^p = e\}$ has $n + 1$ elements. Thus $|G|^{p-1} \equiv n + 1 \pmod p$.

Split on whether p divides $|G|$. If it does, the left side is 0 and $p \mid n + 1$. If it does not, then by Lagrange no element has order p, so $n = 0$. Done.

Corollary 2.5.5 (Cauchy, by McKay's argument)**.** *If a prime p divides $|G|$ then G has an element of order p.*

Proof. In the display above, $p \mid |G|$ gives $p \mid n+1$, and $n+1 \geq 1$, so $n+1 \geq p$ and $n \geq p-1 \geq 1$. □

Bank that construction once and you own both results. For $p = 2$ there is a one line alternative worth knowing: pair each element with its inverse; the unpaired elements are the identity and the involutions, so an even order group has an odd number of involutions, in particular at least one.

Example 2.5.6 (Putnam 2022 A3, in full)**.** For a prime $p > 5$, let $f(p)$ count sequences $a_1, a_2, \ldots$ with $a_n \in \{1, \ldots, p-1\}$ and $a_n a_{n+2} \equiv 1 + a_{n+1} \pmod{p}$. Prove $f(p) \equiv 0 \text{ or } 2 \pmod{5}$.

The cue is the modulus 5 against a recurrence with no visible 5. So look for a group of order 5 acting. The recurrence $a_{n+2} = (1 + a_{n+1})/a_n$ over the field $\mathbb{F}_p$ is the Lyness map, and iterating it five times returns you to the start, so every admissible sequence is periodic with period dividing 5 and is determined by (a_1, a_2). Hence $\mathbb{Z}/5$ acts on the solution set S by shifting the index, and $|S| = f(p)$.

Orbits have size 1 or 5, so $f(p) \equiv |S^{\mathbb{Z}/5}| \pmod{5}$. A fixed sequence is constant, say $a_n = a$, and the recurrence becomes $a^2 = 1 + a$, that is $a^2 - a - 1 = 0$ in $\mathbb{F}_p$. Its discriminant is 5, which is nonzero in $\mathbb{F}_p$ since $p > 5$, so this quadratic has either 0 or 2 roots, and neither root is 0. Therefore $f(p) \equiv 0 \text{ or } 2 \pmod{5}$. The computation of $f(p)$ itself never happens, which is the point.

Proposition 2.5.7 (Free actions)**.** *If G acts on X with all stabilisers trivial, every orbit has size $|G|$, so $|G|$ divides $|X|$ and $|X| = |G| \cdot \#\{\text{orbits}\}$. Any G-invariant subset is a union of orbits, so counting it reduces to counting normalised representatives.*

Example 2.5.8 (Putnam 2017 B6)**.** Count ordered 64-tuples of distinct elements of $\{1, \ldots, 2017\}$ with $x_0 + x_1 + 2x_2 + \cdots + 63x_{63} \equiv 0 \pmod{2017}$. The coefficient sum is $1 + (1 + 2 + \cdots + 63) = 1 + 2016 = 2017 \equiv 0$, so the diagonal translation $x_i \mapsto x_i + t$ preserves the linear form, and it preserves distinctness. That is a free action of $\mathbb{F}_{2017}$ on the solution set, so the count is 2017 times the count of solutions with x_0 fixed. The observation to bank is not the answer but the reflex: when the coefficients of a linear condition sum to 0 modulo p, a diagonal translation acts, and the count factors.

Theorem 2.5.9 (Orbit counting, sometimes called Burnside's lemma)**.** *For a finite group G acting on a finite set X, the number of orbits is $\frac{1}{|G|} \sum_{g \in G} |\mathrm{Fix}(g)|$.*

Proof. Count the pairs (g, x) with $gx = x$ in two ways: by g it is $\sum_g |\mathrm{Fix}(g)|$, by x it is $\sum_x |G_x|$. Group the second sum by orbits; on an orbit O each stabiliser has order $|G|/|O|$, so the orbit contributes $|G|$. Divide. □

Proposition 2.5.10 (Averaging is projection onto invariants)**.** *If a finite group G acts linearly on a vector space V over a field of characteristic 0, then $\pi(v) = \frac{1}{|G|} \sum_g gv$ is a projection onto the subspace V^G of invariants.*

This is the abstract statement of "average over the symmetry". Putnam 2009 B4 asks for the dimension of the space of polynomials of degree at most 2009 in x, y whose average on every circle centred at the origin is zero. Averaging over the rotation group is exactly π, the invariants of the rotation action are the polynomials in $x^2 + y^2$, and each even degree i contributes a one dimensional invariant space while each odd degree contributes none. Hence the conditions number 1005, one for each even degree from 0 to 2008, and the answer is $\binom{2011}{2} - 1005 = 2020050$. The dimension count *explains* the answer instead of computing it.

Example 2.5.11 (Putnam 2020 B4, in full: the multiplier identity)**.** Let V_n be the closed ± 1 walks $s_0 = 0, \dots, s_{2n} = 0$ with $|s_j - s_{j-1}| = 1$, and put $q(\mathbf{v}) = 1 + \sum_{j=1}^{2n-1} 3^{s_j}$. Find the average of $1/q$ over V_n.

Since $s_0 = 0$ contributes $3^0 = 1$, rewrite $q(\mathbf{v}) = \sum_{j=0}^{2n-1} 3^{s_j}$; the leading 1 was never an extra term, it was the $j = 0$ term in disguise, and noticing that is the whole bifurcation point. Now let $\mathbb{Z}/2n$ act by rotating the step word. Rotating by k produces heights $s_{j+k} - s_k$ (read the indices periodically), so

$$q(\varphi^k \mathbf{v}) = \sum_j 3^{s_{j+k} - s_k} = 3^{-s_k} q(\mathbf{v}).$$

The denominator merely scales. Therefore

$$\sum_{k=0}^{2n-1} \frac{1}{q(\varphi^k \mathbf{v})} = \frac{1}{q(\mathbf{v})} \sum_{k=0}^{2n-1} 3^{s_k} = 1$$

for every single $\mathbf{v}$: the reciprocals of q along one group orbit form a partition of unity. Each φ^k is a bijection of V_n, so summing over $\mathbf{v} \in V_n$ gives $2n \sum_{\mathbf{v}} 1/q(\mathbf{v}) = |V_n|$, and the average is $M(n) = 1/(2n)$. Hence $M(2020) = 1/4040$. Check it on $n = 1$: the two walks give $q = 4$ and $q = 4/3$, and $(1/4 + 3/4)/2 = 1/2$.

The general lemma is worth stating on its own, because the cue recurs. Suppose a group G of order N acts on a finite set X and a positive function F satisfies $F(gx) = \lambda(g, x)^{-1} F(x)$ where $F(x) = \sum_g \lambda(g, x)$. Then $\sum_g 1/F(gx) = 1$ for every x, so the average of $1/F$ over X is $1/N$. The definition of q was not guessed; its terms are precisely the multipliers of the rotation action.

- **WHEN** a problem asks for a count modulo a prime p, or asks you to prove p divides a count **DO** find a group of order a power of p acting on the set and count only the points fixed by all of it **BECAUSE** orbit sizes are $1, p, p^2, \dots$, so everything but the fixed points vanishes modulo p (Putnam 2007 A5, 2022 A3).
- **WHEN** a count involves tuples with a product or sum constraint **DO** let the cyclic shift act on the tuples; the constraint is usually shift invariant by conjugation **BECAUSE** $x_2 \cdots x_p x_1 = x_1^{-1}(x_1 \cdots x_p) x_1$, so the shift preserves "product is the identity" in any group (Putnam 2007 A5).
- **WHEN** a recurrence turns out to be periodic with small period T **DO** make $\mathbb{Z}/T$ act on the solution set and read the count modulo T off the constant solutions **BECAUSE** periodicity is exactly a $\mathbb{Z}/T$ action, and its fixed points are the constant sequences (Putnam 2022 A3).
- **WHEN** the coefficients of a linear congruence sum to 0 modulo p **DO** translate all variables by the same t; the solution set is a union of free orbits of size p **BECAUSE** a free action factors the count, so you may normalise one variable to 0 (Putnam 2017 B6).
- **WHEN** a quantity is only defined up to a symmetry and you must average it **DO** compute how the quantity transforms under the group first; if it scales by a multiplier, sum the reciprocals over one orbit **BECAUSE** multipliers that sum to the denominator give a partition of unity, so the average over the whole set is $1/|G|$ (Putnam 2020 B4).
- **WHEN** a condition says "the average over every rotated copy is zero" **DO** recognise averaging as the projection onto the invariant subspace and count dimensions **BECAUSE** the number of conditions is the dimension of the invariants, which is what makes the answer explicable (Putnam 2009 B4).

2.6 Symmetric and alternating groups: cycle type is everything

Definition 2.6.1. S_N is the group of permutations of an N element set. Every permutation is uniquely a product of disjoint cycles, and its *cycle type* is the multiset of cycle lengths, fixed points

included.

Theorem 2.6.2 (Sign). *There is a unique homomorphism* $\operatorname{sgn} : S_N \to \{\pm 1\}$ *sending every transposition to* -1. *A cycle of length* ℓ *has sign* $(-1)^{\ell-1}$, *so a permutation with* c *cycles on* N *points has sign* $(-1)^{N-c}$. *The kernel is the alternating group* A_N, *of index* 2 *for* $N \geq 2$.

Proof. Define $\operatorname{sgn}(\sigma) = \prod_{i<j} \frac{\sigma(j)-\sigma(i)}{j-i}$. Each unordered pair appears once in numerator and once in denominator up to sign, so the value is ± 1, and multiplicativity is a reindexing of the product. A transposition reverses the sign of exactly one factor together with an even number of others, giving -1. A cycle of length ℓ is a product of $\ell - 1$ transpositions, and disjoint cycles multiply. □

Proposition 2.6.3 (Conjugation is relabelling). *Two permutations are conjugate in* S_N *if and only if they have the same cycle type. Hence any conjugation invariant property, such as sign or being a square, depends only on cycle type, and adjoining fixed points changes nothing.*

This is the licence to transport a permutation through any isomorphism you like and compute in the model where it is easiest. Use it aggressively.

Proposition 2.6.4 (Products of actions). *If* σ *acts on* X *and* τ *on* Y, *then on* $X \times Y$, $\operatorname{sgn}(\sigma \times \tau) = \operatorname{sgn}(\sigma)^{|Y|} \operatorname{sgn}(\tau)^{|X|}$.

Proof. $\sigma \times \tau = (\sigma \times \mathrm{id})(\mathrm{id} \times \tau)$, and $\sigma \times \mathrm{id}$ is $|Y|$ disjoint copies of σ. □

Theorem 2.6.5 (Square roots in S_N). *σ is a square in S_N if and only if, for every* even *length L, the number of L-cycles of σ is even.*

Proof. Squaring an odd cycle of length L gives an L-cycle; squaring an even cycle of length L splits it into two cycles of length $L/2$. So squares have even cycles only in equal pairs. Conversely, an odd cycle is the square of the odd cycle obtained by walking the same points at step $(L+1)/2$, and two equal length cycles are the square of the single cycle obtained by interleaving them. □

The general statement, which you should carry because d-th powers appear as often as squares: walking a cycle of length L in steps of d splits it into $\gcd(d, L)$ cycles of length $L/\gcd(d, L)$. Being a d-th power is therefore always a condition on cycle type alone.

Example 2.6.6 (Putnam 2023 B5, in full). For which n is it true that for every m coprime to n there is a permutation π of $\mathbb{Z}/n$ with $\pi(\pi(k)) \equiv mk$? The answer is $n = 1$ and $n \equiv 2 \pmod 4$.

The problem literally says "$\pi^2 = \mu_m$", so the criterion above is the whole content: multiplication by m must have every even cycle length occurring an even number of times.

Case $4 \mid n$. Take $m = -1$. Its fixed points are 0 and $n/2$, and the remaining $n-2$ residues fall into $(n-2)/2 = n/2 - 1$ transpositions, which is odd because $n/2$ is even. One odd count of 2-cycles kills it, and indeed the sign is -1 so μ_{-1} is not even a square of anything.

Case n odd, $n > 1$. Pick a prime $p \mid n$, write $n = p^a b$ with $p \nmid b$, let g be a primitive root modulo p^a, and use the Chinese remainder theorem to choose $m \equiv g \pmod{p^a}$, $m \equiv 1 \pmod b$. Under $\mathbb{Z}/n \cong \mathbb{Z}/p^a \times \mathbb{Z}/b$, μ_m acts as μ_g on the first factor and trivially on the second, so every cycle of μ_g is repeated b times, and b is odd. On $\mathbb{Z}/p^a$ the set of elements of p-adic valuation i has $\varphi(p^{a-i})$ elements and g acts on it as a single cycle, because g is a primitive root modulo every p^{a-i}. So the cycle lengths $p^{a-i-1}(p-1)$, $i = 0, \ldots, a-1$, are distinct, each occurring exactly once, and each is even because p is odd. Multiplying by the odd b keeps every count odd. Failure.

Case $n \equiv 2 \pmod 4$. Write $n = 2n'$ with n' odd, and note m is odd. Translation by n' swaps the even residues and the odd residues, and it commutes with μ_m, since $mn' - n' = (m-1)n' \equiv 0 \pmod{2n'}$ as $m-1$ is even. So the restrictions of μ_m to the two halves are conjugate, hence have identical cycle types, hence every cycle length occurs an even number of times. A square root exists.

Case $n = 1$ is trivial. Note the contrast that makes the sign test insufficient on its own: for $n = 9$, $m = 2$, the cycle type is a 6-cycle, a 2-cycle and a fixed point, whose sign is $+1$, and yet the single

2-cycle blocks the square root. Sign is a coarse obstruction; the full cycle type criterion is the sharp one.

Example 2.6.7 (Putnam 2012 B6, in full)**.** Let $p \equiv 2 \pmod 3$ be an odd prime and $\pi(x) = x^3$ on $\mathbb{Z}/p$. Show π is even if and only if $p \equiv 3 \pmod 4$.

Since $p \equiv 2 \pmod 3$, $\gcd(3, p-1) = 1$ and cubing is a bijection. The fixed point 0 contributes nothing to the cycle type, so discard it and work on $\mathbb{F}_p^\times$. That group is cyclic of order $m = p-1$, and under the isomorphism $\mathbb{F}_p^\times \cong \mathbb{Z}/m$ cubing becomes multiplication by 3. Conjugation invariance of sign means we may compute there instead.

Write $m = 2^a u$ with u odd, so $\mathbb{Z}/m \cong \mathbb{Z}/2^a \times \mathbb{Z}/u$ and μ_3 acts componentwise. By the product rule, $\operatorname{sgn}(\mu_3) = \operatorname{sgn}(\mu_3|_{\mathbb{Z}/2^a})^u \operatorname{sgn}(\mu_3|_{\mathbb{Z}/u})^{2^a}$. Since p is odd, $a \geq 1$, so 2^a is even and the second factor is $+1$; and u is odd, so only the sign on $\mathbb{Z}/2^a$ survives.

If $a = 1$, multiplication by 3 on $\mathbb{Z}/2$ is the identity, sign $+1$. If $a \geq 2$, count cycles on $\mathbb{Z}/2^a$. The order of 3 modulo 2^j is $1, 2, 2, 4, 8, \ldots$, namely 2^{j-2} for $j \geq 3$. Splitting by the valuation i of the element: for $i \leq a-3$ the set has 2^{a-i-1} elements in cycles of length 2^{a-i-2}, giving 2 cycles; the levels $i = a-2$, $i = a-1$ and $i = a$ give one cycle each. So there are $2(a-2)+3 = 2a-1$ cycles among 2^a points, and the sign is $(-1)^{2^a-(2a-1)} = -1$ because $2^a - 2a + 1$ is odd for $a \geq 2$.

Hence π is even exactly when $a = 1$, that is $p - 1 \equiv 2 \pmod 4$, that is $p \equiv 3 \pmod 4$.

Sign also appears as a bookkeeping device rather than a subject. In Putnam 2016 B4 the expected determinant of $A - A^t$ is computed by expanding over permutations: disjoint cycles use disjoint variables, which licenses factoring the expectation, and the sign $(-1)^n$ of a fixed point free involution cancels the $(-1)^n$ from antisymmetry, so the surviving terms do not cancel each other.

- **WHEN** a problem asks whether a permutation is a square, or a d-th power **DO** compute its cycle type and apply the criterion: every even length must occur an even number of times (for d-th powers, use that a step d walk on an L-cycle gives $\gcd(d, L)$ cycles of length $L/\gcd(d, L)$) **BECAUSE** being a d-th power is conjugation invariant, and cycle type is the complete conjugacy invariant (Putnam 2023 B5).
- **WHEN** you need the parity of a permutation defined by a formula **DO** transport it through an isomorphism to the model where it becomes multiplication on a cyclic group, then split the cyclic group into its prime power factors and use $\operatorname{sgn}(\sigma \times \tau) = \operatorname{sgn}(\sigma)^{|Y|} \operatorname{sgn}(\tau)^{|X|}$ **BECAUSE** sign depends only on cycle type, which is unchanged by relabelling (Putnam 2012 B6).
- **WHEN** sign says a permutation is odd **DO** stop: it has no square root, and you are done **BECAUSE** squares lie in the kernel of sgn. But if the sign is $+1$, that proves nothing, so go to the full cycle type (Putnam 2023 B5, $n = 9$, $m = 2$).
- **WHEN** you are expanding a determinant or permanent over S_n by hand **DO** group the terms by cycle type; disjoint cycles involve disjoint entries, so expectations and products factor **BECAUSE** independence follows from disjointness of the index sets (Putnam 2016 B4).

2.7 Finite abelian groups: structure, elementary abelian pieces, characters

Theorem 2.7.1 (Structure theorem)**.** *Every finite abelian group is isomorphic to a direct product of cyclic groups of prime power order, uniquely up to reordering. Equivalently it is $\mathbb{Z}/d_1 \times \cdots \times \mathbb{Z}/d_r$ with $d_1 \mid d_2 \mid \cdots \mid d_r$, and this chain of* invariant factors *is determined by the group. More generally, every finitely generated abelian group is $\mathbb{Z}^s$ times such a finite part.*

The proof is the theory of finitely generated modules over a principal ideal domain, executed concretely by Smith normal form: any integer matrix A can be written $UAV = \operatorname{diag}(d_1, \ldots, d_r)$ with U, V invertible over $\mathbb{Z}$ and $d_1 \mid \cdots \mid d_r$, and then $\mathbb{Z}^n / A\mathbb{Z}^m$ is read straight off the diagonal. I am not

proving it here; it is long and it is the only genuine black box in this chapter. Know the statement, know that Smith normal form is the algorithm, and know that the invariant factors are what make the answer well defined. That last point is the content of Putnam 2008 A3, where a gcd and lcm rewriting process on a sequence of integers terminates in a divisibility chain independent of the choices: the combinatorial uniqueness argument is secretly proving that invariant factors are well defined.

Proposition 2.7.2 (The p-torsion is a vector space)**.** *For an abelian group G and a prime p, the set $G[p] = \{g : g^p = e\}$ is a subgroup, and it is a vector space over $\mathbb{F}_p$. Hence $|G[p]| = p^e$ and the number of elements of order exactly p is $p^e - 1$. For a finite abelian p-group, e is the number of cyclic factors, so $|G[p]| = p$ forces G to be cyclic.*

Proposition 2.7.3 (Partition by generated cyclic subgroup)**.** *Declare $g \sim h$ when $\langle g\rangle = \langle h\rangle$. The classes are the generator sets of the cyclic subgroups, a cyclic subgroup of order d has $\varphi(d)$ generators, and all of them have order d. Hence for any function F,*

$$\sum_{g\in G} F(\operatorname{ord} g) = \sum_{d} \varphi(d) F(d)\, c_d,$$

where c_d is the number of cyclic subgroups of order d. With $F \equiv 1$ and G cyclic of order n this is the familiar $\sum_{d\mid n} \varphi(d) = n$.

Example 2.7.4 (Putnam 2009 A5, in full)**.** Is there a finite abelian G with $\prod_{g\in G} \operatorname{ord}(g) = 2^{2009}$?

If any element had odd prime order q then q would divide the product, so G is a 2-group and the question is whether $\sum_g v_2(\operatorname{ord} g) = 2009$ is achievable. Apply the partition above with $F(d) = v_2(d)$ and $d = 2^h$:

$$\sum_{g} v_2(\operatorname{ord} g) = \sum_{h\geq 1} 2^{h-1}\, h\, c_{2^h}.$$

Read it modulo 4. The term $h = 1$ contributes c_2; the term $h = 2$ contributes $4c_4$; every $h \geq 3$ contributes a multiple of 4. So the sum is $\equiv c_2 \pmod 4$, and c_2 is the number of elements of order 2, which is $2^e - 1$. Since $2009 \equiv 1 \pmod 4$, we need $2^e - 1 \equiv 1 \pmod 4$, so $2^e \equiv 2$, so $e = 1$: exactly one element of order 2. Only now does the structure theorem earn its keep, converting that into "G is cyclic", say $G = \mathbb{Z}/2^k$.

For a cyclic 2-group the sum is exact: $\sum_{h=1}^{k} h2^{h-1} = (k-1)2^k + 1$. Setting this to 2009 gives $(k-1)2^k = 2008 = 2^3 \cdot 251$, so $2^k \mid 2008$ forces $k \leq 3$, and $k = 1, 2, 3$ give $0, 4, 16$. None is 2008. The answer is no.

Notice the shape: a weighted count over all elements became a weighted count over cyclic subgroups, the weights $h2^{h-1}$ were read modulo 4 to collapse everything but one term, and the structure theorem converted a numerical fact into a structural one.

Definition 2.7.5. A *character* of a finite abelian group G with values in a field k is a homomorphism $\chi : G \to k^\times$. For $k = \mathbb{C}$ the characters form a group $\widehat{G}$ under pointwise multiplication, of order $|G|$, and $\widehat{G} \cong G$; explicitly, under $G \cong \prod_t \mathbb{Z}/d_t$ they are $\chi_{\mathbf{a}}(\mathbf{x}) = \prod_t e^{2\pi i a_t x_t/d_t}$.

Theorem 2.7.6 (Orthogonality)**.** $\sum_{g\in G} \chi(g) = |G| \cdot 1_k$ *if χ is trivial and 0 otherwise. Dually,* $\sum_{\chi\in\widehat{G}} \chi(x) = |G|$ *if $x = e$ and 0 otherwise, and* $\sum_x \chi(x)\overline{\psi(x)} = |G|\,[\chi = \psi]$.

Proof. Let $\Sigma = \sum_g \chi(g)$ and pick h with $\chi(h) \neq 1$. Reindexing by $g \mapsto hg$, which is a bijection, gives $\Sigma = \sum_g \chi(hg) = \chi(h)\Sigma$, so $(\chi(h) - 1)\Sigma = 0$ and $\Sigma = 0$. The dual statement is the same argument in $\widehat{G}$. □

Look at what that proof is: translation is a bijection, one more time. The reindexing move is forced, not lucky, and the same identity is what powers roots of unity filters and Gauss sums elsewhere in the archive.

Example 2.7.7 (Putnam 2021 A5, in full). $A = \{n : 1 \le n \le 2021,\ \gcd(n, 2021) = 1\}$ and $S(j) = \sum_{n \in A} n^j$. For which j is $2021 \mid S(j)$? Note $2021 = 43 \cdot 47$, so work modulo each prime. By the Chinese remainder theorem each nonzero residue modulo 43 is hit by exactly 46 elements of A, so $S(j) \equiv 46 \sum_{a \in \mathbb{F}_{43}^{\times}} a^j \pmod{43}$, and $46 \equiv 3 \not\equiv 0$. The map $a \mapsto a^j$ is a character of the cyclic group $\mathbb{F}_{43}^{\times}$, trivial exactly when $42 \mid j$. So orthogonality gives $\sum a^j = -1$ when $42 \mid j$ and 0 otherwise. Likewise modulo 47 with 46 in place of 42. Hence $2021 \mid S(j)$ if and only if $42 \nmid j$ and $46 \nmid j$.

Characters also diagonalise convolution, which is the entire mechanism of Putnam 2011 A6: a random walk on a finite abelian G driven by a generating set has convolution as its transition operator, the characters are its eigenvectors with eigenvalues $\lambda_\chi = \frac{1}{k} \sum_i \chi(g_i)$, and orthogonality turns the requested sum of squared deviations into $\sum_{\chi \neq 1} |\lambda_\chi|^{2m}$. The rate b is the largest $|\lambda_\chi|$ over nontrivial χ, and $b = 1$ happens exactly when the step set lies in a coset of a proper subgroup. Two supporting facts: a character trivial on a generating set is trivial, because kernels are subgroups; and for every proper subgroup H there is a nontrivial character killing H, lifted from the nontrivial quotient G/H.

- **WHEN** you must sum a multiplicative quantity over all elements of a finite abelian group **DO** recognise it as a character and use orthogonality: the sum is 0 unless the character is trivial **BECAUSE** reindexing by translation multiplies the sum by $\chi(h)$, forcing it to vanish (Putnam 2021 A5).
- **WHEN** a finite abelian group has exactly one element of order 2 **DO** conclude it is cyclic **BECAUSE** $G[2]$ is an $\mathbb{F}_2$ vector space whose dimension counts the cyclic factors (Putnam 2009 A5).
- **WHEN** you need a weighted sum over the orders of all elements **DO** convert it to a sum over cyclic subgroups with weights $\varphi(d)$, then reduce modulo a small number to kill all but one term **BECAUSE** elements of the same generated cyclic subgroup have the same order, and there are $\varphi(d)$ of them (Putnam 2009 A5).
- **WHEN** a process on integers or on a group is claimed to have a unique outcome independent of choices **DO** look for the invariant factors: the outcome is usually the invariant factor decomposition of a fixed abelian group **BECAUSE** invariant factors are determined by the group, not by the presentation (Putnam 2008 A3).
- **WHEN** a random walk on a finite abelian group appears **DO** expand in characters; they are the eigenvectors of convolution **BECAUSE** convolution is diagonalised by characters, and orthogonality removes every cross term (Putnam 2011 A6).

2.8 Generation, presentations, commutators

Definition 2.8.1. $\langle T \rangle$ is the smallest subgroup containing T, equivalently the set of finite products of elements of T and their inverses. A *presentation* $\langle X \mid R \rangle$ is the quotient of the free group on X by the normal subgroup generated by the relators R.

Proposition 2.8.2. *If G is finite and $T \subseteq G$, the set of all finite products of elements of T, with no inverses allowed, already equals $\langle T \rangle$. In an infinite group this fails.*

Proof. The product set is finite, nonempty and closed under multiplication, so it is a subgroup by the earlier proposition; concretely $t^{-1} = t^{\mathrm{ord}(t)-1}$. □

Theorem 2.8.3 (von Dyck). *Let F be free on g, h and $W \in F$. Giving elements g_0, h_0 of a group G with $W(g_0, h_0) = e$ is the same as giving a homomorphism $\langle g, h \mid W \rangle \to G$ sending generators to g_0, h_0. Consequently any identity that holds in $\langle g, h \mid W \rangle$ holds for every such pair in every group.*

This is the licence to work purely formally with words. If a problem says "suppose a relation holds in a group G; prove another relation", then no property of G can possibly be needed, and you should stop looking for one.

Example 2.8.4 (Putnam 2016 A5, in full)**.** G is finite, generated by g and h, and $\operatorname{ord}(g)$ is odd. Show every element is a word $g^{m_1}h^{n_1}\cdots g^{m_r}h^{n_r}$ with all exponents ± 1 and $r \le |G|$.

Let $T = \{g^{\pm 1}h^{\pm 1}\}$, four elements. The representable elements with r syllable pairs are exactly the r-fold products T^r, so the claim is that $\bigcup_{r\le|G|} T^r = G$.

First, $\langle T\rangle = G$. The subgroup $N = \langle T\rangle$ contains $g^2 = (gh)(g^{-1}h)^{-1}$ and $h^2 = (gh^{-1})^{-1}(gh)$. It is normal: conjugating each of the four generators by g or by h lands back in N after inserting g^2 or h^2, for example $g(gh)g^{-1} = g^2(gh^{-1})^{-1}$ and $h(gh)h^{-1} = h^2(g^{-1}h)^{-1}$. In G/N we have $\bar{g}\bar{h} = \bar{e}$ and $\bar{g}^2 = \bar{e}$, so $G/N = \langle\bar{g}\rangle$ with $\bar{g}^2 = \bar{e}$. But $\operatorname{ord}(\bar{g})$ divides the odd number $\operatorname{ord}(g)$, so $\bar{g} = \bar{e}$ and $N = G$. Then the previous proposition removes the inverses: since G is finite, the semigroup generated by T is already $\langle T\rangle = G$.

Second, the bound $r \le |G|$. Put $U_r = T \cup T^2 \cup \cdots \cup T^r$. The U_r increase; if ever $U_r = U_{r+1}$ then $U_rT \subseteq U_r$, so U_r is stable and equals $\langle T\rangle = G$. Each strict step adds at least one element, so $U_{|G|} = G$.

The hypothesis is exactly right, and you can see why: the only obstruction to T generating is a homomorphism $\varphi : G \to \mathbb{Z}/2$ with $\varphi(g) = \varphi(h) = 1$, which cannot exist when $\operatorname{ord}(g)$ is odd. Take $G = \mathbb{Z}/2$ with $g = h = 1$ and the conclusion fails, so oddness is necessary, not decorative.

Example 2.8.5 (Putnam 2018 A4)**.** With $\gcd(m, n) = 1$ and $a_k = \lfloor mk/n\rfloor - \lfloor m(k-1)/n\rfloor$, suppose $gh^{a_1}gh^{a_2}\cdots gh^{a_n} = e$ in a group. Show $gh = hg$.

Two structural facts frame this. By von Dyck, nothing about G matters: you are proving an identity in $\langle g, h \mid W\rangle$ where W is the given relator. And by the rotation principle, every cyclic rotation of W is also a relator for free. That is the entire toolkit, and the closing line of the solution is exactly the rotation move: you exhibit the commutator as PQ where QP is a rotation of the relator, so it is the identity.

See the mechanism on the smallest nontrivial case, $m = 2$, $n = 3$, where $(a_1, a_2, a_3) = (0, 1, 1)$ and $W = g^2hgh$. Rotating, $ghg^2h = e$ as well. Comparing the two relators gives $g^2hg = ghg^2$, and cancelling one g on the right gives $g^2h = ghg$, then one g on the left gives $gh = hg$.

The general case is an induction that runs the Euclidean algorithm on (m, n), replacing (g, h) at each step by a new generating pair that reduces the relator to the same shape with smaller parameters; the sequence (a_k) is the Christoffel word of slope m/n and is stable under exactly those substitutions. I am naming that induction as a knowledge gap rather than pretending the two rotations settle it in general. What you must carry away is the pair of reusable moves: rotate the relator, and quote von Dyck to stop hunting for a property of G.

The sharpened conclusion is worth banking too. If g and h commute and $g^nh^m = e$ with $\gcd(m, n) = 1$, choose u, v with $um + vn = 1$ and set $c = g^vh^{-u}$; then $c^m = g$ and $c^{-n} = h$, so $\langle g, h\rangle$ is cyclic generated by c. A coprime relation between commuting elements always produces a single generator, and that two line derivation is the standard model.

Definition 2.8.6. The *commutator* is $[g, h] = ghg^{-1}h^{-1}$. It is trivial exactly when g and h commute. The subgroup generated by all commutators is normal, and the quotient $G/[G, G]$ is the largest abelian quotient.

Proposition 2.8.7 (Involutions force commuting)**.** *In any group, if $g^2 = h^2 = (gh)^2 = e$ then $gh = hg$. A group in which every element squares to the identity is abelian, and is a vector space over $\mathbb{F}_2$.*

Proof. $(gh)^{-1} = h^{-1}g^{-1} = hg$, and $(gh)^2 = e$ says $(gh)^{-1} = gh$. □

Example 2.8.8 (Putnam 2010 A5, in full)**.** $G \subseteq \mathbb{R}^3$ is a group under $*$, and for all $a, b \in G$ either $a \times b = a * b$ or $a \times b = 0$. Prove $a \times b = 0$ always.

Name the clash first, because it is the reason the problem is possible: a group law is associative with an identity, and the cross product is neither associative nor unital. It is a Lie algebra, anticommutative and satisfying the Jacobi identity, and every vector annihilates itself. Agreement between the two can only survive where the cross product degenerates.

Let e be the identity of G. For any a, $a * e = a$, so the hypothesis gives $a \times e = a$ or $a \times e = 0$. The first forces $a \cdot a = a \cdot (a \times e) = 0$, hence $a = 0$. So every nonzero $a \in G$ satisfies $a \times e = 0$.

If $e \neq 0$, every element of G lies in the line $\mathbb{R}e$, so all elements are parallel and all cross products vanish. Done.

So assume $e = 0$ and suppose, for contradiction, that $a \times b \neq 0$ for some $a, b \in G$. Then $a * b = a \times b =: c \neq 0$, and $b * a = b \times a = -c$, so a and b do not commute; that is the first lemma, and it is pure anticommutativity.

Second lemma: if $x \times y \neq 0$ then $x * x$ is parallel to y. Proof: put $s = x * x$ and $z = x * y = x \times y \neq 0$. Then $z \perp x$ and $z \neq 0$, so $x \times z \neq 0$ and $x * z = x \times z$. By associativity,

$$s * y = (x * x) * y = x * (x * y) = x * z = x \times (x \times y) = (x \cdot y)x - |x|^2 y.$$

Dotting with y gives $(x \cdot y)^2 - |x|^2|y|^2 = -|x \times y|^2 \neq 0$, so $s * y$ is not perpendicular to y; but if $s \times y$ were nonzero the hypothesis would force $s * y = s \times y \perp y$. Hence $s \times y = 0$.

Now apply the second lemma three times. Since $c = a \times b$ is perpendicular to both a and b and is nonzero, $a \times c \neq 0$ and $b \times c \neq 0$, while $a \times b \neq 0$. So $a * a$ is parallel to b and to c; $b * b$ is parallel to a and to c; $c * c$ is parallel to a and to b. In each case the two targets are non-parallel nonzero vectors, so the only vector parallel to both is 0. Hence $a * a = b * b = c * c = 0 = e$, that is, a, b and $a * b$ are all involutions. By the previous proposition, a and b commute, contradicting the first lemma. Therefore $a \times b = 0$ for all a, b.

Example 2.8.9 (Putnam 2008 A6, and where a natural route dies)**.** Prove there is a constant $c > 0$ such that every nontrivial finite group G has a sequence of length at most $c \log |G|$ in which every element is the product of some ordered subsequence.

The reason logarithms can appear at all is doubling: if H is the set of achievable products and $g \notin H$, then H and Hg are disjoint sets of equal size, because right translation is a bijection, so appending g doubles the count. Doubling $\log_2 |G|$ times exhausts G, and that is the whole shape of the bound. Two supporting moves: inversion reverses products, $(x_1 \cdots x_r)^{-1} = x_r^{-1} \cdots x_1^{-1}$, so listing the inverses in decreasing index order makes every h^{-1} an ordered subsequence product of a mirrored block; and appending $x, x^2, x^4, \ldots$ climbs a cyclic quotient in about $\log_2 p$ steps.

Now the part you must know so you do not waste twenty minutes in the arena. The obvious induction, "take a normal subgroup H with G/H cyclic of prime order and append binary powers of a lift", covers only solvable groups: chief factors of a general finite group need not be cyclic, and a nonabelian simple group has no proper nontrivial normal subgroup at all. Knowing exactly where a natural route dies is worth as much as knowing a route that works.

- **WHEN** a problem says "a relation holds in a group G, prove another relation" **DO** work formally in $\langle \text{generators} \mid \text{relator} \rangle$ and stop looking for properties of G **BECAUSE** von Dyck says such a pair is exactly a homomorphism out of the presented group, so only the relator matters (Putnam 2018 A4).

- **WHEN** you are told a set of words generates and asked to represent every element **DO** show the generated subgroup is everything by killing the abelian quotient G/N, then use finiteness to drop the inverses **BECAUSE** the only obstruction to a symmetric letter set generating is a homomorphism to $\mathbb{Z}/2$, which odd order forbids (Putnam 2016 A5).

- **WHEN** three elements g, h, gh all square to the identity **DO** conclude $gh = hg$ at once **BECAUSE** $(gh)^{-1} = hg$, so $(gh)^2 = e$ literally says $gh = hg$ (Putnam 2010 A5).

- **WHEN** a group law is claimed to agree with a non-associative or non-unital operation **DO** use associativity and the identity to force the other operation to degenerate **BECAUSE** the two structures clash on exactly those axioms, so the agreement can only survive on the degenerate locus (Putnam 2010 A5).
- **WHEN** a bound of the shape $\log |G|$ is requested **DO** build the object greedily and prove each step at least doubles a set **BECAUSE** translation is a bijection, so a disjoint translate doubles the count exactly (Putnam 2008 A6).

2.9 Sylow, p-groups and how lightly the Putnam uses them

You already have Cauchy from the McKay action. Sylow is the next layer, and the Putnam uses it far less than a textbook would suggest, so learn the statements and do not invest in the proofs.

Theorem 2.9.1 (Sylow). *Let $|G| = p^a m$ with $p \nmid m$. Then Sylow p-subgroups (subgroups of order p^a) exist; any two are conjugate; every p-subgroup lies in one; and the number n_p of them satisfies $n_p \equiv 1 \pmod{p}$ and $n_p \mid m$.*

The congruence $n_p \equiv 1 \pmod{p}$ is the part that shows up. In Putnam 2007 A5, distinct subgroups of order p meet only in the identity, so if k is their number then the elements of order p number $k(p-1)$, and $p \mid n+1$ becomes $k \equiv 1 \pmod{p}$. Recognising that the problem is a Sylow type congruence is how you choose the tool; the actual proof you should give is the McKay action, because you can reconstruct it.

Two more facts about p-groups are worth carrying. A group of order p^a has nontrivial centre, by the class equation $|G| = |Z(G)| + \sum[G : Z(x_i)]$ with every index a positive power of p. And Frobenius's theorem, that $t \mid |G|$ implies t divides the number of solutions of $x^t = e$, is a genuine knowledge gap: its proof is far harder than any Putnam problem that uses it, so cite it only as a sanity check on an answer, never as a step in a solution.

- **WHEN** a count of subgroups of prime order is congruent to 1 modulo p **DO** recognise the Sylow family, but prove it with the shift action on p-tuples **BECAUSE** the McKay argument is reconstructible in the arena and Sylow's proof is not (Putnam 2007 A5).
- **WHEN** a finite p-group acts and you need a fixed point **DO** use $|X| \equiv |X^G| \bmod p$; if $p \nmid |X|$ there is a fixed point **BECAUSE** orbit sizes are powers of p, so a non-multiple of p forces a singleton orbit.

2.10 Rings: ideals, quotients, and computing inside a quotient

Definition 2.10.1. A *ring* is an abelian group under $+$ with an associative multiplication distributing over $+$; in this chapter rings have 1 and, unless stated, are commutative. A *unit* is an element with a multiplicative inverse; a *zero divisor* is a nonzero a with $ab = 0$ for some nonzero b. An *integral domain* has no zero divisors; a *field* is a ring in which every nonzero element is a unit. The *characteristic* is the least $n \geq 1$ with $n \cdot 1 = 0$, or 0 if there is none.

Definition 2.10.2. An *ideal* $I \trianglelefteq R$ is an additive subgroup with $RI \subseteq I$. The quotient ring R/I has the same universal property as the quotient group: ring homomorphisms out of R killing I are exactly ring homomorphisms out of R/I, and $R/\ker\varphi \cong \operatorname{im}\varphi$.

The reason ideals matter to you is not abstraction. It is that *divisibility statements are equations in a quotient*. "d divides F" means "$F = 0$ in $R/(d)$". That single translation is a reroute you will use several times a paper.

Example 2.10.3 (Putnam 2024 A2, in full)**.** For which real polynomials p is there a real q with $p(p(x)) - x = (p(x) - x)^2 q(x)$?

Set $r = p - x$, so the question is when $r^2 \mid p(p(x)) - x$, that is, when $p(p(x)) - x$ vanishes in $k[x]/(r^2)$. In that quotient the class of r is nilpotent of order 2, so $x + r$ behaves like x plus an infinitesimal and every polynomial evaluated at $x + r$ equals its first order Taylor expansion. This is the conceptual reason the Taylor step is exact, and the reason divisibility by a square is tangency information.

Compute: $p(p(x)) - x = r(p(x)) + r(x)$, and $r(p(x)) = r(x + r) \equiv r + r'r \pmod{r^2}$. So $p(p(x)) - x \equiv r(2 + r') \pmod{r^2}$, and the condition is $r \mid 2 + r'$. If r is a nonzero constant c this always holds, giving $p(x) = x + c$. If $\deg r \geq 1$ then $\deg(2 + r') < \deg r$, so divisibility forces $2 + r' = 0$, that is $r' = -2$ and $r = -2x + c$, giving $p(x) = -x + c$. The answers are $p(x) = x + c$ and $p(x) = -x + c$; the first family includes $p(x) = x$.

Reduction modulo m is a ring homomorphism $\mathbb{Z} \to \mathbb{Z}/m$, so evaluation of an integer polynomial commutes with reduction. That is the structural reason, not a computational accident, that a collapse such as $f(f(n) + 1) \equiv f(1) \pmod{f(n)}$ is legitimate in Putnam 2007 B1: you are evaluating the same polynomial in the quotient ring.

Example 2.10.4 (Putnam 2008 B4)**.** If $h \in \mathbb{Z}[x]$ takes distinct values modulo p^2 on $0, \ldots, p^2 - 1$, show it takes distinct values modulo p^3 on $0, \ldots, p^3 - 1$. Work fibrewise. Distinctness modulo p^2 applied to x and $x + p$ uses $h(x + p) \equiv h(x) + ph'(x) \pmod{p^2}$ and forces $h'(x) \not\equiv 0 \pmod{p}$ for every x. Now $\mathbb{Z}/p^3 \to \mathbb{Z}/p^2$ is a surjection with fibres of size p, h covers the induced map, and on the fibre over x the induced map is $t \mapsto h(x) + p^2 h'(x)t$, an affine map of $\mathbb{Z}/p$ with invertible slope. The base map is a bijection by hypothesis and every fibre map is a bijection, so h is a bijection modulo p^3.

Definition 2.10.5. A *Euclidean domain* has a size function permitting division with remainder; a *principal ideal domain* has every ideal generated by one element; a *unique factorisation domain* factors elements into irreducibles uniquely up to units and order. Euclidean implies PID implies UFD. $\mathbb{Z}$ and $k[x]$ for a field k are Euclidean; $k[x, y]$ is a UFD but not a PID.

You need this hierarchy for two purposes: to know that $k[x]$ has gcds by the Euclidean algorithm and that a degree d polynomial over a field has at most d roots, and to know that $\mathbb{Z}[x]$ and $k[x, y]$ are not principal ideal domains, so an argument that quietly divides in them is wrong. The *field of fractions* of a domain R is built as pairs a/b modulo the obvious relation; it is how $\mathbb{Z}$ sits in $\mathbb{Q}$ and $k[x]$ in $k(x)$, and it is what lets you clear denominators and then argue integrally.

Definition 2.10.6. u is *integral of degree at most k* over R if it satisfies a monic degree k polynomial over R; equivalently $R[u]$ is generated as an R-module by $1, u, \ldots, u^{k-1}$.

Cayley and Hamilton is the case of a matrix over R with its characteristic polynomial; the minimal polynomial of an algebraic number is the case $R = \mathbb{Q}$; and the following is the case that appeared on a paper.

Example 2.10.7 (Putnam 2007 B5, in full)**.** Show there are polynomials $P_0, \ldots, P_{k-1}$ with $\lfloor n/k \rfloor^k = \sum_{i=0}^{k-1} P_i(n) \lfloor n/k \rfloor^i$ for every integer n.

The nominal solution is a case analysis on n mod k, of variable length k, which cannot be written down uniformly. The reroute is the annihilating polynomial. Write $u = \lfloor n/k \rfloor$ and $v_j = (n - j)/k$ for $j = 0, \ldots, k - 1$. For each n, exactly one j makes $u = v_j$, so

$$\prod_{j=0}^{k-1} (u - v_j) = 0$$

identically in n, and expanding gives $u^k = \sum_{i=0}^{k-1} (-1)^{k-i-1} e_{k-i}(v_0, \ldots, v_{k-1}) u^i$, where e_m is the mth elementary symmetric polynomial and each coefficient is a polynomial in n. The identity never

requires knowing which case occurs, which is the entire point: a case analysis of variable length has become a single polynomial identity uniform in k. In the language above, you have shown $\lfloor n/k \rfloor$ is integral of degree at most k over $\mathbb{Q}[n]$.

Proposition 2.10.8 (Multiplicative valuations kill finiteness)**.** *If a finite set X is closed under an operation and carries a function N to the positive reals with $N(a * b) = N(a)N(b)$, then $N \equiv 1$ on the part of X where it is nonzero.*

Proof. If $N(a) > 1$ then $N(a), N(a)^2, N(a)^3, \ldots$ are infinitely many distinct values realised in the finite set X; if $N(a) < 1$ use the same argument or invert. □

That is the abstract statement of the whole of Putnam 2022 B2: the norm is the valuation, and finiteness pins every nonzero element to the unit sphere. Units of $\mathbb{Z}/n$ deserve one more line, because they carry the other recurring finiteness statement: if p is an odd prime and $u^2 \equiv 1 \pmod{p^k}$ then $u \equiv \pm 1$, since p divides at most one of $u - 1$ and $u + 1$; modulo an odd q with r distinct prime factors there are 2^r square roots of unity. Putnam 2015 A5 turns on exactly that: the sign information survives for a prime power and dies for a composite modulus.

- **WHEN** a problem says "d^2 divides F" or "d divides F" **DO** rewrite it as "$F = 0$ in $R/(d)$" or $R/(d^2)$ and compute in the quotient, where d is nilpotent **BECAUSE** in $k[x]/(r^2)$ the class of r is a square zero infinitesimal, so substitution is exactly first order Taylor (Putnam 2024 A2).
- **WHEN** a quantity is known to equal one of k listed expressions, but you cannot tell which **DO** multiply the k differences together to get 0 and expand **BECAUSE** an annihilating polynomial replaces a case analysis with an identity that never asks which case occurred (Putnam 2007 B5).
- **WHEN** an integer polynomial identity is being checked modulo m **DO** push everything through the reduction homomorphism at the start, and evaluate afterwards **BECAUSE** reduction preserves $+$ and $\times$, so evaluation commutes with it (Putnam 2007 B1).
- **WHEN** a finite set carries a multiplicative size function **DO** conclude the size is identically 1 off the zero part **BECAUSE** otherwise the powers of one value would be infinitely many distinct sizes inside a finite set (Putnam 2022 B2).
- **WHEN** you want to divide but the ring is $\mathbb{Z}[x]$ or $k[x, y]$ **DO** stop, pass to the field of fractions, clear denominators, and only then argue about integrality **BECAUSE** those rings are not principal ideal domains, so "divide" is not a legal move in them.

2.11 Fields, finite fields and Frobenius

Proposition 2.11.1. *For p prime, $\mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$ is a field. Consequently a nonzero linear form over $\mathbb{F}_p$ in n variables has exactly p^{n-1} zeros, and any coefficient a with $0 < a < p$ is invertible.*

Proof. If $0 < a < p$ then $\gcd(a, p) = 1$, so $ua + vp = 1$ for some integers, and u is the inverse. A nonzero linear form is a surjective linear map onto $\mathbb{F}_p$, so its kernel is a hyperplane. □

This tiny fact is used more often than anything else in this section. In Putnam 2017 B6 it supplies both that a coefficient sum is invertible and that each individual coefficient is nonzero. In Putnam 2006 B4 the whole problem is moved to $\mathbb{F}_2$: a d dimensional subspace of $\mathbb{F}_2^n$ has exactly 2^d elements, and reduction modulo 2 cannot increase rank because an integer matrix with an odd maximal minor has that minor nonzero over $\mathbb{R}$; hence at most 2^k vertices of the unit cube lie in a k dimensional subspace, which is the answer. In Putnam 2017 A6, identifying three colours with $\mathbb{F}_3$ turns "exactly two of these three edges have the same colour" into the single linear inequation $u + v + w \neq 0$, because in $\mathbb{F}_3$ a triple sums to zero exactly when its entries are all equal or all distinct. Twenty face conditions become twenty linear inequations, and the problem is now linear algebra.

Theorem 2.11.2 (Existence and uniqueness of finite fields)**.** *For every prime p and every $n \geq 1$ there is a field $\mathbb{F}_{p^n}$ with p^n elements, unique up to isomorphism; equivalently there is an irreducible polynomial of degree n over $\mathbb{F}_p$. Every finite field has prime power order.*

I am not proving this; it is the splitting field of $x^{p^n} - x$ over $\mathbb{F}_p$, and that is the sentence to remember. It is a genuine knowledge gap in the sense of this book: name it, cite it, and know that without it Putnam 2014 B5 has no construction at all.

Theorem 2.11.3 (The multiplicative group is cyclic)**.** $\mathbb{F}_q^\times$ *is cyclic of order $q-1$.*

Proof. For each $d \mid q-1$, the polynomial $x^d - 1$ has at most d roots in a field, so the elements of order exactly d number either 0 or $\varphi(d)$: if there is one such element x, then $\langle x \rangle$ already supplies d roots of $x^d - 1$, so every element of order d lies in it and there are exactly $\varphi(d)$ of them. Summing over d and using $\sum_{d \mid q-1} \varphi(d) = q-1$, no count can be 0, so in particular there are $\varphi(q-1) \geq 1$ elements of order $q-1$. □

That proof is worth reconstructing once by hand, because it is the model for every "count by order and compare with $\sum \varphi(d) = n$" argument.

Theorem 2.11.4 (Frobenius, the freshman's dream)**.** *In a commutative ring of characteristic p, the map $a \mapsto a^p$ is a ring homomorphism, so $(a+b)^p = a^p + b^p$ and, iterating, $(a+b)^{p^m} = a^{p^m} + b^{p^m}$. On $\mathbb{F}_p$ it is the identity map.*

Proof. $\binom{p}{i}$ is divisible by p for $0 < i < p$, since p divides the numerator of $p!/(i!(p-i)!)$ and not the denominator. Multiplicativity is free. On $\mathbb{F}_p$, $a^p = a$ is Fermat. □

Two immediate consequences you should treat as reflexes. In $\mathbb{F}_p[x]$, $(x-c)^p = x^p - c^p$, so $x^p - 1 = (x-1)^p$ and $1 + x + \cdots + x^{p-1} = (x-1)^{p-1}$. And for a matrix M over $\mathbb{F}_p$, $M^{p^m} = I$ is equivalent to $(M-I)^{p^m} = 0$, that is, to M being unipotent.

Example 2.11.5 (Putnam 2019 A5, in full)**.** $q(x) = \sum_{k=1}^{p-1} a_k x^k$ with $a_k = k^{(p-1)/2} \bmod p$, so a_k is the Legendre symbol of k. Find the largest n with $(x-1)^n \mid q$ in $\mathbb{F}_p[x]$. The answer is $(p-1)/2$.

The cue is that a sum of $p-1$ monomials with ± 1 coefficients is being tested for vanishing to high order at $x = 1$, and the only structural fact about $\mathbb{F}_p[x]$ that produces high order vanishing at 1 for free is Frobenius: $1 + x + \cdots + x^{p-1} = (x-1)^{p-1}$. Once you have collapsed the full geometric sum, the problem becomes a single statement about the order of vanishing of q at 1, and the Legendre symbol coefficients decide it. Sanity check the answer on small p: for $p = 3$, $q = x - x^2 = -x(x-1)$ and $n = 1$; for $p = 5$, $q = x - x^2 - x^3 + x^4 = x(1-x)^2(1+x)$ and $n = 2$.

Example 2.11.6 (Putnam 2012 A5, in full)**.** $G(x) = v + Mx$ on $\mathbb{F}_p^n$. For which (p, n) can the p^n iterates $G^{(k)}(0)$ be distinct? The answer is $n = 1$ with any p, and $(p, n) = (2, 2)$.

Distinctness of the whole forward orbit of one point in a set of size p^n says that G is a permutation consisting of a single p^n-cycle, so $\operatorname{ord}(G) = p^n$ in the affine group. That is the only form in which the hypothesis can interact with the characteristic, and converting it is the bifurcation point.

Now compute the order. Write $M = I + N$. Then $G^{(k)}(x) = M^k x + \left(\sum_{j<k} M^j\right) v$, and $\sum_{j=0}^{k-1} (I+N)^j = \sum_i \binom{k}{i+1} N^i$. Take $k = p^m$. Every $\binom{p^m}{i+1}$ with $0 < i+1 < p^m$ is divisible by p, so the only surviving term is $i+1 = p^m$, giving N^{p^m-1}. Also $M^{p^m} = I + N^{p^m}$ by Frobenius. Hence if $p^m \geq n+1$, both N^{p^m} and N^{p^m-1} vanish (since N is nilpotent of index at most n once $M^{p^m} = I$ forces unipotence), and $G^{(p^m)} = \mathrm{id}$. So $\operatorname{ord}(G) \leq p^{\lceil \log_p(n+1) \rceil}$, and we need $n \leq \lceil \log_p(n+1) \rceil$.

For $p = 2$ this holds at $n = 1$ and $n = 2$ and fails from $n = 3$ on. For odd p it holds only at $n = 1$. Both surviving cases are realisable: for $n = 1$ take $G(x) = x + 1$; for $(p, n) = (2, 2)$ take $N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $v = (0, 1)^t$, so that $G^{(2)}(x) = x + Nv$ with $Nv \neq 0$, giving $\operatorname{ord}(G) = 4$ and a full orbit.

Example 2.11.7 (Putnam 2011 A4, in full)**.** For which n is there an integer $n \times n$ matrix with every row dotted with itself even and every two distinct rows dotted odd? The answer is n odd.

Reduce modulo 2 at once; the hypothesis is a parity hypothesis and nothing else. Over $\mathbb{F}_2$, Frobenius says $x^2 = x$, so $r \cdot r = \sum_t r_t^2 = \sum_t r_t$. The diagonal condition therefore says every row sums to 0, that is $\overline{M}\mathbf{1} = 0$, so $\overline{M}$ is singular and $\det \overline{M} = 0$. The off diagonal conditions say $\overline{M}\,\overline{M}^t = J - I$ over $\mathbb{F}_2$. If n is even then $J^2 = nJ = 0$, so $(J - I)^2 = I$ and $J - I$ is invertible; but $\det(\overline{M}\,\overline{M}^t) = (\det \overline{M})^2 = 0$. Contradiction. For n odd, $J - I$ itself works: a row dotted with itself is $n - 1$, even; two distinct rows share $n - 2$ ones, odd.

Field extensions appear on the Putnam only in the smallest doses, so keep the syllabus honest. You need: L/K is a K-vector space and $[L : K]$ is its dimension; degrees multiply in towers; an element algebraic over K generates $K(\alpha)$ of degree equal to that of its minimal polynomial; algebraic numbers form a field and algebraic integers form a ring. You do not need Galois theory anywhere in this chapter's corpus. Degree arguments here almost always do the job "this number is irrational because its minimal polynomial has degree greater than one".

- **WHEN** a hypothesis is about parity of integers **DO** reduce the entire problem to $\mathbb{F}_2$ on the first line, where squaring is the identity and quadratic conditions become linear **BECAUSE** $x^2 = x$ in $\mathbb{F}_2$ is Frobenius, and it collapses a quadratic form to a linear form (Putnam 2011 A4).
- **WHEN** you see a p-th power, a characteristic p, or an order condition $M^{p^m} = I$ **DO** apply Frobenius: expand $(a + b)^{p^m} = a^{p^m} + b^{p^m}$, or rewrite the order condition as $(M - I)^{p^m} = 0$ **BECAUSE** the binomial coefficients $\binom{p}{i}$ vanish modulo p, so the p-th power map is additive (Putnam 2012 A5, 2019 A5).
- **WHEN** a full geometric sum $1 + x + \cdots + x^{p-1}$ appears over $\mathbb{F}_p$ **DO** replace it by $(x - 1)^{p-1}$ **BECAUSE** $x^p - 1 = (x - 1)^p$ by Frobenius, and then divide (Putnam 2019 A5).
- **WHEN** a self map of a finite set of size m has a forward orbit with m distinct terms **DO** conclude it is a single m-cycle and that its order is exactly m **BECAUSE** a full orbit forces a permutation with one cycle, which is the only form in which the hypothesis can meet the group theory (Putnam 2012 A5).
- **WHEN** you need an element of a finite group whose centraliser has odd order, or a cyclic group of order $q - 1$ **DO** build the finite field $\mathbb{F}_q$ and use that $\mathbb{F}_q^\times$ is cyclic **BECAUSE** existence of $\mathbb{F}_q$ and cyclicity of its unit group are the two facts that make such constructions possible at all (Putnam 2014 B5).
- **WHEN** a problem has exactly three states or three colours with a "two the same, one different" condition **DO** identify the states with $\mathbb{F}_3$ and use that $u + v + w = 0$ exactly when all equal or all distinct **BECAUSE** the condition becomes one linear inequation per constraint, so the whole problem becomes linear algebra (Putnam 2017 A6).

2.12 The recursive and self-similar shapes in this field

Recursion is a standing order in this book, so here are the recursive and self-similar structures that the algebra items actually contain, stated in general form with the recognition cue attached.

1. Geometric orbit sizes and counting modulo p. A group of order p^k acting on a finite set has orbit sizes drawn from the geometric family $1, p, p^2, \ldots, p^k$. The general method: to compute $|X|$ modulo p, discard every orbit except the singletons. The general invariant: $|X| \equiv |X^G| \pmod{p}$. Recognise the family in the wild whenever a count is only required modulo a prime, or a divisibility of a count is requested. Two specialisations to bank: the cyclic shift on p-tuples with product e gives Cauchy's theorem (Putnam 2007 A5); the same shift on strings of length p over an alphabet of size a

gives $a^p \equiv a \pmod p$, which is Fermat's little theorem. The general form with T in place of p is Putnam 2022 A3 with $T = 5$.

2. Annihilating polynomials as linear recurrences. If u satisfies a monic degree k relation over R, then $u^k = \sum_{i<k} c_i u^i$ is a linear recurrence of order k on the sequence of powers $u^0, u^1, u^2, \ldots$, whose characteristic polynomial is the annihilating polynomial itself. Iterating it reduces every power u^N with $N \geq k$ into the span of $1, u, \ldots, u^{k-1}$: that is exactly the statement that $R[u]$ is a finitely generated R-module. The recognition cue is "a quantity is one of finitely many things but you do not know which". Putnam 2007 B5 is the case $R = \mathbb{Q}[n]$.

3. Iterating Frobenius. $(a+b)^p = a^p + b^p$ iterates to $(a+b)^{p^m} = a^{p^m} + b^{p^m}$, and on matrices $M^{p^m} = (I+N)^{p^m} = I + N^{p^m}$. The self-similarity is the point: each application of Frobenius is the same map, so a condition at level p^m reduces to nilpotency at level m, and the recursion terminates as soon as p^m exceeds the nilpotency index. That termination is what produces the bound $\operatorname{ord}(G) \leq p^{\lceil \log_p(n+1) \rceil}$ in Putnam 2012 A5, and it is why only tiny n survive there.

4. Descending chains of subgroups. A game or a process that adds one constraint at a time produces $Z(S_{t+1}) = Z(S_t) \cap Z(g_{t+1})$, a descending chain of subgroups. The general method is not to iterate the chain, which is hopeless, but to find one monotone invariant of it, then identify the earliest move that commits that invariant. In Putnam 2014 B5 the invariant is the parity of the order and the committing move is the choice of a central involution or of a Singer element. Recognise the family whenever a two player game is played on an algebraic structure.

5. Partition by generated cyclic subgroup, and $\sum \varphi(d)$. Grouping the elements of a finite abelian group by the cyclic subgroup they generate turns any sum $\sum_g F(\operatorname{ord} g)$ into $\sum_d \varphi(d) F(d) c_d$. The base identity is $\sum_{d \mid n} \varphi(d) = n$. This is the recursive skeleton behind Putnam 2009 A5, and it is also what makes the cyclicity proof for $\mathbb{F}_q^\times$ work: both are the same counting recursion over the divisor lattice.

6. Powers on cycles. A step d walk on a cycle of length L splits it into $\gcd(d, L)$ cycles of length $L/\gcd(d, L)$. Iterating this is the self-similar structure of $\sigma \mapsto \sigma^d$ on S_N, and it makes "is a d-th power" a pure cycle type condition at every level (Putnam 2023 B5).

7. Doubling and binary construction. If each greedy step at least doubles a set, $\log_2 |G|$ steps suffice; to climb a cyclic quotient of order p, appending $x, x^2, x^4, \ldots$ does it in $\log_2 p$ steps. The invariant is the size of the reachable set and the recursion is $|H_{t+1}| \geq 2|H_t|$ (Putnam 2008 A6).

8. Closure systems as fixed points of a monotone operator. A class defined by generators and a finite list of closure operations is the least fixed point of "apply the operations once". Two consequences with opposite uses: to prove a target lies in the class, exhibit an explicit finite chain of applications, never an indirect existence argument (Putnam 2012 B1, where fg must be built by hand from $e^x - 1$, $\ln(1+x)$ and the three rules); to prove something lies outside, find a property holding for the generators and preserved by every operation. When the ambient set is finite the chain $U_1 \subseteq U_2 \subseteq \cdots$ stabilises within $|G|$ steps, which is where the bound in Putnam 2016 A5 comes from, and the same increasing-union argument gives Putnam 2009 B1.

9. Periodic maps and rotation multipliers. A rational recurrence that turns out to have finite period T is a $\mathbb{Z}/T$ action in disguise, and the right response is always to count orbits rather than solutions (Putnam 2022 A3; the Lyness recursion $u_{n+2} = (1 + u_{n+1})/u_n$ has period 5). When a cyclic rotation instead multiplies a quantity by a scalar, you have a multiplier system: $F(gx) = \lambda(g, x)^{-1} F(x)$ with $F = \sum_g \lambda$, and the reciprocals sum to 1 along every orbit (Putnam 2020 B4). Both are the same idea, that a finite order symmetry imposes a linear relation on the values of any function you build from it.

- **WHEN** a recurrence over a finite ring is claimed to have finitely many solutions and you must count them modulo T **DO** check for periodicity of period T, then let $\mathbb{Z}/T$ act and count only constant solutions **BECAUSE** periodicity is a group action and the fixed points are the constants (Putnam 2022 A3).
- **WHEN** a chain of subgroups, centralisers or reachable sets is generated by a process **DO** do

not iterate it: find a monotone invariant and the move that commits it **BECAUSE** the chain has at most $\log_2 |G|$ strict steps, so the invariant, not the chain, is the object of interest (Putnam 2014 B5, 2008 A6).

- **WHEN** a class is defined as "the smallest set containing these and closed under those" **DO** to include a target, build an explicit finite chain; to exclude one, find an invariant preserved by every operation **BECAUSE** a generated class is the least fixed point of a monotone operator, so membership is a construction and non-membership is an invariant (Putnam 2012 B1).

2.13 The circuit bank, in priority order

Install in this order, because this is the order in which the Putnam actually asks.

First, and without hesitation: the fixed point lemma for p-groups and the shift action on tuples (Putnam 2007 A5, 2022 A3, 2017 B6); Lagrange as a divisibility engine including the order of an element (2007 A5, 2014 B5, 2016 A5); Frobenius in characteristic p (2011 A4, 2012 A5, 2019 A5); cycle type and sign in S_N (2012 B6, 2023 B5, 2016 B4); homomorphisms as transport, kernels as fibres, and $\mathbb{Z}^n$ free (2019 B6); orbit stabiliser and free actions (2017 B6, 2020 B4).

Second: the structure theorem for finite abelian groups and the $\mathbb{F}_p$-vector space $G[p]$ (2009 A5); characters and orthogonality (2021 A5, 2011 A6); existence of $\mathbb{F}_{p^n}$ and cyclicity of $\mathbb{F}_{p^n}^{\times}$ (2014 B5); quotient rings as the home of divisibility statements (2024 A2); generation, presentations and von Dyck (2016 A5, 2018 A4); integrality and annihilating polynomials (2007 B5); finiteness upgrading closure to a subgroup (2009 B1, 2016 A5, 2022 B2).

Third, and only after the above are automatic: Sylow congruences, which appear as a recognition device rather than a tool (2007 A5); invariant factors as the hidden object behind a uniqueness claim (2008 A3); complete mappings of cyclic groups, that is, a bijection σ with $x \mapsto x\sigma(x)$ also bijective, which exists for $\mathbb{Z}/n$ exactly when n is odd; this settled Putnam 2024 B1, where summing $i + \sigma(i)$ over all i forces $k = (n+1)/2$ and n odd, and the doubling map supplies σ.

Learn these last, because each appeared once. The quaternion description of the cross product and the Lie algebra $\mathfrak{so}(3)$ (Putnam 2022 B2, 2010 A5). The Laurent phenomenon from cluster algebras, of which $u_n u_{n-2} = u_{n-1}^2 - 1$ is a toy instance (Putnam 2017 A2); it explains why such recursions keep producing integers, and you will never need to prove it. Frobenius's theorem on the number of solutions of $x^t = e$, cited only as a sanity check (Putnam 2007 A5). Rotation invariants of polynomials in the plane (Putnam 2009 B4). Affine groups over finite fields and their Sylow structure (Putnam 2012 A5).

Two closing instructions. When a problem names a group, a ring or a field, the intended solution is almost never a computation inside it; it is a structural statement about it, and the available structural statements are exactly the ones in this chapter. When a problem names none of them, ask which one is hiding: a count modulo a prime is a p-group action, a uniqueness claim is an invariant factor, a permutation defined by a formula is a cyclic group in disguise, and a parity hypothesis is $\mathbb{F}_2$. That second habit is what separates solving two of these a paper from solving five.

Chapter 3

Combinatorics, counting and discrete structures

3.1 What this chapter is for

A-level gave you $\binom{n}{r}$ as a calculator button, the binomial theorem, series, the method of differences, and induction. That is your entire discrete inheritance. It contains no pigeonhole principle, no inclusion and exclusion, no bijective proof, no graph, no partition, no poset, no game, no invariant, and Further Maths adds none of them.

Now look at what the Putnam asks. Across the 2006 to 2025 papers the mathematicians who dismantled every problem for this archive tagged 205 separate pieces of required knowledge as combinatorics or discrete structure, the second largest field in the archive. It is also the field where the gap is widest, because the others start from something: calculus starts from A-level calculus, algebra from A-level algebra, combinatorics from nothing.

The compensation is that combinatorics is the most circuit-friendly field on the paper. A combinatorics problem is almost always a translation problem: the answer is short once the objects are named correctly, and naming them correctly is a finite list of moves. That list is this chapter.

At the end you will count a finite set by deciding which of four principles applies, knowing that all four are statements about maps; convert a count into a recursion and a recursion into a closed form or a transfer matrix; recognise Catalan, Fibonacci, Stirling of both kinds, Bell, derangement and Euler zigzag from their recursions rather than their first terms; use pigeonhole in four forms; run a bound plus construction argument to full marks instead of half; and translate a matrix into a bipartite graph, an array into a poset, a process into a monovariant, a tiling into a weighting and a game into a pairing.

One warning first. Nearly every problem below has a nominal solution which is a brute-force count or a case bash, and an ideal solution which is one bijection, one double count or one invariant. The bifurcation point is always the same line: the moment you name the objects. Name them as "the ways of doing the thing" and you grind. Name them as functions, words, lattice paths or permutations and the circuit fires.

3.2 The four counting principles

What counting is

Definition 3.2.1. A finite set X has cardinality n if there is a bijection $X \to \{1, \dots, n\}$.

Hold on to that. Every counting principle below is a statement about a map, because cardinality is defined by a map. So when you are stuck on a count, the question is never "how many are there"

but "what is the map". That is the same move as the circle: the centre falls out because a circle is defined as the points equidistant from it.

Proposition 3.2.2 (The four principles)**.** *Let X be finite.*

1. Addition. *If X is partitioned into disjoint blocks $X_1, \dots, X_k$ then $|X| = \sum_i |X_i|$.*
2. Multiplication. *If an element of X is built by k successive decisions and, whatever the earlier decisions were, decision i has exactly n_i options, then $|X| = n_1 \cdots n_k$.*
3. Bijection. *A bijection $X \to Y$ gives $|X| = |Y|$.*
4. Division. *If $f : X \to Y$ is surjective with every fibre of size d then $|X| = d|Y|$.*

Proof. Addition and bijection are the definition. Multiplication is induction on k using addition on the blocks indexed by the first decision, and the clause "whatever the earlier decisions were" is exactly the requirement that those blocks have equal size. Division is addition applied to the fibres. □

Marks are lost on multiplication when the option count depends on the earlier decisions. When it does, you do not abandon multiplication: you split by the offending feature and multiply inside each block. That forced split is the seed of every recursion in Section 7.

Bijection: the workhorse

A bijective proof outranks an algebraic one because it survives generalisation. Three transfers recur throughout the evidence.

Selections are permutations. Choosing n cells of an $n \times n$ grid, no two in a row or column, is a permutation σ, the cell of row i being $(i, \sigma(i))$. That is the whole first move of Putnam 2024 B1: once σ is written the entries become a formula in i and $\sigma(i)$, and both the invariant and the construction become algebra. Without the name there is only a picture.

Words are paths are partitions. A monotone lattice path across an $a \times b$ box is a word with a letters of one kind and b of the other, and also a partition whose Young diagram fits in the box; there are $\binom{a+b}{a}$ of each. In Putnam 2023 B1 the reachable configurations are in bijection with the monotone paths cut out by the holes, so the answer is $\binom{m+n-2}{m-1}$ and nothing is ever listed.

Height sequences are step words. Tuples $(s_0, \dots, s_{2n})$ with $s_0 = s_{2n} = 0$ and $|s_j - s_{j-1}| = 1$ correspond to words $w \in \{\pm 1\}^{2n}$ with n entries $+1$, via $w_j = s_j - s_{j-1}$. Putnam 2020 B4 needs this direction because the cyclic rotation acts on the step word, not on the heights. General lesson: choose the encoding in which the symmetry you intend to exploit is visible.

Division and the complement

Division handles a deliberate overcount. The standard instance: the number of ways to split $2n$ labelled elements into n unordered pairs is

$$(2n-1)!! = \frac{(2n)!}{2^n n!},$$

since each matching arises from $2^n n!$ of the $(2n)!$ orderings. That count is the whole answer to Putnam 2016 B4 once the expectation has killed every non-matching permutation.

The complement principle is division's cheap cousin: a subset and its complement carry the same information, so describe whichever is smaller or more rigid. Putnam 2023 B1 has $(m-1)(n-1)$ coins but only $m+n-1$ holes, and the holes carry the path structure the coins hide. Putnam 2024 A4 uses $p-4$ terms out of p, so the problem lives on the four omitted elements.

- **WHEN** a configuration is described in words ("ways of placing", "arrangements such that") **DO** name a set and a map: write it as a function, a word, a path or a permutation before counting anything **BECAUSE** cardinality is defined by a bijection, so a count you cannot phrase as a map is a count you cannot manipulate
- **WHEN** the number of options at a step depends on what you did earlier **DO** split by the feature it depends on, then multiply inside each case **BECAUSE** multiplication needs constant fibres, and the forced split is the first line of the recursion
- **WHEN** the selection is "one per row, one per column" or a bijection between labelled families **DO** write σ and convert the picture into $\sum_i M_{i,\sigma(i)}$ **BECAUSE** Putnam 2024 B1 is unreachable in pictures and immediate in permutation notation
- **WHEN** a condition constrains a large subset of a small ground set **DO** transport it to the complement **BECAUSE** four omitted elements are checkable by hand, $p-4$ present ones are not (Putnam 2024 A4, Putnam 2023 B1)
- **WHEN** you counted ordered objects and the problem wants unordered **DO** divide by the symmetry group, having first checked every fibre has that size **BECAUSE** division is only valid with constant fibres; $(2n)!/(2^n n!)$ in Putnam 2016 B4 is that check carried out

Two multiplicative disguises

Multinomial placement. Once you fix which of n ordered positions each labelled group occupies, the fillings number the product of the factorials of the group sizes. Putnam 2007 A3 is this exactly: choose which k of the $3k$ available positions carry the multiples of 3 in $\binom{3k}{k}$ ways (one position is excluded in advance, which is why it is $\binom{3k}{k}$), then fill in $k!(k+1)!k!$ ways.

Tuple spaces with one relation. In a group G of order m the p-tuples with product the identity number exactly m^{p-1}: the first $p-1$ coordinates are free and the last is forced. Putnam 2007 A5 uses nothing else about G. A set of exactly known size inside an unknown structure is a set you can run a divisibility or group action argument on.

- **WHEN** a probability over a uniform random labelled object **DO** count favourable, divide by total, and keep both in factorial form until the very end **BECAUSE** the simplification in Putnam 2007 A3 is one cancellation if postponed and a mess if performed early; ink costs time
- **WHEN** you need a set of known size inside an unknown algebraic structure **DO** impose one relation on a tuple and let the last coordinate absorb it **BECAUSE** one relation is the cheapest way to buy an exact cardinality (Putnam 2007 A5)

3.3 Double counting and finite Fubini

Theorem 3.3.1 (Finite Fubini)**.** *For finite F, K and $P \subseteq F \times K$,*

$$\sum_{f \in F} \#\{k : (f,k) \in P\} = |P| = \sum_{k \in K} \#\{f : (f,k) \in P\}.$$

Proof. Both sides count P by addition, partitioning by the first coordinate on the left and the second on the right. □

That is the whole theorem and one of the three highest-yield tools in the chapter. One summation order is usually cheap and the other is the one the problem hands you, so the circuit is: build the incidence set, sum the other way.

Example 3.3.2 (Putnam 2013 A1)**.** Twenty faces of an icosahedron carry nonnegative integers totalling 39; show two faces sharing a vertex are equal. Negate: at each of the 12 vertices the five surrounding faces are distinct nonnegative integers, so they total at least $0+1+2+3+4=10$, and the grand total of all vertex sums is at least 120. Sum the other way: each face lies at exactly 3 vertices, so the same grand total is $3\times 39=117$. Contradiction.

Two reusable sub-lemmas fall out. Any k distinct nonnegative integers total at least $\binom{k}{2}$, with equality only for $\{0,\dots,k-1\}$. And the general threshold: if n labelled objects lie in m groups of size t, each object in exactly d groups, with labels totalling T, then two objects in a group share a label as soon as $T < mt(t-1)/(2d)$. For the icosahedron that threshold is 40 and you are given 39. A given number sitting exactly one below a threshold is strong evidence you have found the intended route.

Handshake, incidences, windows

The handshake lemma, $\sum_v \deg v = 2|E|$, is finite Fubini on vertex-edge incidence; you will use it inside Euler's formula (Putnam 2007 A6) and whenever a degree hypothesis must meet an edge count (Putnam 2014 B3).

Install this as a formula. For $F:\mathbb{Z}/m\mathbb{Z}\to\mathbb{R}$ and window length L, with $S_L(p)=\sum_{t=0}^{L-1}F(p+t)$,

$$\sum_p S_L(p) = L\sum_p F(p),$$

because each value is picked up by exactly L windows. Putnam 2013 A4 is built on it with F the indicator of a zero digit around a circle: it supplies the exact mean Lq/m, the only global invariant the circle has.

Averaging is double counting

An average is a double count divided by a count, and the use is always: some object is at least as good as the average.

Example 3.3.3 (Putnam 2008 A6)**.** For $H\subseteq G$ with $|H|=m$, $|G|=n$, double counting gives $\sum_g |H\cup Hg| = 2mn-m^2$, so some g achieves $|H\cup Hg|\geq 2m-m^2/n$. In terms of the deficiency $s=1-m/n$ the step is $s\mapsto s^2$, so after k rounds $s\leq(1-1/n)^{2^k}$ and $k=\lfloor 2\log_2 n\rfloor$ forces $H=G$. That is a doubling argument needing no disjointness hypothesis at all.

Example 3.3.4 (Putnam 2025 B4)**.** To prove $\sum_P a\leq c|P|$ it suffices to find m pointwise bounds b_t with $a\leq b_t$ and $\sum_t b_t = mc$ pointwise. In B4 the bounds $(i+j-n)$, $(n+1-j)$, $(n+1-i)$ add to $n+2$ everywhere, giving $3S\leq(n+2)N$. The denominator 3 in the target counts the bounds averaged, and that is how to read any target with a small integer denominator.

A caveat worth clarifying, because careless proofs die here. The row bound $j-Z_i$ depends on its row and the substitute $n+1-j$ does not; they agree only after summing along a whole line, since $\sum_{j=Z+1}^{n}(j-Z)=\sum_{j=Z+1}^{n}(n+1-j)$ relabels an arithmetic run from the other end. Pointwise the substitution is false. Sum first, then substitute.

Group actions

Proposition 3.3.5. *If a finite group G acts on a finite set X and $f:X\to\mathbb{R}$, then $\sum_x\sum_g f(gx) = |G|\sum_x f(x)$.*

Proof. For each fixed g, $x\mapsto gx$ is a bijection of X, so the inner sum is $\sum_x f(x)$ for every g. □

No orbit lengths, no stabilisers, no short-orbit cases. Putnam 2020 B4 uses exactly this to turn a per-cycle identity into a global average in one line; the temptation to classify rotation orbits on words of length $2n$ is a page of work for nothing.

When you must count orbits rather than average, use Burnside: the number of orbits is $\frac{1}{|G|}\sum_g |X^g|$, proved by double counting the pairs (g, x) with $gx = x$ and applying orbit-stabiliser. Putnam 2022 A3 is such a count. Putnam 2017 A6 is the trap in the other direction: the edges there are explicitly labelled, so colourings are functions and are counted as functions. The sentence that labels the objects is the sentence telling you not to divide by the symmetry group. Read it.

- **WHEN** a quantity is described from the point of view of one family (vertices, rows, players, positions) **DO** build the incidence set and sum from the other family **BECAUSE** Putnam 2013 A1 is a contradiction between the vertex sum and the face sum of one total, and neither alone says anything
- **WHEN** the words "average" or "expected" appear, or the target has denominator 2 or 3 **DO** find that many pointwise bounds with constant sum, or double count and divide **BECAUSE** the denominator counts the bounds averaged (Putnam 2025 B4); some object beats the mean (Putnam 2008 A6)
- **WHEN** a statistic is summed over all rotations, translations or windows **DO** use $\sum_p S_L(p) = L\sum_p F(p)$ or $\sum_x \sum_g f(gx) = |G|\sum_x f(x)$ **BECAUSE** every value is counted equally often, so no orbit analysis is needed (Putnam 2013 A4, Putnam 2020 B4)
- **WHEN** objects are counted up to a symmetry **DO** use Burnside, but first check whether the problem labelled the objects **BECAUSE** labelled objects are functions and must not be divided by $|G|$ (Putnam 2017 A6); unlabelled ones are orbits (Putnam 2022 A3)

3.4 Binomial and multinomial coefficients

Definition 3.4.1. $\binom{n}{k}$ counts the k-subsets of an n-set and equals $n!/(k!(n-k)!)$. For indeterminate z, $\binom{z}{k} = z(z-1)\cdots(z-k+1)/k!$ is a polynomial of degree k. The multinomial $\binom{n}{k_1,\ldots,k_r} = n!/(k_1!\cdots k_r!)$ counts distributions of n labelled objects into labelled groups of those sizes.

The polynomial extension matters: it lets you locate roots, which is what Putnam 2024 B5 needs.

Proposition 3.4.2 (Pascal, committee and chair, hockey stick, Vandermonde)**.**

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}, \tag{3.1}$$

$$k\binom{n}{k} = n\binom{n-1}{k-1}, \tag{3.2}$$

$$\sum_{j=0}^{k-1} \binom{j}{i} = \binom{k}{i+1}, \tag{3.3}$$

$$\sum_k \binom{m}{k}\binom{n}{r-k} = \binom{m+n}{r}. \tag{3.4}$$

Proof. (3.1): split the k-subsets by whether they contain n. (3.2): both sides count pairs (a k-subset, a distinguished element of it), choosing the subset first on the left and the chair first on the right. (3.3): telescope $\binom{j+1}{i+1} - \binom{j}{i+1} = \binom{j}{i}$, which is (3.1). (3.4): choose r elements from a ground set split into blocks of sizes m and n, summing over how many came from the first. □

Each is a named move with a named cue. Pascal is the split-by-the-last-element move and the source of every subset recursion; Putnam 2018 A2 is the extreme case, where the nonempty subsets of $[n]$ split as $(2^{n-1}-1)+2^{n-1}$ and that split is the block structure of the matrix. Committee and

chair removes a stray linear factor: in Putnam 2013 B5 a factor j in front of $\binom{p}{j}$ blocks the binomial theorem and (3.2) clears it. Hockey stick is summation in the binomial basis, the discrete analogue of $\int t^i\,dt$, and it closes Putnam 2012 A5; install the habit of rewriting a polynomial in the binomial basis before summing it over consecutive integers, never expanding powers and quoting Faulhaber, since $\sum_{z\le n} z^4 = (6n^5+15n^4+10n^3-n)/30$ has a negative coefficient and destroys any positivity you were proving (the warning in Putnam 2024 B5). Vandermonde is the split-the-ground-set move, usually disguised as a sum of two binomials whose indices add to a constant.

The binomial theorem as a target shape

$$\sum_{i=0}^{m}\binom{m}{i}q^i p^{m-i} = (p+q)^m.$$

Read it as a target, not a fact. Any sum of a binomial against two powers with exponents adding to a constant should be steered into this shape, and the steering moves are: clear linear factors with (3.2), shift the index so the lower index matches the exponent, extend or truncate using vanishing terms. Putnam 2013 B5 closes in one line this way.

Stars and bars

Proposition 3.4.3. *The weakly increasing sequences $1 \le u_1 \le \cdots \le u_j \le z$, equivalently the multisets of size j from z types, equivalently the nonnegative solutions of $a_1+\cdots+a_z = j$, number*

$$\binom{z+j-1}{j} = R(z,j) = \frac{z(z+1)\cdots(z+j-1)}{j!}.$$

Compositions of n into exactly k positive parts number $\binom{n-1}{k-1}$.

Proof. Encode a multiset as j stars split by $z-1$ bars: the row is a word of length $z+j-1$ and the multiset is the choice of bar positions. For compositions place $k-1$ bars in the $n-1$ gaps. □

$R(z,j)$ satisfies $R(z,j)-R(z-1,j) = R(z,j-1)$ and $\sum_{z\le n} R(z,j) = R(n,j+1)$, which are Pascal and hockey stick in disguise. What makes it the right object in Putnam 2024 B5 is not the count but the factorisation: as a polynomial in z it vanishes at $0,-1,\dots,-(j-1)$, and knowing where a counting polynomial vanishes is what proves positivity of coefficients by root location instead of expansion.

The honest boundary: sequences of k positive parts summing to n with parts in a set A number $[x^n](\sum_{j\in A} x^j)^k$ and have no closed form. Putnam 2018 B6 is where the unrestricted count $\binom{3859}{2017} \approx 10^{1158}$ overshoots the target $\approx 10^{1149}$ by nine orders of magnitude, and that overshoot is the evidence that only a structural bound can work.

Ratios, and the snake oil method

When two binomial coefficients have indices differing by bounded amounts, their ratio is a rational function after cancellation: for $m > n+1$, $\binom{m}{n-1}/\binom{m-1}{n} = mn/((m-n+1)(m-n))$, so an inequality between them cross-multiplies into a quadratic. That is the reduction in Putnam 2016 A2. In product form, $n!/i! = \prod_{j=i+1}^{n} j$, so a product of factorial ratios regroups by counting how many admissible i each factor j survives: $\prod_{i=1}^{n} n!/i! = \prod_{j=2}^{n} j^{j-1}$, the hyperfactorial shape Putnam 2007 B6 estimates.

Proposition 3.4.4 (Snake oil)**.** *To evaluate $a_n = \sum_k f(n,k)$, form $A(x) = \sum_n a_n x^n$, exchange the summations, perform the inner sum over n in closed form, and read off the coefficient.*

It works because the sum over n is often geometric or binomial even when the sum over k is not.

Example 3.4.5. Evaluate $a_n = \sum_k \binom{k}{n-k}$. Exchanging,

$$A(x) = \sum_{k\geq 0}\sum_n \binom{k}{n-k}x^n = \sum_{k\geq 0} x^k \sum_{m\geq 0}\binom{k}{m}x^m = \sum_{k\geq 0}\big(x(1+x)\big)^k = \frac{1}{1-x-x^2},$$

so $a_n = F_{n+1}$. Bijectively, $\binom{k}{n-k}$ counts compositions of n into k parts of size 1 or 2, and F_{n+1} counts square-and-domino tilings of a $1 \times n$ strip. Putnam 2017 A2 lives on that identification.

The tiling reading also gives the identities behind Putnam 2019 B5,

$$\sum_{k=0}^{M}\binom{M}{k}F_k = F_{2M}, \qquad \sum_{k=0}^{M}\binom{M}{k}F_{k+1} = F_{2M+1},$$

both from $\sum_k \binom{M}{k}\gamma^k = (1+\gamma)^M = \gamma^{2M}$ at the two roots of $t^2 = t+1$, since $1+\gamma = \gamma^2$. In B5 the extrapolated sum falls exactly one term short of F_{2M}, which produces both terms of the answer at once.

- **WHEN** a binomial sum carries a stray factor k in front of $\binom{n}{k}$ **DO** apply $k\binom{n}{k} = n\binom{n-1}{k-1}$ and reindex **BECAUSE** the binomial theorem cannot fire while a linear factor sits in the way (Putnam 2013 B5)
- **WHEN** you must sum a polynomial over consecutive integers **DO** rewrite it in the basis $\binom{z}{i}$ and apply hockey stick **BECAUSE** binomial-basis summation is one index shift; the power basis gives Faulhaber and destroys sign information (Putnam 2012 A5, Putnam 2024 B5)
- **WHEN** two binomials appear whose indices add to a constant **DO** recognise Vandermonde and read it as splitting one ground set **BECAUSE** it is a single choice made in two stages
- **WHEN** a sum over k has no telescoping and no combinatorial reading **DO** snake oil: multiply by x^n, sum over n first, identify the closed form **BECAUSE** the outer variable often makes a geometric series where the inner one makes nothing
- **WHEN** an inequality between binomial coefficients with nearby indices **DO** cancel factorials into a rational function and cross-multiply **BECAUSE** the result is a polynomial inequality of bounded degree (Putnam 2016 A2)
- **WHEN** you count nonnegative solutions of $a_1 + \cdots + a_z = j$, or weakly increasing sequences **DO** stars and bars, $\binom{z+j-1}{j}$ **BECAUSE** a multiset is the bar positions in a longer word

3.5 Inclusion and exclusion

Theorem 3.5.1. *For finite $A_1, \ldots, A_n$ inside U,*

$$\Big|\bigcup_i A_i\Big| = \sum_{\emptyset\neq S}(-1)^{|S|+1}\Big|\bigcap_{i\in S}A_i\Big|, \qquad \Big|U\setminus\bigcup_i A_i\Big| = \sum_S(-1)^{|S|}\Big|\bigcap_{i\in S}A_i\Big|.$$

Proof. Fix x lying in exactly m of the sets. On the right of the second identity it contributes $\sum_j(-1)^j\binom{m}{j} = (1-1)^m$, which is 1 for $m = 0$ and 0 otherwise, the indicator of the left side. □

The proof is the technique: the sieve is the binomial theorem applied to an indicator. Which tells you when it is cheap, namely when $|A_S|$ depends only on $|S|$, since then the sum collapses to one binomial sum.

Example 3.5.2 (Derangements; Putnam 2020 A4, Putnam 2023 A6)**.** With A_i the permutations of $[m]$ fixing i, $|A_S| = (m-|S|)!$, so

$$D_m = \sum_j(-1)^j\binom{m}{j}(m-j)! = m!\sum_{j=0}^{m}\frac{(-1)^j}{j!},$$

and $D_m/m! \to 1/e$. The permutations with exactly j fixed points number $\binom{m}{j}D_{m-j}$, so the fixed point count tends to Poisson of mean one. In Putnam 2020 A4 the quantity $w(N)$ is $D_{N+1}/N!$, so $w(N)/(N+1)$ is a derangement probability and the limit $1/e$ is forced rather than fitted.

Two recursions to be able to produce without the sieve: $D_m = mD_{m-1} + (-1)^m$ and $D_m = (m-1)(D_{m-1} + D_{m-2})$. The second reads combinatorially: the image of 1 is one of $m-1$ letters j, and then either $j \mapsto 1$ (a derangement of the other $m-2$) or $j \not\mapsto 1$ (a derangement of $m-1$ after identifying the forbidden value).

Three further sieve applications you must recognise. *Surjections*: $[n] \to [k]$ onto number $\sum_j (-1)^j \binom{k}{j}(k-j)^n = k!\, S(n,k)$, same shape with A_j the maps missing j. *Descent sets*: the permutations of $[n]$ with descent set contained in $D = \{d_1 < \cdots < d_k\}$ number $n!/(d_1!(d_2-d_1)!\cdots(n-d_k)!)$, since the permutation is determined by which values fill each block, and sieving over subsets of D converts "contained in" to "exactly" (MacMahon). Putnam 2025 A5 is stated in that language, and the sieve is its nominal route; the ideal route is the peak recursion of Section 7, but you must know the sieve version to see what it costs. *Abstract sieve*: Putnam 2013 B3 imposes $f(S \cup S') = f(S) + f(S') - f(S \cap S')$ as an axiom on an abstract real-valued f. That is the two-set case promoted to a hypothesis, and a function satisfying it is a valuation or modular function; naming it selects the right toolkit (Section 12) instead of trying to prove f is a measure.

The general form replaces subsets of $[n]$ by a finite poset and $(-1)^{|S|}$ by that poset's Mobius function. Putnam 2017 B6 needs the partition lattice: the events $x_i = x_j$ have intersections indexed by set partitions, and comparing degree-one coefficients in $x(x-1)\cdots(x-n+1) = \sum_\pi \mu_\pi x^{|\pi|}$ gives μ of the one-block partition as $(-1)^{n-1}(n-1)!$, which for $n = 64$ is $-63!$, exactly the bias $2016 \cdot 63!$ in the answer. That is the deepest sieve on the paper in twenty years and it appeared once. Learn it last in this section.

- **WHEN** a count is "none of these bad events" or "at least one of them" **DO** sieve, checking first whether $|A_S|$ depends only on $|S|$ **BECAUSE** if it does the whole computation is one binomial sum
- **WHEN** fixed points, forbidden positions, "no letter in its own envelope" **DO** write $D_m = m!\sum_j (-1)^j/j!$ and expect $1/e$ **BECAUSE** the sieve over the m fixing conditions gives it immediately (Putnam 2020 A4)
- **WHEN** a problem hypothesises $f(A \cup B) + f(A \cap B) = f(A) + f(B)$ **DO** call it a valuation and prove it is determined by its singleton values by induction along covers **BECAUSE** the hypothesis is the two-set sieve as an axiom (Putnam 2013 B3)
- **WHEN** a statistic is naturally "contained in D" but you want "exactly D" **DO** sieve over subsets of D **BECAUSE** containment counts are multinomials and exactness counts are not (Putnam 2025 A5)

3.6 Pigeonhole in its four forms

Theorem 3.6.1. *Let $f : X \to Y$ with X, Y finite.*

1. *If $|X| > |Y|$ some fibre has at least two elements.*
2. Strong form. *Some fibre has at least $\lceil |X|/|Y| \rceil$ elements and some has at most $\lfloor |X|/|Y| \rfloor$.*
3. Capacity form. *If $x_1, \ldots, x_k \geq 0$ are integers with $\sum x_i > \sum c_i$ then $x_i > c_i$ for some i.*
4. Infinite form. *If X is infinite and Y finite, some fibre is infinite.*

Proof. All four are the contrapositive of addition: bounding every fibre bounds $\sum_y |f^{-1}(y)| = |X|$. □

The strong form is the one that gets forgotten and the one that produces exact thresholds. The capacity form is the same statement for boxes with individually prescribed limits, which is the shape of Putnam 2010 B3: each box sits below its own label threshold, and one inequality supplies both the frozen counterexample below the threshold and the guarantee that some box is eligible above it. A pointed variant: if n objects go into k nonempty boxes then at most $n-k$ boxes can hold two or more, so at least $2k-n$ boxes are singletons, a positive count as soon as $n < 2k$ (Putnam 2010 A1).

Proposition 3.6.2 (Continuous form)**.** *If $N+1$ points lie in an interval of length ℓ, two differ by less than ℓ/N.*

Proof. Cut into N pieces of length ℓ/N and apply the finite form. □

Applied to the fractional parts of $0, \alpha, \ldots, N\alpha$ this is Dirichlet's approximation theorem. Putnam 2011 A5 is the pure form: it produces $i \neq j$ with $|G(x_i) - G(x_j)| < \ell/n$, and the factor $1/n$ in the problem's conclusion is exactly that denominator. When a Putnam conclusion carries $1/n$ with n the number of objects supplied, continuous pigeonhole fires first.

Proposition 3.6.3 (State space form)**.** *A deterministic process on a finite state space must repeat a state, so it is eventually periodic; if the transition map is a bijection the orbit is purely periodic.*

Proof. The first $|S|+1$ states cannot be distinct. Given a repeat $v_{n_1} = v_{n_2}$, apply the inverse n_1 times to slide it back to the start, giving $v_0 = v_{n_2-n_1}$. □

That last sentence is the content of Putnam 2018 B4, where the problem says periodic and not eventually periodic, so you must record that the Fibonacci pair map is invertible modulo d. Pigeonhole alone gives the shape of the letter rho, a tail plus a cycle; invertibility deletes the tail. Putnam 2006 A3 is the same lemma on windows of n consecutive residues modulo N, of which there are N^n.

Two specialisations that are cues rather than facts. *Partial products*: if $s = t_1 \cdots t_r$ is a shortest representation and $r > |G|$, two partial products coincide and the segment between them deletes, contradicting minimality; that is the alternative route to the bound in Putnam 2016 A5, and the index range matters because the shortened word must stay nonempty. *Cyclic pigeonhole*: if a cyclic sequence has strictly more marks of type L than of type H, two Ls are cyclically adjacent, since otherwise each L is followed by an H. Putnam 2007 A6 uses that to produce the run of boundary degrees $3, 4, \ldots, 4, 3$ which then gets peeled.

The fourth form is a proof shape rather than a count: to show a quantity is at least m, assume it is at most $m-1$, use a crude total to manufacture one explicit exceptional object, and extract m new facts from that object's behaviour. Putnam 2012 B3 runs on this: the counting step produces one team, and that team alone generates the m distinct winners. Naming the shape tells you where to spend effort, on the witness rather than on the global count.

- **WHEN** more objects than categories, or a total exceeding a sum of caps **DO** pigeonhole, using the strong form to get the exact threshold rather than mere existence **BECAUSE** $\lceil |X|/|Y| \rceil$ is free once you have the fibre decomposition (Putnam 2010 A1, Putnam 2010 B3)
- **WHEN** the conclusion contains $1/n$ where n objects are given **DO** cut the range into n or $n-1$ equal pieces and apply continuous pigeonhole **BECAUSE** Dirichlet's mechanism puts the object count into the denominator (Putnam 2011 A5)
- **WHEN** a deterministic process on a finite state space **DO** conclude eventual periodicity, then check invertibility if the problem said "periodic" **BECAUSE** invertibility removes the tail of the rho (Putnam 2018 B4, Putnam 2006 A3)

- **WHEN** a minimal-length representation is assumed **DO** look at partial products and delete a repeated segment **BECAUSE** minimality contradicts any repeat, so the length is bounded by the state count (Putnam 2016 A5)
- **WHEN** a cyclic arrangement has strictly more of one type than the other **DO** two of the majority type are cyclically adjacent **BECAUSE** otherwise the minority would separate them all and be at least as numerous (Putnam 2007 A6)

3.7 Recursions and what they mean

A recursion is a bijection you have not written down

Every combinatorial recursion here arises the same way: choose a canonical feature (the largest element, the last letter, the first return to a level, whether the new ground-set element is used), split the family by it, and observe that each block is in bijection with a product of smaller families. The recursion is then addition and multiplication applied to that split. So when you meet a recursion, ask what the split was, because the split generalises and the formula does not.

This is a standing order for the rest of your preparation. Whenever a recursion appears, record (i) the split, (ii) the general form, (iii) the solution method for that form, (iv) the recognition cue. Section 15 collects the families.

Fibonacci and the linear constant-coefficient family

$F_1 = F_2 = 1$, $F_{n+1} = F_n + F_{n-1}$, and F_{n+1} counts tilings of a $1 \times n$ strip by squares and dominoes. The split is the last tile, and everything else follows, including $\sum_k \binom{k}{n-k} = F_{n+1}$ and the transforms of Putnam 2019 B5.

General family. $a_n = c_1 a_{n-1} + \cdots + c_d a_{n-d}$ is solved by the characteristic polynomial: simple roots give $a_n = \sum_j \alpha_j \lambda_j^n$ and a root of multiplicity r contributes λ^n times an arbitrary polynomial of degree $r-1$; equivalently $v_n = Tv_{n-1}$ on the last d terms, diagonalised. Putnam 2022 A3 makes it explicit: $x_{n+2} = F(x_n, x_{n+1})$ is iteration of $T(x,y) = (y, F(x,y))$ on pairs, so periods, orbits, fixed points and invariants of the sequence are those of T. That is the default reading of any order-k recursion.

Catalan and the convolution family

$C_0 = 1$, $C_{n+1} = \sum_{i=0}^{n} C_i C_{n-i}$, $C_n = \frac{1}{n+1}\binom{2n}{n}$; C_n counts Dyck paths of length $2n$, binary trees with n internal nodes, triangulations of an $(n+2)$-gon and standard Young tableaux of shape (n,n). The split is the first return to the axis: a nonempty Dyck path is an up step, a Dyck path, a down step, a Dyck path. That is where the convolution comes from and hence where the closed form comes from.

Proposition 3.7.1 (Reflection principle)**.** *Monotone paths from $(0,0)$ to (n,n) never rising above the diagonal number $\binom{2n}{n} - \binom{2n}{n-1} = C_n$.*

Proof. Reflect a bad path in the first line above the diagonal that it touches; this bijects bad paths with all paths to $(n-1, n+1)$. □

General family. An ordinary convolution $a_{n+1} = \sum_i b_i a_{n-i}$ is multiplication of ordinary generating functions, $A = 1 + xBA$ for Catalan, solved by the quadratic formula or, when nonlinear, by Lagrange inversion. The cue is "first block, then the rest, independently". Two adjacent facts: the cycle lemma of Dvoretzky and Motzkin, that if each $w_i \le 1$ and $\sum_{i=1}^{m} w_i = r > 0$ then exactly r of the m cyclic rotations have all partial sums positive, which is what to recall when a problem rotates a nearly closed path and counts good rotations (the family of Putnam 2020 B4); and the Hankel

determinant route through Lindstrom, Gessel and Viennot, an optional route in Putnam 2024 A6 that appeared once, so learn it last.

Stirling numbers of both kinds, and Bell

$S(n,k)$ counts partitions of an n-set into k nonempty unlabelled blocks, $c(n,k)$ counts permutations of $[n]$ with k cycles, and $B_n = \sum_k S(n,k)$ counts all set partitions.

Proposition 3.7.2. $S(n,k) = S(n-1,k-1) + kS(n-1,k)$; $c(n,k) = (n-1)c(n-1,k) + c(n-1,k-1)$; $B_{n+1} = \sum_k \binom{n}{k} B_k$; $x(x+1)\cdots(x+n-1) = \sum_k c(n,k)x^k$; $x^n = \sum_k S(n,k)x(x-1)\cdots(x-k+1)$.

Proof. For S: either n is alone or it joins one of k blocks. For c: either n is a fixed point, or it is inserted immediately after one of the $n-1$ other elements in some cycle. For B: split by the block containing $n+1$. The rising factorial identity follows from the c recursion by induction, since multiplying by $(x+n-1)$ sends the coefficient vector to $(n-1)$ times itself plus its own shift. The last identity is the inverse change of basis, which is why S is the tool for turning powers into falling factorials. □

Putnam 2007 B5 wants the first-kind expansion explicitly, with coefficients $c(k,i)$. Putnam 2017 A5 wants the probabilistic reading: $x(x+1)\cdots(x+n-1)/n!$ is the probability generating function of independent Bernoulli variables with parameters $1, \frac{1}{2}, \ldots, \frac{1}{n}$, so the cycle count of a uniform permutation has that distribution, mean H_n, concentrated near $\log n$; the Chinese restaurant construction, which opens a new cycle at stage j with probability $1/j$, makes the product form obvious for every n at once. Both recursions have the shape $T(n,k) = \alpha(n,k)T(n-1,k) + T(n-1,k-1)$ with α polynomial, which is what any "the new element becomes a new part or joins an old one" split gives. Do not chase a closed form; identify the change of basis it encodes.

Splitting at the maximum: binomial convolution

Proposition 3.7.3. *If a family of labelled objects on $[n]$ splits by the position i of the largest value into independent left and right structures, then*

$$F_n = \sum_{i=1}^{n} \binom{n-1}{i-1} L_{i-1} R_{n-i},$$

which for exponential generating functions $F(x) = \sum F_n x^n/n!$ reads $F' = LR$.

The binomial coefficient is the choice of which labels go left; the derivative is the removal of the maximum. This template covers much of the evidence.

Putnam 2025 A5. For any property determined by comparisons of consecutive entries, classify by the position of n. The comparisons adjacent to n are automatically up then down, so n can only sit at a peak of the sign pattern, and symmetrically 1 only at a valley. The recursion $f(s) = \sum_{\text{peaks } i} \binom{n-1}{i-1} f(\text{left}) f(\text{right})$ is uniform in s and both blocks are strictly shorter, so it drives a strong induction. On an alternating pattern it gives the Euler zigzag recursion

$$2E_n = \sum_{i=1}^{n} \binom{n-1}{i-1} E_{i-1}E_{n-i} \quad (n \geq 2), \qquad E_0 = E_1 = 1,$$

whose exponential generating function satisfies $2E'(x) = 1 + E(x)^2$, so $E(x) = \sec x + \tan x$ (Andre), giving $1, 1, 1, 2, 5, 16, 61, 272, 1385, \ldots$, the secant and tangent numbers interleaved. Niven's theorem, which A5 asks you to prove, is that the two alternating patterns are the unique maximisers of $f(s)$, with maximum E_n.

Putnam 2006 A4. The same split is the recursive route to the peak count, and it is worth knowing because it stalls: the two blocks inherit one-sided boundary rules, forcing a second coupled statistic.

That is why the ideal route there is linearity of expectation, and knowing the cost of the recursive route is what lets you rank routes in the arena.

Putnam 2013 B5. The first-entry layer peel: for disjoint Y, S with $|Y| = p$, $|S| = q \geq 1$, the maps $h : Y \to Y \cup S$ all of whose orbits reach S number $q(p+q)^{p-1}$. Condition on $T = h^{-1}(S)$, nonempty, and note h on $Y \setminus T$ is again absorbing with sink T. The recursion $F(p,q) = \sum_{j \geq 1} \binom{p}{j} q^j F(p-j, j)$, $F(0,q) = 1$, has the general shape $a_p = \sum_j \binom{p}{j} b_j a_{p-j}$, which arises whenever a labelled structure is cut into a canonical first layer plus a smaller copy of itself; finish by clearing linear factors with committee and chair, or by exponential generating functions with Lagrange inversion when nonlinear.

The exponential formula. If a_j weights connected structures on a j-set then $k![x^k] \exp(\sum_j a_j x^j / j!)$ weights disjoint unions of them, which is Putnam 2022 B1's third official route and explains why the quantity there is an integer.

Deletion, insertion and transfer systems

Proposition 3.7.4. *Suppose an element can always be deleted from a valid object of size $n+1$ to leave a valid object of size n. Then the count at level $n+1$ is the sum over level-n objects of the number of legal reinsertion slots; and if that slot count depends only on a bounded feature set F, the counts refined by F satisfy $x(n+1) = Mx(n) + b$, hence a constant-coefficient recursion with characteristic polynomial that of M.*

Putnam 2015 B5 is the archetype: deleting $n+1$ is always legal because its only possible neighbours n and $n-1$ are themselves adjacent, and the slot count depends only on the first letter, the last letter, and whether n and $n-1$ are adjacent, a finite feature set. The same protocol gives the Eulerian recursion $A(n,k) = (k+1)A(n-1,k) + (n-k)A(n-1,k-1)$ for permutations with k descents, by inserting n and asking whether the slot creates a descent (Putnam 2006 A4); in general a statistic controlled under insertion of the maximum gives $P_n(t) = \sum_\pi t^{\mathrm{st}(\pi)}$ a recursion with coefficients polynomial in t and ∂_t, for descents $P_n(t) = ntP_{n-1}(t) + t(1-t)P'_{n-1}(t)$.

The extremal version: define a restriction map from n-object configurations to $(n-1)$-object ones, bound its fibre size, and separately bound how many outputs have large fibres, giving $f(n) \leq f(n-1) + (\text{correction})$. Putnam 2006 B3 is this, with the correction supplied by a geometric lemma: a line through a fixed point P, parametrised by direction, partitions the other points in a way that changes only at the directions from P to those points, so at most $n-1$ partitions occur.

First passage, self-reference, and digit programmes

Example 3.7.5 (Putnam 2020 A2)**.** Strings of length $n+m+1$ over $\{H, T\}$ with at least $n+1$ copies of H number $\sum_{j=0}^{m} \binom{n+j}{j} 2^{m-j}$: condition on the position of the $(n+1)$st H, at place $n+j+1$ where j counts the preceding Ts, with the last $m-j$ letters free. With $m = n = k$ that is the given sum. Now the involution swapping H and T: in odd length exactly one letter has a majority, so the two classes are equinumerous and each has $2^{2k} = 4^k$ elements.

The nominal route deserves storage as a general circuit. Applying Pascal inside the sum and reindexing yields $S_k = 2S_{k-1} + \frac{1}{2}S_k$, with S_k on both sides. Any relation $u_k = \alpha u_{k-1} + \beta u_k$ with $\beta \neq 1$ rearranges to $u_k = \frac{\alpha}{1-\beta} u_{k-1}$, here ratio 4, so $S_k = 4^k$. Do not panic when the unknown reappears; solve for it.

Example 3.7.6 (Putnam 2020 A5)**.** Let $r(n)$ count index sets whose weights sum to n for an increasing weight sequence with partial sums U_m. If $u_m \leq n$ and no index above m can occur, splitting on whether index m is used gives

$$r(n) = r(n - u_m) + r(U_{m-1} - n),$$

the second term by complementing inside $\{1, \dots, m-1\}$, legitimate because $S \mapsto S^c$ is an involution carrying weight sum t to $U_{m-1} - t$. With Fibonacci weights this is the engine of A5; with $u = 9 \cdot 10^j$ it is the nines complement.

Store the complementation involution separately: if the weights total T, the subset-sum counting function is symmetric about $T/2$, and that is what keeps the A5 recursion closed instead of spawning restricted subproblems. The additive sibling is the denomination recursion $f_m(N) = f_{m-1}(N) + f_m(N - w_m)$, $f_0(N) = [N = 0]$, which is exactly the factorisation $\prod_i (1 - x^{w_i})^{-1}$ and is the general solution method for every linear Diophantine counting problem (Putnam 2007 B6).

- **WHEN** a family is indexed by n and you can delete a canonical element **DO** write the deletion map, count reinsertion slots, and check whether the count depends on boundedly many features **BECAUSE** bounded features give a transfer matrix and a constant-coefficient recursion (Putnam 2015 B5)
- **WHEN** objects are labelled and split at their largest element **DO** write $F_n = \sum_i \binom{n-1}{i-1} L_{i-1} R_{n-i}$ and pass to exponential generating functions, where it is $F' = LR$ **BECAUSE** the binomial coefficient is the label split and the derivative removes the maximum (Putnam 2025 A5, Putnam 2013 B5, Putnam 2006 A4)
- **WHEN** a sequential object has a decisive first occurrence **DO** condition on its position and let the tail be free **BECAUSE** the count becomes $\sum_j \binom{n+j}{j} (\text{free})^{m-j}$, and an alphabet involution often finishes it (Putnam 2020 A2)
- **WHEN** the unknown reappears on the right of your own recursion **DO** solve the linear equation for it **BECAUSE** $u_k = \alpha u_{k-1} + \beta u_k$ is first order in disguise with ratio $\alpha/(1-\beta)$
- **WHEN** representations by a fixed increasing weight sequence **DO** use $r(n) = r(n - u_m) + r(U_{m-1} - n)$ via the complementation involution **BECAUSE** complementation keeps the recursion inside the same family (Putnam 2020 A5)
- **WHEN** a recursion of order k appears anywhere **DO** rewrite it as a map on k-tuples and ask for that map's fixed points, invariants and periods **BECAUSE** periodicity of the sequence is finite order of the map (Putnam 2022 A3)

3.8 Permutations, cycle type, sign, and the symmetric group

Definition 3.8.1. S_n is the group of bijections $[n] \to [n]$. Every σ factors uniquely into disjoint cycles, whose multiset of lengths is the *cycle type*, a partition of n. An *inversion* is a pair $i < j$ with $\sigma(i) > \sigma(j)$; a *descent* is k with $\sigma(k) > \sigma(k+1)$; a *record* is a left-to-right maximum.

Those statistics plus the cycle type are what Putnam problems ask about. Install the standard means as free sanity checks: for uniform $\sigma \in S_n$ the expected number of descents is $(n-1)/2$, of peaks $(n+1)/3$, of records H_n, of cycles H_n, of fixed points 1.

Proposition 3.8.2. *The permutations with m_k cycles of length k number $n!/\prod_k (k^{m_k} m_k!)$, and two permutations are conjugate in S_n exactly when they share a cycle type.*

Proof. Write out the $n!$ linear orders and read each as a cycle diagram of the prescribed shape; each permutation arises $\prod_k k^{m_k} m_k!$ times, from rotating each cycle and permuting cycles of equal length. Conjugating by τ relabels the entries of every cycle, preserving lengths and realising any relabelling. □

Sign, two ways

Definition 3.8.3. With $\Delta = \prod_{i<j}(x_i - x_j)$, the action $\sigma \cdot \Delta = \prod_{i<j}(x_{\sigma(i)} - x_{\sigma(j)}) = \pm\Delta$ defines $\operatorname{sgn}\sigma$.

Proposition 3.8.4. sgn *is a homomorphism* $S_n \to \{\pm 1\}$*;* $\operatorname{sgn}\sigma = (-1)^{\operatorname{inv}(\sigma)} = (-1)^{n-c(\sigma)}$ *with* $c(\sigma)$ *the number of cycles including fixed points; transpositions have sign* -1 *and generate* S_n*.*

Proof. The homomorphism property is immediate from the action on Δ. The factor $(x_{\sigma(i)} - x_{\sigma(j)})$ is reversed exactly at the inversions. A k-cycle is a product of $k-1$ transpositions, so a permutation with c cycles is a product of $n-c$ of them, and sorting by swaps shows transpositions generate. □

The inversion formula converts an algebraic invariant into a finite count you can organise. Putnam 2012 B6 does that for $\sigma(i) = ai \bmod q$: write $\sigma(i) = ai - q\lfloor ai/q \rfloor$, note σ is increasing on each level set of $\lfloor ai/q \rfloor$ so every inversion is cross-level, and for i in level r and j in level $r' > r$ the condition is the single inequality $a(j-i) < (r'-r)q$. The count reduces to three triangular numbers, and that level-set organisation is the reusable shape. The generation statement has a cheap corollary: a property invariant under one transposition is invariant under every permutation, which is how Putnam 2018 A2 reduces an order-independence claim to one case.

Involutions and the parity circuit

Proposition 3.8.5. *If a finite* X *carries an involution* ι *then* $|X| \equiv \#\mathrm{Fix}(\iota) \pmod 2$*; if* ι *is fixed-point-free,* $|X|$ *is twice the number of orbits.*

Proof. The non-fixed points fall into orbits of size 2. □

This is one of the highest-return circuits in the chapter, because it converts a parity question into a classification of fixed points, usually a finite check. In Putnam 2008 B6 the map $\sigma \mapsto \sigma^{-1}$ has the involutions as fixed points and preserves k-limitedness, so the count has the parity of the number of k-limited involutions. In Putnam 2015 B5 word reversal is fixed-point-free on the valid permutations for $n \geq 2$, since the first and last letters differ, giving $P_n = 2\Sigma_n$ and halving a classification from ten oriented shapes to five.

A sharper version, Putnam 2025 B5: if $T(\iota x) = c - T(x)$ then $\#\{T > c/2 + w\} = \#\{T < c/2 - w\}$, so

$$\#\{T > c/2 + w\} = \tfrac{1}{2}\big(|X| - \#\{|T - c/2| \leq w\}\big).$$

There the involution is $m \mapsto p - m$ with $T(m) = m + I(m)$ and $c = 2p$; it halves the count and isolates a small exceptional window, and the rest of the problem is bounding that window. Store this as the antisymmetric involution circuit; it is the same device as pairing k with $p-k$ in a sum over $\mathbb{F}_p$.

Finally, Putnam 2025 A3: a perfect matching is a fixed-point-free involution τ with v adjacent to $\tau(v)$ always. On $\{0,1,2\}^n$ minus the zero string, transposing the values 1 and 2 in the first nonzero coordinate is such an involution, self-inverse precisely because the transposition does not move which coordinate is first nonzero.

Patterns, uniformity and runs

The *pattern* of $(a_1, \ldots, a_n)$ is the sign vector of consecutive differences, equivalently the descent set, and it depends only on relative order, so pattern counts also count arrangements of any fixed set of distinct reals. Three facts make patterns manageable. Descent-set-contained counts are multinomials and exactness comes from the sieve (Section 5). The symmetry group has order four: complementation $a \mapsto n+1-a$ negates every sign and reversal negates and reverses, so the two alternating patterns lie in one orbit (Putnam 2025 A5). And every sign pattern is realised: induct on n, appending n when the last sign is $+1$ and otherwise adding one to every entry and appending 1.

That third fact is load-bearing, since A5's main inequality is strict only because the omitted term is at least 1.

Proposition 3.8.6. *If π is uniform on S_n and $i_1 < \cdots < i_m$ are fixed positions, the relative order of $\pi(i_1), \ldots, \pi(i_m)$ is uniform over the $m!$ orderings.*

Proof. For fixed $\tau \in S_m$, rearranging the values in those positions by τ is a bijection of S_n and preserves the uniform measure. □

With $m = 2$ at the ends and $m = 3$ inside, this gives probabilities $\frac{1}{2}, \frac{1}{3}, \frac{1}{2}$ and the expected peak count $2 \cdot \frac{1}{2} + (n-2) \cdot \frac{1}{3} = (n+1)/3$ in one line (Putnam 2006 A4).

Theorem 3.8.7 (Putnam 2023 B3)**.** *For a finite sequence of distinct reals with N maximal monotone runs, the longest alternating subsequence has length exactly $N + 1$.*

Proof. The run endpoints give the lower bound. For the upper bound, encode the sequence by the word of signs of consecutive differences. Deleting one term deletes exactly one letter of that word, because the two merged differences add to a nonzero number whose sign is one of theirs; and deleting a letter cannot increase the number of adjacent unequal pairs. So the run count is monotone under passing to subsequences, which bounds every subsequence at once. □

That is the model of a structure theorem: it removes a maximum from a definition and makes the statistic local.

- **WHEN** a permutation appears anywhere **DO** name its cycle type, sign, descent set and inversion count immediately **BECAUSE** those four carry essentially everything a Putnam problem asks (Putnam 2012 B6, Putnam 2024 B1, Putnam 2025 A5)
- **WHEN** a parity question about a finite set **DO** find an involution and count its fixed points **BECAUSE** $|X| \equiv$ #Fix modulo 2 and the fixed points are a short classification (Putnam 2008 B6, Putnam 2015 B5)
- **WHEN** a statistic is antisymmetric under an involution, $T(\iota x) = c - T(x)$ **DO** halve the count and bound the exceptional window near $c/2$ **BECAUSE** the two tails are equinumerous by the involution (Putnam 2025 B5)
- **WHEN** a claim must hold for every permutation **DO** check it for one transposition **BECAUSE** transpositions generate S_n (Putnam 2018 A2)
- **WHEN** a probability involves only relative orders at fixed positions **DO** use uniformity of the induced pattern and multiply by the number of positions **BECAUSE** rearranging values at those positions is measure-preserving (Putnam 2006 A4)
- **WHEN** alternating subsequences, zigzags, or "up then down" appear **DO** count maximal monotone runs and use length = runs + 1; if the pattern is itself alternating expect the zigzag numbers $\sec x + \tan x$ **BECAUSE** run count is monotone under deletion, which makes the bound uniform (Putnam 2023 B3, Putnam 2025 A5)

3.9 Partitions, Young diagrams and the hook length formula

Definition 3.9.1. A partition $\lambda = (\lambda_1 \geq \cdots \geq \lambda_r > 0)$ of N is drawn as its Young diagram, left-aligned rows of λ_i cells. The conjugate λ' is the transpose, so λ'_j counts rows of length at least j.

Conjugation is why a sorted sequence is determined by its threshold counts: a nondecreasing sequence of nonnegative integers is recoverable from "how many terms are at least m", because that function is the conjugate diagram. Putnam 2008 A3 is exactly this duality applied at every prime, and it is what upgrades "the threshold counts are invariant" to "the sorted sequence is unique".

Partitions with at most a parts each at most b are the diagrams inside an $a \times b$ box, hence the monotone paths of Section 2, counted by $\binom{a+b}{a}$. Refining by area gives the Gaussian binomial coefficient $\begin{bmatrix} a+b \\ a \end{bmatrix}_q$ with the q-Pascal recursion

$$\begin{bmatrix} a+b \\ a \end{bmatrix}_q = \begin{bmatrix} a+b-1 \\ a-1 \end{bmatrix}_q + q^a \begin{bmatrix} a+b-1 \\ a \end{bmatrix}_q .$$

In Putnam 2023 B1 the moved coins form a diagram anchored at a corner and the number of moves is its area, so the q-refinement is the structure behind an answer that only needs $q = 1$.

Definition 3.9.2. A standard Young tableau of shape $\lambda \vdash N$ fills the diagram with $1, \dots, N$ increasing along rows and down columns; equivalently it is a linear extension of the cell poset ordered by "weakly above and weakly left". Write f^λ for the number of them.

Two consequences are immediate: the cell holding k has at most $k-1$ cells weakly above-left, so 1 sits at $(1,1)$ and 2 at $(1,2)$ or $(2,1)$; and recording the row index of $1, 2, \dots, N$ gives a lattice word in which every prefix has at least as many is as $(i+1)$s, which for two rows is the ballot problem, giving $f^{(n,n)} = C_n$ by reflection.

Theorem 3.9.3 (Frame, Robinson and Thrall, 1954)**.** *$f^\lambda = N! / \prod_c h(c)$, where the hook length $h(c)$ is 1 plus the cells strictly right of c in its row plus those strictly below in its column.*

The proof is long and you will not reconstruct it in the arena. That is a named knowledge gap, and the honest way to carry it is as a quoted theorem plus two memorised cases, $f^{(n,n)} = C_n$ and $f^{(n,n,n)} = 2(3n)!/(n!(n+1)!(n+2)!)$. What you can reconstruct is the branching recursion, and it is what to reach for first.

Proposition 3.9.4. *$f^\lambda = \sum_c f^{\lambda - c}$ over removable corners c, with $f^\varnothing = 1$; more generally the linear extensions of a finite poset satisfy $L(P) = \sum_m L(P - m)$ over maximal elements.*

Proof. The largest entry occupies a removable corner, and deleting it leaves a standard tableau of the smaller shape. □

So the probability that the maximum of a uniform tableau sits at corner c is $f^{\lambda - c} / f^\lambda$, and any candidate closed form is verified by checking the recursion and the initial condition; that is how the three-row formula is confirmed without proving the general theorem. Putnam 2024 A3 is built from these pieces: S is the tableaux of shape (n,n,n); locating the entry 2 turns $T(2,1) < T(1,2)$ into the event $T(2,1) = 2$; and the complementation involution $T^*(i,j) = N + 1 - T(\text{rows} + 1 - i, \text{cols} + 1 - j)$, which reverses both orders and so preserves standardness, rotates the leftover skew shape into the straight shape $(n, n-1, n-1)$ where the hook formula applies. The answer is a ratio of hook products, $(n+1)/(3n-1)$.

- **WHEN** a sorted sequence or multiset must be shown unique or invariant **DO** pass to the conjugate: record how many terms are at least m, for every m **BECAUSE** a partition is determined by its transpose (Putnam 2008 A3)
- **WHEN** a filling increases along rows and down columns **DO** it is a standard Young tableau: count by the hook formula and locate small entries by counting cells weakly above-left **BECAUSE** k needs $k-1$ predecessors, so 1 and 2 have at most two possible cells (Putnam 2024 A3)
- **WHEN** the hook formula does not apply to the shape you have **DO** use branching $f^\lambda = \sum_c f^{\lambda - c}$, or rotate a skew shape by 180 degrees and complement the entries **BECAUSE** the maximum sits at a corner, and complementation reverses both orders at once (Putnam 2024 A3)
- **WHEN** a configuration is a monotone staircase inside a box **DO** read it as a partition in an $a \times b$ box, and track moves by the area with the q-Pascal recursion **BECAUSE** diagrams, monotone paths and two-letter words are the same object (Putnam 2023 B1)

3.10 Extremal arguments, Ramsey, and the two methods

The extremal principle and the two-sided contract

Any nonempty finite set of reals attains a maximum and a minimum. That triviality is a construction device: it manufactures a fixed object with a strong property and no compactness argument. Putnam 2016 B3 takes the triple of maximum area among all triples of a finite point set and builds the whole covering argument around it; Putnam 2010 A1 takes the maximum element of a partitioned totally ordered set, whose group then has invariant at least that maximum, which is the crux bound.

Remark 3.10.1 (Bound plus construction). An answer to "what is the largest or smallest possible number of X" is two independent proofs: that no configuration beats the value, and one explicit configuration attaining it. Neither half implies the other and each is worth roughly half the credit.

The evidence is overwhelming: Putnam 2016 A4 (lower bound by colouring, then an explicit tiling), Putnam 2021 A1 (a universal bound, then the walk $288(3,4) + 288(4,3) + (5,0) + (0,5)$ absorbing the slack of 4), Putnam 2022 A2 (the free bound $2n-1$ closes to $2n-2$ only via sign propagation, and the value is confirmed only by a palindromic example), and Putnam 2022 A5, whose minimax induction needs both directions at every step and supplies four explicit attainment splits. The construction habit: saturate the bound's extremal case almost everywhere and absorb the arithmetic slack explicitly and locally.

Two smaller extremal shapes. *Extremal induction*: peel the largest and smallest elements separately, note each remainder inherits the hypothesis, and glue on the overlap; that is how Putnam 2009 B3 classifies sets closed under integer midpoints as arithmetic progressions of odd common difference. *Sort then induct on a sufficient sub-family*: to find a k-subset with property P, sort and show P holds for some window of k consecutive terms, which proves the full claim because windows are a subfamily of all k-subsets; Putnam 2012 A1 reduces $\binom{12}{3} = 220$ triples to ten windows this way.

Ramsey type arguments

$R(3,3) = 6$: a vertex of K_6 has five incident edges so three share a colour, say red to a, b, c; if any of ab, bc, ca is red there is a red triangle, otherwise abc is blue, and the pentagon plus pentagram colouring of K_5 has none. Ramsey numbers as such do not appear in the 2006 to 2025 evidence for these fields, so do not study bounds for $R(s,t)$. What appears is the Ramsey *shape*, a forced-colour chase. Putnam 2025 B1 is Euclidean Ramsey: the apexes of the two equilateral triangles on a segment cannot both share the colour of their common centre, and the unit equilateral triangle forces a monochromatic unit distance by pigeonhole; the official first solution chases eight such points, which is why you rank it below the counting route of Section 3. Putnam 2022 B3 is the same genre on the line, where for fixed d the graph joining t to $t+d$ is a union of two-way infinite paths, hence bipartite, so a distance can in fact fail to be monochromatic and the problem becomes a rigidity cascade. Two lemmas from it generalise cleanly. *Doubling*: with exactly two colours, $\chi(t) \neq \chi(t+d)$ for all t forces $\chi(t+2d) = \chi(t)$, so $2d$ is monochromatic; with m colours the hypothesis $\chi(t+d) = \rho(\chi(t))$ gives $\chi(t+kd) = \rho^k(\chi(t))$ and period $\operatorname{ord}(\rho) \cdot d$, which is why the conclusion fails for $m \geq 3$. *Splitting*: if $\chi(t) \neq \chi(t+d)$ and $0 < s < d$ then $\chi(t+s)$ agrees with exactly one of them, so one of s and $d-s$ is monochromatic, and taking $s = d/2$, the fixed point of $s \mapsto d-s$, makes the alternatives identical and forces $d/2$.

The linear algebra method

Proposition 3.10.2 (Oddtown)**.** *If $A_1, \ldots, A_m \subseteq [n]$ have odd sizes and pairwise even intersections then $m \leq n$.*

Proof. Take characteristic vectors in $\mathbb{F}_2^n$; then $v_i \cdot v_j = |A_i \cap A_j| \bmod 2$, so the Gram matrix is the identity and the v_i are independent in an n-dimensional space. □

Eventown, with even sizes and even intersections, allows $2^{\lfloor n/2 \rfloor}$ sets, so the parity of the diagonal does all the work. Putnam 2011 A4 is in this family, and recognising the family is what makes the Gram matrix your first object rather than your last. The general method: map the family into a space of known dimension and prove independence, usually via a dual family whose pairing is nonzero exactly on the diagonal. The standard named instance is the Bollobas set pair inequality: pairs (A_i, B_i) with $A_i \cap B_i = \emptyset$ and $A_i \cap B_j \neq \emptyset$ for $i \neq j$, sizes a and b, number at most $\binom{a+b}{a}$. Putnam 2014 A6 is in this family and also illustrates the multiplicative principle for independent channels: a condition factoring into m independent conditions in spaces of dimension d_t has maximal family size $\prod_t d_t$, since $K(m) = nK(m-1)$ with $K(0) = 1$.

The probabilistic method at Putnam level

At this level it means linearity of expectation used as a counting device: E of a sum is the sum of Es with no independence needed, and some object is at least as good as the average. Putnam 2006 A4 is the model, with the expected peak count a sum of uniformity probabilities. Putnam 2016 B4 is expectation used to kill terms, leaving only the perfect matchings and the count $(2n)!/(2^n n!)$.

The weighted form looks probabilistic and is really the distributive law.

Proposition 3.10.3. *If $w(s) = \prod_{i=1}^{k} c_{s_i}$ is multiplicative across coordinates then $\sum_{s \in A^k} w(s) = (\sum_{j \in A} c_j)^k$; if also $w \equiv W$ on S and all $c_j > 0$ then $|S| \leq (\sum_j c_j)^k / W$.*

That is Putnam 2018 B6 in one sentence: the weight $2^{-\sigma(s)}$ is multiplicative, its total factorises, and it is constant on the set to be bounded. Reach for a multiplicative weight whenever a crude count overshoots by orders of magnitude. Beyond linearity of expectation the probabilistic method is rare on this paper; the only appearance of concentration is Putnam 2017 A5, where the cycle count of a random permutation concentrates near $\log n$. Do not over-invest.

- **WHEN** a question asks for the largest or smallest possible value **DO** write the two headings "bound" and "construction" before doing any mathematics **BECAUSE** they are independent proofs and half the marks live in the one you were going to skip (Putnam 2016 A4, Putnam 2021 A1, Putnam 2022 A2)
- **WHEN** you need a distinguished object with a strong property **DO** take the extremal one: maximum area, largest element, longest path, shortest representation **BECAUSE** finite nonempty sets attain extremes, and the extremal object cannot be improved, which is the hypothesis you wanted (Putnam 2016 B3, Putnam 2010 A1)
- **WHEN** a family of finite sets must be bounded in number **DO** take characteristic vectors and compute the Gram matrix over $\mathbb{F}_2$ or $\mathbb{R}$ **BECAUSE** prescribed intersection parities make it invertible, forcing independence and a dimension bound (Putnam 2011 A4, Putnam 2014 A6)
- **WHEN** some object must have at least or at most a given value of a statistic **DO** compute the expectation by linearity and quote that some object beats the mean **BECAUSE** no independence is needed and the per-position probabilities are uniformity lemmas (Putnam 2006 A4, Putnam 2016 B4)
- **WHEN** a crude count overshoots the target by orders of magnitude **DO** find a multiplicative weight constant on the target set and factorise the total **BECAUSE** the factorised total is a k-th power and gives the bound in one line (Putnam 2018 B6)
- **WHEN** a colouring of a line or plane must contain a monochromatic pattern **DO** run local forcing: doubling and splitting lemmas first, then a rigid configuration chase **BECAUSE** two colours make the alternation cascade close, which is exactly why the statement fails for three (Putnam 2022 B3, Putnam 2025 B1)

3.11 Graph theory

Definition 3.11.1. A simple graph $G = (V, E)$ has E a set of unordered pairs; $\deg v$ counts edges at v. A path repeats no vertex, a cycle is a closed path, G is connected if every pair is joined by a path. A forest is acyclic, a tree is a connected forest, a leaf has degree 1. G is bipartite if V splits into two parts with every edge crossing. A matching is a set of disjoint edges, perfect if it covers V.

The two counting lemmas are handshake (Section 3) and the following threshold, which you will use most.

Proposition 3.11.2. *A forest on V vertices with c components has exactly $V - c$ edges. Hence V vertices with V or more edges force a cycle. In a bipartite graph every cycle has even length at least* 4*.*

Proof. Induct by removing a leaf: a finite forest with an edge has one, since an endpoint of a maximal path can have no further neighbour without closing a cycle or extending the path. Removing it drops V and E by one and leaves the component count unchanged. □

Putnam 2014 B3 uses that threshold as a trap door: model the matrix as a bipartite graph with rows and columns as the two vertex classes, so at least $m + n$ distinct primes become at least $m + n$ edges on $m + n$ vertices, exactly where a cycle is unavoidable, and the cycle supplies the closed alternating sequence the rest of the proof multiplies around. Store the translation as the circuit: a matrix hypothesis counting "rows plus columns" means build the bipartite graph.

Trees, Cayley, functional graphs

Theorem 3.11.3 (Cayley)**.** *There are N^{N-2} labelled trees on N vertices, and qN^{N-q-1} rooted forests whose roots form a prescribed set of q vertices.*

Four standard proofs exist (Prufer codes, the matrix tree theorem, Joyal's bijection, the first-entry peel). The one to be able to execute is the Prufer code, because it is a bijection you can state in a sentence: repeatedly delete the least leaf and record its neighbour, giving a sequence in $[N]^{N-2}$, invertibly. Putnam 2013 B5 is where this pays: the residual count after the free factor is the number of forests rooted at $\{1, \dots, k\}$, namely kn^{n-k-1}, so recognising the shape of the answer identifies the structure to build. The free factor is its own circuit: if a counting condition never reads certain coordinates, those coordinates contribute a bare power, here n^k.

Proposition 3.11.4 (Functional graphs)**.** *Drawing $x \to f(x)$ for a map on a finite set gives every out-degree* 1*, so each component is a directed cycle with rooted trees attached and every orbit is eventually periodic. Hence every point reaches K under iteration if and only if every cycle of f meets K.*

That converts the orbit condition of Putnam 2013 B5 into the finite statement that no cycle lies inside $\{k + 1, \dots, n\}$, which is what the forest description rests on.

Matchings, Hall, connectivity, orientation

Theorem 3.11.5 (Hall)**.** *A bipartite graph with parts X, Y has a matching saturating X if and only if $|N(S)| \geq |S|$ for every $S \subseteq X$.*

Proof. Necessity is clear. Induct on $|X|$: if every proper nonempty S has $|N(S)| > |S|$ strictly, match one vertex to any neighbour and delete both, since Hall's condition survives with slack. Otherwise some proper nonempty S is tight; match S to $N(S)$ by induction and check Hall's condition on $X \setminus S$ against $Y \setminus N(S)$, since a violation there would combine with S to violate it originally. □

A system of distinct representatives for $A_1, \dots, A_k$ is a choice of distinct $a_i \in A_i$, existing exactly when the membership bipartite graph satisfies Hall's condition. Putnam 2012 B3 is that translation: one winner per day, all distinct, so the whole problem is one counting lemma about neighbourhood sizes. The same problem names a structure worth carrying: a round-robin schedule of $2n$ teams over $2n-1$ days is a 1-factorisation of K_{2n} into $2n-1$ perfect matchings, and $2n-1$ is the edge chromatic number of K_{2n}. Perfect matchings also appear as fixed-point-free involutions (Putnam 2025 A3) and, counted, as the double factorial (Putnam 2016 B4).

Proposition 3.11.6 (Monotone closure)**.** *If $S_1 \subseteq S_2 \subseteq \cdots \subseteq X$ with $|X| = N$ and $S_r = S_{r+1}$ implies $S_r = X$, then $|S_r| \geq \min\{r, N\}$, so $S_N = X$. Equivalently a strongly connected Cayley digraph on G has diameter at most $|G| - 1$.*

Putnam 2016 A5 uses this to convert "the letters generate" into the bound $r \leq |G|$ with no minimality argument and no pigeonhole. The general statement to store: iterating any monotone map on subsets of an N-set produces an increasing chain stabilising within N steps at the least fixed point containing the start, and when the only fixed points are trivial, stabilisation means completion. Transitive closure, automaton reachability and closure under moves are all this skeleton.

Theorem 3.11.7 (Topological order)**.** *A finite digraph has no directed cycle if and only if its vertices can be listed so that every edge points forwards.*

Proof. An acyclic finite digraph has a vertex of indegree zero, namely the initial vertex of a directed path of maximal length, since an incoming edge would close a cycle or lengthen the path. Delete it and induct. □

Putnam 2021 B5 needs the topological order as the relabelling that makes a matrix unitriangular, so this is the bridge from a combinatorial statement back to linear algebra; the same peeling proves every finite strict partial order extends to a linear order. A connectivity fact of another flavour, from Putnam 2008 A1: on a connected graph an edge labelling summing to zero around every cycle is a difference of a vertex potential, the discrete analogue of a closed form being exact on a simply connected domain. That names why a three-variable identity carries exactly enough information to pin a function down as a single potential difference.

Eulerian and Hamiltonian structure, planarity, colouring

An Eulerian circuit exists in a connected graph exactly when every degree is even, and an Eulerian trail exactly when at most two degrees are odd; the proof is the maximal-trail argument and you should be able to produce it. Hamiltonian structure has no such criterion, and Putnam 2015 B5 is the honest Putnam version: the condition "$|i-j| = 1$ implies $|\pi(i) - \pi(j)| \leq 2$" says the word is a directed Hamiltonian path of the graph G_n on $[n]$ with $i \sim j$ iff $0 < |i-j| \leq 2$. What makes it tractable is local structure and not a Hamiltonicity theorem: vertex n has only neighbours $n-1, n-2$, vertex $n-1$ only $n-3, n-2, n$, and the cut between $\{1, \dots, k\}$ and the rest has three edges, which supply the endpoints, degrees and deletion the induction runs on.

Theorem 3.11.8 (Euler's formula)**.** *For a connected plane graph $V - E + F = 2$, counting the unbounded face.*

Proof. Induct on E: a tree has $E = V - 1$, $F = 1$; otherwise delete an edge on a cycle, dropping E and F by one each and preserving connectivity. □

Corollaries: a simple planar graph with $V \geq 3$ has $E \leq 3V - 6$, a bipartite one $E \leq 2V - 4$, so K_5 and $K_{3,3}$ are not planar; and every planar graph has a vertex of degree at most 5, which gives the five colour theorem by induction (the four colour theorem is quoted, never proved).

Putnam 2007 A6 is the arena version. For a triangulation of an n-gon with I internal vertices, $F = T + 1$ and counting triangle-edge incidences gives $3T = 2E - n$, so $T = n + 2I - 2$ and $E = 3I + 2n - 3$; feeding $\sum_v \deg v = 2E$ into that gives the discrete Gauss and Bonnet identity

$$\sum_{\text{internal } v} (6 - \deg v) \; + \sum_{\text{boundary } v} (4 - \deg v) \; = \; 6.$$

Read what it says. Under the hypothesis that every internal degree is at least 6 the internal contribution cancels, so the only global constraint is $\sum_{\text{boundary}}(4 - \deg v) \geq 6$, which must therefore be used as a local existence device (find a boundary run of low degrees and peel it) and not as a global bound. That is the strategic content of A6, and it came from computing the identity rather than guessing.

Colouring appears mostly as a tiling lower bound (Section 14) and once as a rank question: Putnam 2017 A6 counts labelled edge colourings of the icosahedron (12 vertices, 30 edges, 20 faces, degree 5 at every vertex, two faces at every edge). Since each edge lies in two faces, the sum of all face sums is twice the sum of all edge values, which vanishes identically in characteristic 2; and a bipartite dual would force the two part sums always to agree. Neither obstruction applies over $\mathbb{F}_3$ here, so the face-sum map is surjective and the count is $2^F 3^{E-F} = 2^{20}3^{10}$. Remember it in general form: for any triangulation $3F = 2E$, so surjectivity gives $2^F 3^{F/2}$.

- **WHEN** a hypothesis counts something as "rows plus columns" or "vertices" **DO** build a bipartite graph and compare edges with vertices **BECAUSE** V edges on V vertices force a cycle, and the cycle is what you actually wanted (Putnam 2014 B3)
- **WHEN** the answer has the shape kn^{n-k-1} or N^{N-2} **DO** you are counting labelled trees or rooted forests; build the Prufer or first-entry bijection **BECAUSE** the shape of the answer identifies the structure to construct (Putnam 2013 B5)
- **WHEN** one item per day, per person or per set, all distinct **DO** check Hall's condition on the bipartite membership graph **BECAUSE** a system of distinct representatives exists exactly then, so the problem becomes one neighbourhood count (Putnam 2012 B3)
- **WHEN** a set grows under a closure rule and you need a step bound **DO** use the monotone chain: it grows or it is finished, so N steps suffice **BECAUSE** no minimality or pigeonhole argument is needed (Putnam 2016 A5)
- **WHEN** a hypothesis says "no directed cycles" or "no cyclic dependency" **DO** take a topological order and relabel **BECAUSE** the matrix becomes unitriangular and determinants become trivial (Putnam 2021 B5)
- **WHEN** a planar or polyhedral configuration with a degree hypothesis **DO** write $V - E + F = 2$, count face-edge incidences, convert the hypothesis into the Gauss and Bonnet sum **BECAUSE** the internal terms cancel, so the hypothesis is local existence and not a global bound (Putnam 2007 A6)
- **WHEN** a permutation condition constrains $|\pi(i) - \pi(j)|$ for adjacent i, j **DO** redraw it as a Hamiltonian path in an explicit graph and read off degrees, endpoints and cuts **BECAUSE** the local structure supplies the deletion the induction needs (Putnam 2015 B5)

3.12 Posets, chains and antichains

Definition 3.12.1. A partially ordered set carries a reflexive, antisymmetric, transitive relation. A *chain* is a totally ordered subset, an *antichain* a pairwise incomparable one, an *order ideal* a downward closed subset, a *linear extension* a total order refining the partial one. x is covered by y if $x < y$ with nothing between, and the poset is *graded* if every cover raises a rank function by one. A *lattice* has $a \vee b$ and $a \wedge b$ for every pair.

The Putnam use is always the same move: an array, a family of sets or a matrix is secretly a poset, and saying so makes the right induction available.

Order ideals. If an array is nondecreasing along rows and columns then $\{(i,j) : a_{i,j} \le t\}$ is an order ideal, so the zeros of each row are a prefix of columns and the zeros of each column a prefix of rows. Putnam 2025 B4 cannot even name its row sums without this.

Linear extensions. Standard Young tableaux are linear extensions (Section 9). Putnam 2024 B5's function is the order polynomial $\Omega_P(n)$, counting order-preserving maps $P \to [n]$, a polynomial of degree $|P|$ with leading coefficient $e(P)/|P|!$ where $e(P)$ counts linear extensions; for the tree with $k+m+1$ elements, $e(P) = \binom{k+m}{k}$, which is the leading-coefficient check. Putnam 2023 B6 uses the other consequence: a matrix with entry 1 exactly when $i \le j$ in a partial order becomes unitriangular under a linear extension, so its determinant is 1 and its columns form a basis over $\mathbb{Z}$.

Lattices and valuations. A family of subsets closed under union and intersection is a distributive lattice, and by Birkhoff's representation theorem every finite distributive lattice arises this way, so such a hypothesis is fully general rather than special. Putnam 2013 B3 combines that with two more facts: the hypothesis $f(S \cup S') + f(S \cap S') = f(S) + f(S')$ says f is a valuation, and its second hypothesis says the poset is graded by cardinality with every nonempty element covering one exactly one size smaller, which licenses strong induction on rank with the base case obtained by iterating covers down to $\emptyset$. The contrast showing why equality is essential: a matroid rank function is only submodular, and two parallel elements give $\text{rank}\{1,2\} = 1 \ne 2$, so a merely submodular set function need not be additive over elements.

The named theorems are Dilworth's (minimum chain cover equals maximum antichain), its dual Mirsky's (minimum antichain cover equals longest chain), whose standard corollary is Erdos and Szekeres, that any sequence of $mn+1$ distinct reals has an increasing subsequence of length $m+1$ or a decreasing one of length $n+1$; and Sperner's, that the largest antichain in $2^{[n]}$ has size $\binom{n}{\lfloor n/2 \rfloor}$, which follows from the LYM inequality $\sum_{A \in \mathcal{F}} \binom{n}{|A|}^{-1} \le 1$, itself a double count of maximal chains, since each maximal chain meets an antichain at most once and $|A|!(n-|A|)!$ chains pass through A. Now the honest ranking: Dilworth and Sperner do not appear in the 2006 to 2025 evidence for these fields. The poset material that does appear is order ideals, linear extensions, graded induction and valuations. Learn those four first and carry the rest as quotable one-liners.

- **WHEN** an array or family is monotone in two directions **DO** name the order ideal and describe rows and columns as prefixes **BECAUSE** monotonicity is downward closure, which is what makes row statistics definable (Putnam 2025 B4)
- **WHEN** a family of sets is closed under union and intersection **DO** say "finite distributive lattice" and look for a valuation and a rank **BECAUSE** Birkhoff says nothing is lost, and grading licenses induction along covers (Putnam 2013 B3)
- **WHEN** a matrix is indexed by a partial order or by divisibility **DO** relabel by a linear extension to make it unitriangular **BECAUSE** determinant 1 and integrality come free (Putnam 2023 B6, Putnam 2021 B5)
- **WHEN** you must count order-preserving maps or fillings **DO** it is a linear extension count: use branching, or the order polynomial and its leading coefficient **BECAUSE** the maximum sits at a maximal element in every such count (Putnam 2024 A3, Putnam 2024 B5)

3.13 Combinatorial games

Definition 3.13.1. In a finite two-player perfect-information game with no draws and no repeated positions, played under the *normal* convention (a player unable to move loses), a position is a *P-position* if the Previous player wins with correct play and an *N-position* if the Next player does.

Proposition 3.13.2. *Terminal positions are P-positions; a position is an N-position iff some move reaches a P-position, and a P-position iff every move reaches an N-position.*

Proof. Induct on the length of the longest remaining play, finite by hypothesis. □

For sums of independent games there is Sprague and Grundy theory: $g(x) = \mathrm{mex}\{g(y) : x \to y\}$, where mex is the least nonnegative integer not in the set; x is a P-position iff $g(x) = 0$, and the Grundy value of a disjoint sum is the bitwise XOR of the parts, which is Nim. State the position plainly: in the 2006 to 2025 evidence for these fields no problem is decided by Grundy values. The games that appear are decided by parity, by pairing, or by direct minimax induction, so Sprague and Grundy is a one-page insurance policy, not a core circuit.

What does appear:

Parity of a terminal clique (Putnam 2013 B6). If a set Z of positions is closed under moves, every two of its elements are joined by a move, and positions may not repeat, then play inside Z is a bare enumeration of Z, so the player who first enters Z loses when $|Z|$ is even and wins when $|Z|$ is odd. With $Z = P_{n-1} \cup P_n$ and $|Z| = n + 1$ even, that is the only place the hypothesis that n is odd is used.

Pairing strategies (Putnam 2025 A3). If the positions partition into pairs joined by legal moves, the second player answers every move by moving to its partner and can therefore always move. A pairing is a perfect matching, hence a fixed-point-free involution along edges; on $\{0, 1, 2\}^n$ minus the zero string the explicit involution transposes the values 1 and 2 in the first nonzero coordinate. Naming the game graph as the n-fold Cartesian product of the path on three vertices converts "legal move" into "edge to an unvisited vertex" and exposes the counts 3^n and $3^n - 1$ that decide who can pair. The symmetry strategy is the same circuit with the pairing given by a reflection or a central involution.

The play is a permutation (Putnam 2023 A6). A sequence of distinct choices exhausting $[n]$ is a bijection from turns to numbers, and "k was chosen on turn k" says k is a fixed point. Renaming the game data as a permutation is what makes the pairing argument expressible, since pairs of numbers are then hit or missed together. The same problem carries a caution: the random model, in which fixed points are Poisson of mean one, calibrates intuition and is irrelevant to an adversarial game.

Minimax value induction (Putnam 2022 A5). To prove an optimal value equals v, show no move gives the mover more than v and exhibit a move attaining v, at every step. It is the bound plus construction contract applied recursively.

Discernment for token games (Putnam 2010 B3). Distinguish games where the amount moved is fixed by the source's own label, forcing a capacity or threshold analysis, from games where the amount is free or constant, which usually admit greedy balancing but can carry a parity obstruction instead. The label-tied family is the chip firing or abelian sandpile family, and knowing the name means recognising the next one on sight.

- **WHEN** a finite two-player game with no draws **DO** compute P and N positions by backward induction from the terminal ones **BECAUSE** the labels are forced, and a small table usually reveals the invariant (parity, a residue, a matching)
- **WHEN** the second player must be shown to survive every move **DO** find a pairing of positions along legal moves, that is, a fixed-point-free involution **BECAUSE** answering into the partner is always legal, so the second player never runs out (Putnam 2025 A3)
- **WHEN** play must eventually enter a small closed set of positions **DO** count that set and read off the parity of who enters first **BECAUSE** inside a closed clique the play is a bare enumeration (Putnam 2013 B6)
- **WHEN** a game asks for the optimal value rather than the winner **DO** induct with two clauses at every step: no move beats v, and one move attains v **BECAUSE** a minimax induction with only the inequality half proves nothing (Putnam 2022 A5)

- **WHEN** tokens are redistributed by an amount tied to the source's own label **DO** run a capacity or threshold analysis; it is a chip firing game **BECAUSE** the label-tied move size is what forces a capacity bound rather than a greedy balance (Putnam 2010 B3)

3.14 Invariants, monovariants and colourings

Definition 3.14.1. An *invariant* of a process is unchanged by every legal move; a *monovariant* (potential) changes strictly in one direction.

An invariant proves impossibility: different values at the two ends means no sequence of moves connects them. A monovariant proves termination and bounds the number of steps. Both are read off the move, never guessed, and the reading is always the same question: what does one move do to this quantity?

Proposition 3.14.2 (Descent)**.** *If a state carries a potential in* $\mathbb{N}$ *that strictly decreases at every move, the process terminates and the terminal state is locally optimal. The parity of the potential is itself invariant when every move changes it by* 2*.*

Putnam 2017 A4 is the model: for an equal-split problem the potential $|\sum A - \sum B|$ is congruent to the total modulo 2, and swapping a pair differing by one decreases it by exactly 2, so the terminal value is 0 or 1 according to the parity of the total.

The two invariants that appear most are inversion counts. In Putnam 2020 B2, for a peg position set $I(P) = \#\{(v, p) : v \text{ vacant}, p \in P, v > p\}$; a move from p to v decreases I by exactly $v - p$, so $\sum_t d_t = I(P_0) - I(\text{terminal}) = k(n-k)$. The total distance is conserved while the number of moves is not, which is what the problem is about; generalise to any game whose move shifts one token past a computable number of obstacles. In Putnam 2023 B1, swapping an adjacent LD into DL in a two-letter word reduces inversions by exactly one, so any process of adjacent sorting swaps terminates after exactly $\text{inv}(w)$ steps, bounding every play by $(m-1)(n-1)$ and driving the surjectivity induction.

Two more shapes. *Bounded increasing statistic*: if each node of a binary refinement tree carries a positive integer statistic strictly increasing from parent to child, and a node is split only while the statistic is at most N, every root-to-leaf path has length at most $N-1$ and the tree is finite (Putnam 2014 B6). *Compressing a lexicographic monovariant*: if a process improves a later coordinate while earlier ones are secondary, choose strictly increasing weights and combine into a weighted sum or product, so the layered induction becomes one displayed inequality, provided the weighted gain later dominates any simultaneous loss earlier (Putnam 2008 A3). That converts "prove termination layer by layer" into a single line and applies to any finite process with a lexicographic improvement.

Confluence is the companion property. A process that repeatedly finds any out-of-order pair and replaces it by its minimum and maximum in position order terminates at the same sorted arrangement whatever choices are made, because for each threshold the number of elements at least that large is invariant under every comparator firing (Putnam 2008 A3). Threshold counting is the template: to show a process has a unique outcome, find a family of counts each of which is invariant.

Colouring and weighting for tilings

Theorem 3.14.3. *If* w *is a weight function on cells with* $\sum_{c \in T} w(c) \le 1$ *for every legal placement* T *of every allowed tile, then any tiling uses at least* $\sum_c w(c)$ *tiles. The* $0/1$ *case: if no legal tile covers two cells of* R*, any tiling uses at least* $|R|$ *tiles.*

Proof. Sum w over the cells of each tile and add over the tiles: the total is the region's weight and each tile contributes at most 1. □

This is linear programming duality for the covering problem, and it is the only reason a tiling lower bound can beat the area bound. Putnam 2016 A4 is the complete case, and it also shows how to *find* the colouring instead of guessing: a polyomino inside a $p \times q$ bounding box covers at most one cell of any set meeting every translate of that box at most once, so for a 2×2 box the candidates are forced to be the four parity cosets of $2\mathbb{Z} \times 2\mathbb{Z}$, and on a $(2m-1) \times (2n-1)$ board the coset with both coordinates odd is largest, with mn cells. Area accounting then fixes the census before you draw anything: with a tiles of area 3, b of area 4 and $a + b = N$ prescribed, $b = A - 3N$ and $a = 4N - A$, so a minimum tiling is almost all trominoes. The construction assembles from a pinwheel decomposition of the odd square about its centre plus a strip lemma, at cost $N(m,n) = N(m-1,n) + n = N(m,n-1) + m$ with $N(4,4) = 16$, solving to mn.

The density version: if a set is partitioned into congruent pieces of size V, the piece centres have density $1/V$, so V is an invariant of any valid construction. Putnam 2019 B6 extracts the modulus $2n+1$ from this before any construction exists, which is why the eventual homomorphism looks forced rather than invented. The general principle it carries: when a family of problems indexed by n has an invariant whose values at consecutive n are coprime, as $2n+1$ and $2n+3$ are, no layering construction can pass from one to the next and a uniform construction must be sought. The same problem is the honest boundary marker: perfect Lee codes of radius 1 exist in every $\mathbb{Z}^n$ (the ball has $2n+1$ points), and Golomb and Welch conjectured in 1970 that for $n \geq 3$ and radius at least 2 none exist, so generalising it in the arena is wasted time.

Lattice points and Pick

Counting lattice points is counting solutions of linear equations and inequalities, and there are two circuits: substitute to kill a congruence, and bracket between boxes. For the first, to count pairs (k,b) with $d \mid k$, $k \geq 1$, $b \geq b_0$, $k + b \leq M$, put $k = dj$ so the congruence disappears; for each j the admissible b form an interval and the total is an arithmetic series with no floor functions. Putnam 2020 A1 is exactly this, and the other summation order produces floors and is strictly worse. For the second, the unordered coin collections of denominations $w_1, \ldots, w_k$ totalling N are the lattice points of the simplex $\{x \geq 0 : \sum x_i w_i = N\}$, equal to $[x^N] \prod_i (1 - x^{w_i})^{-1}$, and a lattice point count can be bracketed between two boxes. Putnam 2007 B6 does that, the lower bound coming from the fact that if one weight is 1 then any choice of the other coordinates not overshooting N extends uniquely, so the solution set contains an injective image of a box. For fixed k the count is a quasi-polynomial in N of degree $k-1$ with leading term $N^{k-1}/((k-1)!\, w_1 \cdots w_k)$ (Ehrhart), which gives the compass direction before the rigorous work; in B6 the number of weights grows with n, so it is a heuristic only and you must say so.

Pick's theorem, that a simple lattice polygon has area $I + B/2 - 1$ with I interior and B boundary lattice points, did not appear anywhere in the 2006 to 2025 evidence for these fields. Learn it last, as a one-line insurance policy, and know that its proof is Euler's formula applied to a triangulation into unit-area triangles.

- **WHEN** a process asks whether a target is reachable **DO** find a quantity every move preserves and evaluate it at both ends **BECAUSE** different values prove impossibility with no search (Putnam 2018 B1, Putnam 2020 B2)
- **WHEN** a process must be shown to terminate **DO** find a nonnegative integer potential that strictly decreases, or a bounded statistic that strictly increases **BECAUSE** well-founded descent terminates, and the potential's parity gives the terminal value (Putnam 2017 A4, Putnam 2014 B6)
- **WHEN** a process improves a later coordinate at the cost of earlier ones **DO** weight the coordinates increasingly and take a weighted sum or product **BECAUSE** one scalar monovariant replaces a layered induction (Putnam 2008 A3)

- **WHEN** a tiling lower bound must beat the area bound **DO** find a weighting with total at most 1 per tile, searching among the parity cosets forced by the tile's bounding box **BECAUSE** that is LP duality for covering, and the bounding box forces the candidate colourings (Putnam 2016 A4)
- **WHEN** a region is partitioned into congruent pieces **DO** compute the piece size as a density invariant before attempting a construction **BECAUSE** the invariant fixes the modulus, and coprimality across n rules out recursive constructions (Putnam 2019 B6)
- **WHEN** you are counting lattice points under a congruence **DO** substitute to remove the congruence, then sum the resulting intervals **BECAUSE** floors appear only if you sum in the wrong order (Putnam 2020 A1)

3.15 The recursion compendium

The collected form of the standing order. For each family: the general shape, the solution method, the cue.

1. *Linear, constant coefficients.* $a_n = c_1 a_{n-1} + \cdots + c_d a_{n-d}$. Solve by the characteristic polynomial, or by $v_n = Tv_{n-1}$ on d-tuples; a root of multiplicity r carries a polynomial factor of degree $r-1$. Cue: deleting bounded structure at one end leaves a member of finitely many smaller families. Putnam 2015 B5, Putnam 2022 A3, Putnam 2019 B5.
2. *First order with additive forcing.* $s_j = s_{j-1} + w_j$, so $s_j = s_0 + \sum_{i \le j} w_i$; it extends to a sequence of period m exactly when $\sum_{i=1}^{m} w_i = 0$. Cue: any partial-sum or window statistic. Putnam 2020 B4, Putnam 2013 A4, Putnam 2017 A4.
3. *Self-referential linear.* $u_k = \alpha u_{k-1} + \beta u_k$ with $\beta \neq 1$ has ratio $\alpha/(1-\beta)$. Cue: an identity applied inside a sum reproduces the sum. Putnam 2020 A2.
4. *Binomial convolution.* $F_n = \sum_i \binom{n-1}{i-1} L_{i-1} R_{n-i}$, that is $F' = LR$ for exponential generating functions; nonlinear versions need Lagrange inversion. Cue: labelled objects split at their maximum, or a canonical first layer. Putnam 2025 A5, Putnam 2013 B5, Putnam 2006 A4, Putnam 2022 B1.
5. *Ordinary convolution.* $a_{n+1} = \sum_i b_i a_{n-i}$, that is $A = 1 + xBA$. Cue: first return to a level; unlabelled first block then the rest. Catalan, Putnam 2020 B4.
6. *Two-parameter insertion tables.* $T(n,k) = \alpha(n,k)T(n-1,k) + T(n-1,k-1)$. Solve as a change of polynomial basis, not by a closed form. Cue: the new element becomes a new part or joins an old one. Stirling numbers (Putnam 2007 B5, Putnam 2017 A5), Eulerian numbers (Putnam 2006 A4).
7. *Transfer systems.* $x(n+1) = Mx(n) + b$ on counts refined by a bounded feature set; solve by the characteristic polynomial of M. Cue: the reinsertion slot count depends on boundedly many features. Putnam 2015 B5.
8. *Digit and weight programmes.* $r(n) = r(n - u_m) + r(U_{m-1} - n)$ by complementation, or $f_m(N) = f_{m-1}(N) + f_m(N - w_m)$ for denominations. Cue: representations by a fixed increasing weight sequence; base-b digits are the case $u = 9 \cdot 10^j$. Putnam 2020 A5, Putnam 2007 B6.
9. *Additive geometric growth.* $R(k) = R(k-1) + k$, $R(0) = 1$, giving $\binom{n}{2} + n + 1$ regions for n lines in general position. Cue: a new generic object crosses all previous ones in distinct points. Putnam 2006 B3.

10. *Peel a slab.* $N(m,n) = N(m-1,n) + f(n)$ with f free of m, so $N(m,n) = N(m_0,n) + (m - m_0)f(n)$. Cue: a construction built by appending a repeating strip. Putnam 2016 A4.

11. *Extremal removal with a correction.* $f(n) \le f(n-1) + (\text{correction})$ from a restriction map whose fibre sizes and large fibres are separately bounded. Cue: an extremal count with a natural remove-one operation. Putnam 2006 B3.

12. *Facet splitting on the cube.* $f(n,k) \le \max\{2f(n-1,k-1), f(n-1,k)\}$ with $f(n,0) = 1$ solves to 2^k; the additive sibling $g(n,d) \le g(n-1,d) + g(n-1,d-1)$ solves to a partial sum of binomial coefficients, the Sauer and Shelah bound. Cue: a family of subsets of $\{0,1\}^n$ split by the last coordinate. Putnam 2006 B4.

13. *Monotone closure.* $S_{r+1} = S_r \cup S_r T$; an increasing chain in an N-set stabilises within N steps at the least fixed point. Cue: closure under moves, reachability, transitive closure. Putnam 2016 A5.

14. *Well-founded descent.* A strictly decreasing sequence in $\mathbb{N}$ terminates, and the potential's parity decides the terminal value. Cue: every move strictly improves something. Putnam 2017 A4, Putnam 2020 B2, Putnam 2023 B1, Putnam 2014 B6.

15. *Finite differences of counting sequences.* If $A(n)$ counts configurations on a ground set grown one element at a time and the property is translation invariant, $A(n+1) - A(n)$ counts configurations using the new element and the k-th difference counts those pinned at k new elements. Cue: the problem asks about a second difference. Putnam 2009 B3.

16. *Branching over maximal elements.* $L(P) = \sum_m L(P - m)$, for tableaux $f^\lambda = \sum_c f^{\lambda - c}$; verify a closed form against the recursion and the initial condition. Cue: linear extensions, standard tableaux, order-preserving fillings. Putnam 2024 A3.

17. *Doubling on a deficiency.* $s \mapsto s^2$ from an averaging step, giving $s \le (1 - 1/n)^{2^k}$ and completion in $O(\log n)$ rounds. Cue: each round at least squares a defect. Putnam 2008 A6.

18. *Colour dynamics along a coset.* $u_{k+1} = \rho(u_k)$ on colours along $t + d\mathbb{Z}$, so $u_k = \rho^k(u_0)$ with period $\mathrm{ord}(\rho)$; two colours with ρ the transposition give period 2. Cue: a hypothesis relating $\chi(t)$ and $\chi(t+d)$. Putnam 2022 B3.

19. *q-refinements.* Any Pascal-type recursion refines to $P_q(a,b) = P_q(a-1,b) + q^a P_q(a,b-1)$ when you track an area or inversion statistic, solved by the Gaussian binomial coefficient. Cue: the problem also counts moves, area or inversions. Putnam 2023 B1.

The meta-circuit: when you meet a recursion in the arena, do not solve it first. Identify which of these nineteen families it belongs to, because the family names the method.

3.16 Learning order and the rarity ledger

Learn in this order, because it matches the frequency in the evidence.

First, the tools that appear every year: the four counting principles and bijective encoding, double counting and finite Fubini, pigeonhole in all four forms, the binomial identities with Pascal and hockey stick, the bound plus construction contract, invariants and monovariants, and the deletion-insertion recursion protocol. If you install nothing else, install those seven; between them they carry the 54 priority-five items behind this chapter.

Second, the structural translations: bipartite graphs and Hall, forests and the edge-vertex threshold, functional graphs, order ideals and linear extensions, involutions and parity, generating functions for weight and digit recursions, Euler's formula.

Third, the named objects you must recognise but rarely construct: Catalan, Fibonacci, Stirling of both kinds, Bell, derangements, Euler zigzag and Eulerian numbers, Cayley's formula and Prufer codes, the hook length formula, Gaussian binomials.

Last, the exotica, with their single appearances named so you can rank them. The Mobius function of the partition lattice: Putnam 2017 B6. Fourier analysis on a finite abelian group: Putnam 2011 A6, and characters of $\mathbb{Z}^d$ in Putnam 2013 A6, both the same recognition step, that a difference kernel is a convolution and characters are its eigenvectors. Sturmian and Christoffel words: Putnam 2013 A4, Putnam 2018 A4 and Putnam 2020 B6, always the fact that a balanced word carries no information beyond its density. The Sauer and Shelah lemma: Putnam 2006 B4. Lindstrom, Gessel and Viennot: an optional route in Putnam 2024 A6. Perfect Lee codes and Golomb and Welch: Putnam 2019 B6. Addition chains and straight-line programs: Putnam 2009 B6, where the binary digit sum $s_2(n)$ is unbounded and eliminates every purely additive route in one line. Dilworth, Mirsky, Sperner, Ramsey numbers, Sprague and Grundy theory and Pick's theorem did not appear at all in these fields between 2006 and 2025; carry them as quotable statements and spend no drilling time on them until everything above is automatic.

One closing instruction, and it decides whether this chapter is worth anything to you. Every circuit is written WHEN, DO, BECAUSE, and the BECAUSE is the only part that transfers. A circuit memorised without its reason fires on the wrong cue and costs you more than not having it. So when you drill these, drill the third clause: say out loud why the move is forced by the definition of the object, the way the centre falls out of the definition of a circle. That sentence has to be self-generated, in your head, in the arena.

Chapter 4

Number theory

4.1 What this chapter is for

Edexcel hands you the integers as a place to do arithmetic and almost nothing else. You can factorise, you can divide polynomials, you can run an induction, you have the binomial theorem. What 9FM0 never hands you is the integers as a *structure*: the division algorithm stated as a theorem with a uniqueness clause, the greatest common divisor as an object with a universal property rather than a thing you find by listing factors, congruence as an equivalence relation that respects addition and multiplication, unique factorisation as a theorem rather than a slogan, and the whole apparatus that sits on top of those four things: valuations, orders, primitive roots, quadratic characters, arithmetic functions, Gaussian integers, Hensel lifting, continued fractions.

That gap matters more here than anywhere else in this book. Number theory is the single largest field in the corpus of harvested Putnam items, larger than real analysis, larger than combinatorics. On a typical paper at least two problems are pure arithmetic, and two or three more are problems labelled algebra, combinatorics or analysis in which an arithmetic fact is the bifurcation point. Putnam 2018 B4 is a problem about a real sequence and a trigonometric limit whose entire content turns out to be the Pisano period of the Fibonacci sequence modulo $4F_N$. Putnam 2022 B1 is a problem about a power series whose entire content is that $v_2(n!) < n$. Putnam 2014 B6 is a problem about Lipschitz functions whose entire content is the mediant of a unimodular pair. If your arithmetic is weak you will lose those problems without ever realising they were arithmetic.

By the end of this chapter you will do the following without deliberation. You will read a divisibility statement and write the congruence in the same breath. You will see a factorial or a binomial coefficient inside a divisibility question and reach for Legendre and Kummer. You will see an exponent that varies and reach for multiplicative order. You will see the phrase "is a perfect square" and reach for valuations modulo two. You will see a sum of two squares and move into $\mathbb{Z}[i]$. You will see a bilinear Diophantine equation and complete the rectangle. You will see an irrational rotation and reach for pigeonhole on the circle. Above all, you will stop treating each of these as a separate piece of folklore and start treating them as what they are: consequences of four or five definitions, each of which pays for itself many times over.

The ordering below is the order in which the Putnam actually needs them. Sections 2 to 8 are the constantly recurring machinery and you should be fluent in every line of them. Sections 9 to 13 are common but specialised. Section 14 is the standing recursion section: every recursive, inductive or self-similar structure the corpus threw up, generalised. Section 15 is the exotica, and I say plainly there which items appeared exactly once.

4.2 Divisibility, the division algorithm, Euclid and Bezout

The division algorithm, stated properly

Theorem 4.2.1 (Division algorithm). *For every integer n and every integer $k \geq 1$ there exist unique integers q and r with $n = kq + r$ and $0 \leq r < k$, and then $q = \lfloor n/k \rfloor$ and $r = n - k\lfloor n/k \rfloor$.*

Proof. Existence: the set $\{n - kq : q \in \mathbb{Z},\ n - kq \geq 0\}$ is a nonempty set of nonnegative integers (take q very negative), so by well ordering it has a least element $r = n - kq$. If $r \geq k$ then $r - k = n - k(q+1)$ is a smaller nonnegative member, a contradiction, so $0 \leq r < k$. Uniqueness: if $kq + r = kq' + r'$ with both remainders in $[0, k)$ then $k \mid r - r'$ and $|r - r'| < k$, so $r = r'$ and hence $q = q'$. The identification $q = \lfloor n/k \rfloor$ is the definition of the floor applied to n/k. □

Two things in that statement are load bearing and A-level never says either. First, *uniqueness*. To identify b mod c with a candidate r you must verify two things, the congruence and the size bound $0 \leq r < c$; a congruence alone does not name a remainder. Putnam 2009 B6 is a construction problem in which every single reduction step has to be certified this way, and a write up that gives the congruence without the bound is not a proof. Second, *no sign hypothesis on n*. Putnam 2007 B5 needs $\lfloor n/k \rfloor$ for possibly negative n, and the whole point is that $n - k\lfloor n/k \rfloor$ lands in $[0, k)$ regardless of the sign of n, so the values $(n - j)/k$ for $j = 0, \ldots, k-1$ contain exactly one integer.

The uniqueness clause has a second face that you will use constantly, and it is the single most economical principle in elementary number theory.

Lemma 4.2.2 (Size beats modulus). *If $q \mid N$ and $|N| < q$ then $N = 0$. More generally, if $a \equiv b \pmod{q}$ and a, b both lie in an interval of length less than q, then $a = b$.*

Proof. N is a multiple of q of absolute value below q, so it is the multiple $0 \cdot q$. For the second form apply the first to $N = a - b$. □

This is how you convert a congruence, which carries no order information, into an equality, which does. Putnam 2012 A4 uses it to upgrade a congruence between two cross products into genuine collinearity of lattice points. Putnam 2025 B5 uses it twice: two integers in $\{1, \ldots, p-1\}$ sum to something in $\{2, \ldots, 2p-2\}$, and the only member of that range congruent to 1 modulo p is $p + 1$, so a congruence about $m + I(m)$ becomes the exact equation $m + I(m) = p + 1$. Putnam 2007 B1 uses the degenerate case: if $0 < r < m$ then $r \not\equiv 0 \pmod{m}$, because the least positive integer congruent to 0 is m itself. Putnam 2014 B6 uses it in the shape "an integer of absolute value strictly less than 1 is zero".

- **WHEN** you have proved a congruence but the conclusion you want is an equation or an ordering **DO** bound both sides into an interval shorter than the modulus and invoke uniqueness of the remainder **BECAUSE** a congruence is a statement in $\mathbb{Z}/q$, which has no order and no size; only a size bound can lift it back to $\mathbb{Z}$.
- **WHEN** a problem asks you to show some integer quantity vanishes **DO** find a modulus q dividing it and prove the quantity is smaller than q in absolute value **BECAUSE** 0 is the only multiple of q inside $(-q, q)$, so divisibility plus a size bound is a proof of vanishing that costs two lines instead of an identity hunt.

Greatest common divisors, Euclid, Bezout

Definition 4.2.3. For integers a, b not both zero, $\gcd(a, b)$ is the largest positive integer dividing both. Equivalently, and far more usefully, it is the unique positive generator of the set $a\mathbb{Z} + b\mathbb{Z} = \{ax + by : x, y \in \mathbb{Z}\}$.

Take the second description as the definition and everything falls out, in exactly the way the centre falls out of the definition of a circle. The set $I = a\mathbb{Z} + b\mathbb{Z}$ is closed under subtraction, so by the division algorithm its least positive element d divides every element of I: if $c \in I$ then $c = qd + r$ with $0 \leq r < d$ and $r = c - qd \in I$, forcing $r = 0$. Since $a, b \in I$, d is a common divisor. Since $d \in I$, $d = ax + by$ for some integers x, y, so every common divisor of a and b divides d. Hence d is the greatest common divisor and it is simultaneously an integer combination. That last sentence is Bezout's identity, and it is a theorem you get for free from the right definition rather than a separate fact to memorise.

Theorem 4.2.4 (Bezout)**.** *There exist integers x, y with $ax + by = \gcd(a, b)$. In particular $\gcd(m, n) = 1$ if and only if $mx + ny = 1$ is solvable, if and only if m is invertible modulo n.*

Corollary 4.2.5 (Euclid's lemma and the shift rule)**.** *$\gcd(a, b) = \gcd(a, b - qa)$ for every integer q; every common divisor of a and b divides every integer combination of them; if $\gcd(m, n) = 1$ and $n \mid mk$ then $n \mid k$; if $u \mid m$, $v \mid m$ and $\gcd(u, v) = 1$ then $uv \mid m$.*

The shift rule is the Euclidean algorithm: replace the pair (a, b) by $(b \bmod a, a)$ and recurse. The extended version tracks the coefficients and returns x and y as well as d. You must be able to run it by hand at speed, because a modular inverse is a Putnam subroutine, not a Putnam problem: Putnam 2015 B2 needs $3q \equiv 200 \pmod{1000}$ solved, Putnam 2018 A4 needs the residue x with $mx \equiv 1 \pmod{n}$ as the index shift that is the entire mechanism of the ideal solution.

Putnam 2025 A1 is the cleanest demonstration of the shift rule in the corpus. You are handed a recursion on pairs and you need to control $d_k = \gcd(2m_k + 1, 2n_k + 1)$. Do not expand anything. Both arguments are odd, so their difference $2(m_k - n_k)$ is a multiple of d_k; but d_k is odd, so $d_k \mid m_k - n_k$; and therefore $\gcd(2m + 1, 2n + 1) = \gcd(2m + 1, m - n)$. That single line links the gcd to the difference, and the difference obeys $\delta_{k+1} = 2\delta_k / d_k$, which is a recursion you can actually solve. Putnam 2009 A4 uses the same shift rule in the form $\gcd(p - q, q) = \gcd(p, q)$ to prove that a rule producing $q^2/(p(p - q))$ from a reduced p/q produces something already reduced, which is what makes "numerator modulo denominator" a well defined invariant rather than an artefact of how you wrote the fraction.

- **WHEN** a gcd appears whose arguments are large or are given by a recursion **DO** apply $\gcd(a, b) = \gcd(a, b - qa)$ to reduce to the smallest expression you can reach, and pull out any parity or unit factor first **BECAUSE** the gcd only sees the lattice $a\mathbb{Z} + b\mathbb{Z}$, so any combination of the arguments is fair game; expanding the arguments costs ink and destroys the structure.
- **WHEN** you need to divide by m in a congruence modulo n with $\gcd(m, n) = 1$ **DO** run the extended Euclidean algorithm to get x with $mx \equiv 1$, and reindex by x rather than clearing denominators **BECAUSE** invertibility modulo n is exactly Bezout, and reindexing by a unit is a bijection, so no information and no solutions are lost.
- **WHEN** two coprime numbers each divide the same quantity **DO** conclude their product divides it **BECAUSE** $\operatorname{lcm}(u, v) = uv/\gcd(u, v)$, so coprimality lets you check divisibility one coprime factor at a time; Putnam 2025 B3 assembles $n = 3^a 5^b d$ exactly this way.

4.3 Unique factorisation and the p-adic valuation

The theorem and the object it creates

Theorem 4.3.1 (Fundamental theorem of arithmetic)**.** *Every positive integer is a product of primes, uniquely up to order. Consequently every positive rational is uniquely a finite product $\prod_p p^{e_p}$ with integer exponents e_p, positive or negative, and all but finitely many zero.*

Existence is strong induction on the integer. Uniqueness is Euclid's lemma: if $p \mid ab$ and $p \nmid a$ then $\gcd(p, a) = 1$, so $p \mid b$ by the corollary above; induct on the number of factors. Note where the

work is. Existence is trivial; uniqueness is where Bezout is spent. That is why unique factorisation fails in rings without a division algorithm, a point that becomes live in Section 10.

The object the theorem creates, and the real reason you care, is the valuation.

Definition 4.3.2. For a prime p and a nonzero rational x, the p-adic valuation $v_p(x)$ is the exponent of p in the factorisation of x. Set $v_p(0) = +\infty$.

Theorem 4.3.3 (Properties of v_p)**.** *For nonzero x, y:*

$$v_p(xy) = v_p(x) + v_p(y), \qquad v_p(x+y) \geq \min(v_p(x), v_p(y)),$$

with equality in the second whenever $v_p(x) \neq v_p(y)$. Also $v_p(\gcd(a,b)) = \min(v_p(a), v_p(b))$ and $v_p(\operatorname{lcm}(a,b)) = \max(v_p(a), v_p(b))$, so $\gcd(a,b)\operatorname{lcm}(a,b) = ab$ because $\min(s,t) + \max(s,t) = s + t$ at every prime at once.

Proof. Multiplicativity is unique factorisation. For the ultrametric inequality write $x = p^a u$, $y = p^b w$ with u, w of valuation zero and $a \leq b$; then $x + y = p^a(u + p^{b-a}w)$ and the bracket is a rational of nonnegative valuation, so $v_p(x+y) \geq a$. If $a < b$ then the bracket is u plus something of positive valuation, hence has valuation exactly zero, giving equality. □

That equality clause is worth more than the inequality. It is what turns valuations into a decision procedure rather than a bound. Its sharpest form is the following lemma, which is the closed form of infinite descent and which you should install as a reflex.

Lemma 4.3.4 (Tie at the minimum)**.** *If $x + y = z$ with x, y, z nonzero integers and p is prime, then $\min(v_p(x), v_p(y), v_p(z))$ is attained by at least two of the three.*

Proof. If $v_p(x) < v_p(y)$ then by the equality clause $v_p(z) = v_p(x)$ and the minimum is attained at x and z. If $v_p(x) = v_p(y)$ the minimum is already attained twice. The remaining case is symmetric. □

Putnam 2024 A1 asks for the integers $n \geq 2$ such that $2a^n + 3b^n = 4c^n$ has a solution in positive integers, and this lemma *is* the solution. Look at the 2-adic valuations of the three terms: $1 + nv_2(a)$, $0 + nv_2(b)$, $2 + nv_2(c)$. They are congruent to $1, 0, 2$ modulo n, which are pairwise distinct once $n \geq 3$, so no two can be equal, so the lemma is violated and there is no solution. For $n = 2$ the residues $1, 0, 2$ modulo 2 collide, so you move to $p = 3$: the coefficient exponents are $0, 1, 0$, forcing the tie between the a term and the c term, and after stripping the common power of 3 the reduced equation reads $2 \equiv 1 \pmod 3$ against the residue table of squares. The general principle is worth stating on its own.

Proposition 4.3.5 (Homogeneous valuation obstruction)**.** *In $\sum_i k_i t_i^n = 0$ with fixed integer coefficients k_i, we have $v_p(k_i t_i^n) \equiv v_p(k_i) \pmod n$. If the residues $v_p(k_i) \bmod n$ are pairwise distinct for some prime p, there is no solution in nonzero integers. If exactly two tie, strip the common power of p from the tied terms and finish with the residue table of nth powers modulo p.*

Notice what I did not do: I did not attempt a descent. The nominal solution to a homogeneous Diophantine equation is to assume a primitive solution and manufacture a smaller one. The ideal solution is to observe that descent is nothing but repeated division by p, and that the fixed point of that process is the statement about valuations. Compare the valuations once and the descent never happens. That reroute costs three lines instead of a page.

- **WHEN** a Diophantine equation is homogeneous of degree n with distinct small integer coefficients **DO** compute v_p of each term for a prime dividing some coefficient, reduce the coefficient exponents modulo n, and look for pairwise distinctness **BECAUSE** the term valuations are pinned modulo n by the coefficients alone; a sum of three nonzero things cannot have a unique minimum valuation.
- **WHEN** you catch yourself starting an infinite descent **DO** stop and ask what quantity the descent is decreasing, then compute it directly **BECAUSE** descent is division by p repeated, and v_p is the closed form of that repetition; the closed form is always cheaper than the loop.

Squares, squarefree parts and the $\mathbb{F}_2$ reading

Proposition 4.3.6. *A positive integer k is a perfect square if and only if $v_p(k)$ is even for every prime p. Every k factors uniquely as $k = u^2 d$ with d squarefree, and d is the squarefree kernel. The map $k \mapsto (v_p(k) \bmod 2)_p$ turns multiplication of positive integers into addition in the $\mathbb{F}_2$ vector space indexed by primes, with squares exactly the zero vector.*

Three immediate consequences you will use in the arena. The product of two squares is a square, because even plus even is even. If N is a square and v^2 exactly divides N then N/v^2 is a square, which is the cancellation lemma that Putnam 2013 A2 uses at every step of its minimality argument. If k^3 is a square then k is a square, because $3v_p(k)$ even forces $v_p(k)$ even, which is the last line of Putnam 2021 A3. And a product of pairwise distinct primes each to the first power is never a rational square, which is the contradiction that closes Putnam 2014 B3, using also that a positive integer equal to the square of a rational is the square of an integer.

The $\mathbb{F}_2$ reading is not decoration. It says that "choose a subset of these numbers whose product is a square" is a linear algebra problem over $\mathbb{F}_2$, which is precisely how the quadratic sieve factors integers. Putnam 2013 A2 is that structure in miniature.

The 2-adic valuation and the odd part

Every positive integer is uniquely $n = 2^{v_2(n)} \cdot \mathrm{odd}(n)$. Both factors are multiplicative, and an odd number divides $2^k n$ exactly when it divides $\mathrm{odd}(n)$. This tiny decomposition carries an astonishing amount of the corpus.

The odd divisors of n are exactly the divisors of $\mathrm{odd}(n)$, so n has exactly one odd divisor precisely when n is a power of two (Putnam 2009 B3). Factorisations of n into two factors of prescribed opposite parity biject with divisors of the odd part, because the whole power of two is forced into one side (Putnam 2017 B2). The set of pairs (k, m) with $k2^m = N$ is a chain of $v_2(N) + 1$ pairs (Putnam 2016 B6). Choosing $d = 2^{v_2(n)}$ is exactly what makes n/d odd (Putnam 2010 A4). And in Putnam 2025 A1 the odd part is the monovariant: the doubling in the recursion lives entirely in the 2-part while the divisions live entirely in the odd part, so the odd part of $|m_k - n_k|$ never increases and the product of all the gcds divides the odd part of the initial difference. That is the whole problem, and it is invisible unless you have trained yourself to split off the 2-part on sight.

- **WHEN** a divisibility or counting question distinguishes odd from even divisors, or a parity condition sits on a factorisation **DO** write $n = 2^{v_2(n)}\mathrm{odd}(n)$ immediately and re-express the condition in terms of divisors of the odd part **BECAUSE** the power of two is forced by parity constraints, so all the freedom is in the odd part; splitting removes the constraint instead of carrying it.
- **WHEN** a process doubles a quantity and divides it by odd factors **DO** track the odd part as the monovariant and the exponent of 2 separately **BECAUSE** v_2 and odd are independent multiplicative coordinates, so a process that only touches one of them is monotone in the other.

Valuations of factorials and binomials: Legendre and Kummer

Theorem 4.3.7 (Legendre's formula)**.** *For a prime p and $n \geq 1$,*

$$v_p(n!) = \sum_{i \geq 1} \left\lfloor \frac{n}{p^i} \right\rfloor = \frac{n - s_p(n)}{p - 1},$$

where $s_p(n)$ is the digit sum of n written in base p.

Proof. Count the multiples of p in $[1, n]$, then the multiples of p^2, and so on; each multiple of p^i but not p^{i+1} is counted exactly i times, which is its valuation. For the closed form write $n = \sum_j a_j p^j$; then $\lfloor n/p^i \rfloor = \sum_{j \geq i} a_j p^{j-i}$, and summing over $i \geq 1$ gives $\sum_j a_j (p^j - 1)/(p-1) = (n - s_p(n))/(p-1)$. □

Putnam 2016 A1 asks for the least j such that 2016 divides the jth derivative of every integer polynomial at every integer, and Legendre is how you find and cross check the least j with 2^5, 3^2 and 7 all dividing $j!$. Putnam 2022 B1 needs only the corollary $v_2(n!) = n - s_2(n) < n$, which says $2^n/n!$ has even numerator in lowest terms for $n \geq 1$. Putnam 2014 A1 needs only the qualitative half: $n!$ is divisible by exactly the primes up to n.

Theorem 4.3.8 (Kummer). *$v_p\binom{a+b}{a}$ equals the number of carries when a and b are added in base p. In particular, for $1 \leq t \leq p^m - 1$, $v_p\binom{p^m}{t} = m - v_p(t)$, so p divides every intermediate binomial coefficient at a prime power.*

The prime power case is immediate from $t\binom{p^m}{t} = p^m \binom{p^m - 1}{t-1}$ together with $\binom{p^m - 1}{j} \equiv (-1)^j \pmod{p}$. Putnam 2012 A5 chooses $k = p^{n-1}$ precisely so that every surviving binomial coefficient vanishes modulo p, which forces an orbit to close far too early and produces the contradiction.

Lifting the exponent

Theorem 4.3.9 (LTE). *Let p be an odd prime with $p \mid x - y$ and $p \nmid xy$. Then $v_p(x^n - y^n) = v_p(x - y) + v_p(n)$. For $p = 2$ with x, y odd and n even, $v_2(x^n - y^n) = v_2(x - y) + v_2(x + y) + v_2(n) - 1$.*

You should be able to reconstruct the $p = 2$ statement from scratch even if you forget the general one, because the reconstruction is one line: $x^2 - y^2 = (x - y)(x + y)$, and for odd x, y exactly one of $x - y, x + y$ is $\equiv 2 \pmod 4$. Putnam 2025 A6 uses exactly this in the iterated form: if B is odd with $v_2(B - 1) \geq 2$ then $v_2(B^2 - 1) = v_2(B - 1) + 1$, so $v_2(B^{2^j} - 1) = v_2(B - 1) + j$; the derivative of an iterate squares at each doubling of the index and LTE reads off the valuation exactly. Putnam 2023 B2 uses LTE in the order-lifting form $\operatorname{ord}_{p^e}(g) = d\, p^{\max(0, e - v_p(g^d - 1))}$ where $d = \operatorname{ord}_p(g)$, which is what produces $\operatorname{ord}_{289}(2) = 136$. Putnam 2020 A1 uses the repunit case $v_3(R_k) = v_3(k)$, which is LTE for $p = 3$ applied to $(10^k - 1)/(10 - 1)$.

- **WHEN** you must compute or bound v_p of $a^n \pm b^n$ where p already divides $a \mp b$ **DO** apply LTE; do not expand the geometric factor **BECAUSE** the factorisation $a^n - b^n = (a - b) \sum a^i b^{n-1-i}$ has a second factor congruent to nb^{n-1} modulo p, so all the extra p-content is $v_p(n)$ and nothing else.
- **WHEN** a doubling recursion is claimed to gain exactly one power of 2 per step **DO** prove it with the $p = 2$ LTE one-step identity rather than by an induction on explicit formulae **BECAUSE** $B^2 - 1 = (B - 1)(B + 1)$ with $B + 1 \equiv 2 \pmod 4$ is an exact statement, so the induction closes with equality rather than an inequality.

Divisor counts

If $N = \prod p_i^{e_i}$ then $\tau(N) = \prod (e_i + 1)$, because a divisor is a choice of exponent for each prime. Hence $\tau(N) = 2$ characterises primality (Putnam 2017 B2), and enumerating the nine divisors of $2018^2 = 2^2 \cdot 1009^2$ systematically rather than by trial is what makes Putnam 2018 A1 a five minute problem. To confirm a factorisation is complete you check no prime up to $\lfloor \sqrt{N} \rfloor$ divides N: for 1009 that means primes up to 31, for 2017 up to 43. Say so in the write up. A grader who cannot see that you verified primality will assume you did not.

4.4 Congruences and the Chinese remainder theorem

Congruence as a ring structure

Definition 4.4.1. $a \equiv b \pmod m$ means $m \mid a - b$. The classes form the ring $\mathbb{Z}/m\mathbb{Z}$, and the reduction map $\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ is a ring homomorphism.

The first move in an enormous fraction of arithmetic problems is the translation $a \mid b \iff b \equiv 0 \pmod a$. It looks like notation. It is not: it converts a statement about the existence of a hidden cofactor into a computation in a finite ring, and finite rings can be exhausted. Putnam 2007 B1 opens with exactly this translation and closes three lines later.

Theorem 4.4.2 (Congruence compatibility of polynomials)**.** *If $f \in \mathbb{Z}[x]$ and $a \equiv b \pmod m$ then $f(a) \equiv f(b) \pmod m$. Equivalently $(u - v) \mid f(u) - f(v)$ for all integers u, v, and f descends to a well defined map on $\mathbb{Z}/m\mathbb{Z}$.*

Proof. Evaluation of f is built only from addition and multiplication of the argument with fixed integers, and reduction modulo m respects both. For the factor form, $u^k - v^k = (u - v)\sum u^i v^{k-1-i}$, so $u - v$ divides $f(u) - f(v)$ termwise. □

This one engine solves Putnam 2007 B1 outright: $f(f(n) + 1) \equiv f(1) \pmod{f(n)}$, because $f(n) + 1 \equiv 1$, and a general-degree problem collapses to a one-line congruence. Putnam 2008 B4 uses the factor form to extend $p \nmid h'(x)$ from the finite range $0 \le x \le p - 1$, where the hypothesis has bite, to every integer.

A related principle: if a condition on integers is phrased purely through divisibility by a fixed m, it depends only on residues, so you may replace the integers by their residue multiset before counting anything. Putnam 2007 A3 replaces $1, \dots, 3k + 1$ by k zeros, $k + 1$ ones and k twos before a single count is performed. Any m consecutive integers form a complete residue system modulo m; a list of m integers pairwise distinct modulo m is a complete residue system, and for maps of finite sets of equal size injective, surjective and bijective coincide (Putnam 2008 B4, Putnam 2017 B6).

Splitting a modulus: the Chinese remainder theorem

Theorem 4.4.3 (CRT)**.** *If $m_1, \dots, m_r$ are pairwise coprime with product M, the map $\mathbb{Z}/M\mathbb{Z} \to \prod_i \mathbb{Z}/m_i\mathbb{Z}$ is a ring isomorphism, and it restricts to a group isomorphism on unit groups $(\mathbb{Z}/M\mathbb{Z})^\times \cong \prod_i (\mathbb{Z}/m_i\mathbb{Z})^\times$.*

Proof. The map is a ring homomorphism with kernel the classes divisible by every m_i, hence by M by coprimality, so it is injective; both sides have M elements, so it is bijective. A ring isomorphism carries units to units. □

Three distinct uses, all of them frequent.

Divisibility splits. $N \mid X$ if and only if $p^e \mid X$ for every prime power p^e exactly dividing N, and these conditions are independent. Putnam 2016 A1 splits "$2016 \mid j!$" into conditions at 2, 3 and 7. Putnam 2020 A1 splits $2020 = 2^2 \cdot 5 \cdot 101$ and finds that the 4 and 5 conditions are carried by the factor 10^b and the 101 condition by the repunit, decoupling the two parameters completely. Putnam 2021 A5 splits $2021 = 43 \cdot 47$. Putnam 2023 B2 splits $2023 = 7 \cdot 17^2$.

Counting. Each unit class modulo m_1 is realised by exactly $\varphi(m_2)$ reduced residues modulo $m_1 m_2$. Putnam 2021 A5 turns a power sum over the reduced residues modulo 2021 into $46\sum_{a=1}^{42} a^j$ modulo 43 and $42\sum_{b=1}^{46} b^j$ modulo 47 by exactly this multiplicity count, and the multiplicities 46 and 42 are not decoration: you must check they are invertible modulo the respective prime or the condition at that prime becomes vacuous.

Prescribing independently. Putnam 2023 B5 constructs a hostile multiplier m that is a generator on one prime power factor and the identity on the rest, which is only possible because CRT lets you prescribe m on each factor separately. Putnam 2023 B2 prescribes an exponent modulo 3 and modulo 272 independently. Putnam 2015 B2 intersects $q \equiv 200 \cdot 3^{-1} \pmod{1000}$ with $q \equiv 2 \pmod 3$.

- **WHEN** a composite modulus appears, especially one that factors as a year number **DO** factor it and split the problem into independent prime power conditions before doing anything else **BECAUSE** $\mathbb{Z}/M$ is literally the product ring, so a condition modulo M is a conjunction of conditions modulo the prime powers, and the conjuncts are usually easier and often decoupled.
- **WHEN** you must construct an integer with prescribed multiplicative behaviour on several primes at once **DO** prescribe it separately on each prime power and glue by CRT **BECAUSE** surjectivity of the CRT map means any tuple of local prescriptions is realised by a global integer.

Linear congruences and residue tables

$ra \equiv b \pmod q$ is solvable in a if and only if $\gcd(q,r) \mid b$, and then the solutions form one class modulo $q/\gcd(q,r)$ (Putnam 2012 A4). If $\gcd(a,m) = 1$ the solution is unique modulo m and equals $a^{-1}b$. More generally the map $b \mapsto ab$ on $\mathbb{Z}/n$ has image of size $n/\gcd(a,n)$ with every value hit $\gcd(a,n)$ times, and is a bijection exactly when $\gcd(a,n) = 1$ (Putnam 2015 A3, Putnam 2024 B1). That last is why "$i \mapsto 2i$ is a bijection of $\mathbb{Z}_n$" is equivalent to n odd, which is the hinge of both directions of Putnam 2024 B1.

The cheapest weapon in the whole subject is the residue table. Squares are $\{0,1\}$ modulo 3 and modulo 4, $\{0,1,4\}$ modulo 8, and $\{0,1,4\}$ modulo 5; odd squares are $\equiv 1 \pmod 8$ (Putnam 2011 B2); cubes are $\{0,1,-1\}$ modulo 7 and modulo 9 (Putnam 2019 A1); nth powers modulo p take at most $1 + (p-1)/\gcd(n,p-1)$ values. A local obstruction, meaning no solution modulo some p^k, is a complete proof of no integer solution, and for a homogeneous equation you may always normalise to a primitive solution first, which is the same as fixing the minimum valuation at 0 (Putnam 2024 A1).

Two further translation devices. A base b representation makes "N ends in the digit string D" the congruence $N \equiv D \pmod{b^{|D|}}$, which is how "ends in 2015" becomes $s_N \equiv 2015 \pmod{10^4}$ in Putnam 2015 B2. And the divisibility tests generalise: if $N = \sum d_i B^i$ then $N \equiv \sum d_i \pmod{B-1}$ and $N \equiv \sum (-1)^i d_i \pmod{B+1}$. Putnam 2010 A4 is precisely the alternating digit sum test run in base $B = 10^{2^m}$, where the number in question has digits 1, 1 and k copies of $B-1$.

- **WHEN** an equation in integers resists direct attack, or you suspect there is no solution **DO** test it against small moduli, choosing the modulus so that the powers present have small image: modulo 8 or 3 or 4 for squares, modulo 9 or 7 for cubes, modulo p with $p \equiv 1 \pmod n$ for nth powers **BECAUSE** the image of $x \mapsto x^n$ modulo p has size about $(p-1)/\gcd(n,p-1)$, so a modulus with $\gcd(n,p-1)$ large makes the table short and the obstruction visible.
- **WHEN** a condition is about the last few digits, or about digit sums, in some base **DO** write it as a congruence modulo b^k, $b-1$ or $b+1$ as appropriate **BECAUSE** positional notation is a polynomial evaluated at b, so every digit statement is a congruence statement about that polynomial.

4.5 Fermat, Euler, Wilson and the multiplicative order

The three theorems

Theorem 4.5.1 (Euler, with Fermat as the case $m = p$)**.** *If* $\gcd(c,m) = 1$ *then* $c^{\varphi(m)} \equiv 1 \pmod m$, *where* $\varphi(m) = |(\mathbb{Z}/m\mathbb{Z})^\times|$. *For prime* p: $c^{p-1} \equiv 1$ *for* $p \nmid c$, *and* $c^p \equiv c$ *for all* c.

Proof. $(\mathbb{Z}/m\mathbb{Z})^\times$ is a finite abelian group of order $\varphi(m)$; multiplication by c permutes it, so $\prod_u u = \prod_u (cu) = c^{\varphi(m)} \prod_u u$, and cancelling the unit $\prod_u u$ gives the result. This is Lagrange's theorem in the concrete case, and you should know both proofs. □

A-level gives you nothing here, and Fermat alone is not enough: Putnam 2025 B3 needs $d \mid 135^k - 1$ whenever $\varphi(d) \mid k$, for d an arbitrary integer coprime to 15, and Fermat would require d prime. Euler also gives you division for free: $x^{-1} \equiv x^{\varphi(m)-1}$, which is how Putnam 2009 B6 realises a quotient when only remainders are available.

Theorem 4.5.2 (Wilson and its converse)**.** *For a prime p, $(p-1)! \equiv -1 \pmod{p}$. Conversely, if N is composite and $N > 4$ then $N \mid (N-1)!$.*

Proof. In $(\mathbb{Z}/p)^\times$ pair each element with its inverse; the only self-paired elements solve $x^2 = 1$, and over a field that means $x = \pm 1$, so the product of all elements is $1 \cdot (-1) = -1$. For the converse, let N be composite with smallest prime factor q and write $N = qs$. If $q < s$ then q and s are distinct factors below N and both appear in $(N-1)!$. If $q = s$ then $N = q^2$ with $q \geq 3$, and q and $2q$ are distinct factors below N, so $q^2 \mid (N-1)!$. Only $N = 4$ escapes, since $4 \nmid 6$. □

Putnam 2014 A1 is exactly the statement that a composite $m \neq 4$ divides $(m-1)!$, and the proof above, via the smallest prime factor trichotomy, is what turns an assertion into a proof. Putnam 2021 B4 uses Wilson after symmetrising a product with varying exponents: reindexing $k \mapsto p-k$ in $A = \prod_{k=1}^{p-1} k^k$ gives $A \equiv (-1)^{p(p-1)/2} \prod k^{p-k}$, and multiplying the two expressions produces $A^2 \equiv (-1)^{p(p-1)/2}((p-1)!)^p \equiv -1$ when $p \equiv 1 \pmod 4$. Putnam 2011 B6 uses the reflection corollary $k!\,(p-1-k)! \equiv (-1)^{k+1} \pmod{p}$, which converts a coefficient list $k!$ into the reciprocal list $1/k!$ and turns a mysterious polynomial into the truncated exponential.

Order: the single most useful multiplicative object

Definition 4.5.3. For $\gcd(a, m) = 1$, $\operatorname{ord}_m(a)$ is the least $d \geq 1$ with $a^d \equiv 1 \pmod{m}$.

Theorem 4.5.4. *$\{k \in \mathbb{Z} : a^k \equiv 1\}$ is a subgroup of $\mathbb{Z}$, hence equal to $d\mathbb{Z}$ with $d = \operatorname{ord}_m(a)$. Consequently $m \mid a^k - 1$ if and only if $d \mid k$, $a^s \equiv a^t$ if and only if $s \equiv t \pmod{d}$, and $d \mid \varphi(m)$.*

Proof. The set is closed under subtraction because a is invertible, so it is a subgroup of $\mathbb{Z}$, and every subgroup of $\mathbb{Z}$ is $d\mathbb{Z}$ for its least positive element. Divisibility of $\varphi(m)$ is Euler plus the previous sentence. □

That theorem is the reason exponents in arithmetic problems live modulo something. Once you see a varying exponent, you ask for the order. Putnam 2020 A1: $101 \mid R_k$ iff $101 \mid 10^k - 1$ iff $4 \mid k$, because $10^2 \equiv -1 \pmod{101}$ forces $\operatorname{ord}_{101}(10) = 4$. Putnam 2010 A4: modulo $M = a^d + 1$ one has $a^d \equiv -1$, so $\operatorname{ord}_M(a) = 2d$ and $a^{dj} \equiv (-1)^j$; every congruence in that problem is the sign character applied to three exponents written as multiples of $d = 2^m$. Putnam 2023 B2: the binary weight question becomes conditions on exponents modulo $\operatorname{ord}_7(2) = 3$ and $\operatorname{ord}_{289}(2) = 136$.

Two corollaries that come up repeatedly.

Proposition 4.5.5. *For integers $a \geq 2$ and $m, n \geq 1$: $a^m - 1 \mid a^n - 1$ if and only if $m \mid n$; more precisely $\gcd(a^m - 1, a^n - 1) = a^{\gcd(m,n)} - 1$. Also, for odd k, $x + 1 \mid x^k + 1$, so $a^d + 1 \mid a^n + 1$ whenever n/d is an odd integer; hence $a^n + 1$ can be prime only if n is a power of two.*

Proof. Write $n = qm + r$ with $0 \leq r < m$; then $a^n \equiv a^r \pmod{a^m - 1}$, and $0 \leq a^r - 1 < a^m - 1$, so divisibility forces $r = 0$. The gcd statement is the same computation run as a Euclidean recursion on the exponents. The odd case follows from $x^k + 1 = (x+1)(x^{k-1} - x^{k-2} + \cdots + 1)$. □

Putnam 2018 B3 applies the Mersenne half twice, once to each of two remaining divisibility conditions, and each application reduces to the same test one level down. Putnam 2010 A4 applies the Fermat-number half to strip powers of two out of an exponent. Note the recursion buried in the gcd identity; Section 14 takes that up properly.

Proposition 4.5.6. *For an odd prime p with $p \nmid g$, some power of g is $\equiv -1 \pmod p$ if and only if $\operatorname{ord}_p(g)$ is even.*

Proof. -1 is the unique element of order 2 in the cyclic group $\langle g \rangle$, and a cyclic group of order d contains an element of order 2 exactly when $2 \mid d$. □

Putnam 2023 B2 uses this as its entire lower bound: binary weight two would require $2^c \equiv -1 \pmod 7$, and $\operatorname{ord}_7(2) = 3$ is odd, so it is impossible.

Power sums over the residues

Theorem 4.5.7. *For a prime p and an integer $j \geq 0$,*

$$\sum_{n=1}^{p-1} n^j \equiv \begin{cases} -1 \pmod p & \text{if } (p-1) \mid j, \\ 0 \pmod p & \text{otherwise.} \end{cases}$$

Proof. If $(p-1) \mid j$ every term is 1 by Fermat and there are $p - 1 \equiv -1$ terms. Otherwise take a primitive root g (Section 6); the sum is the geometric series $\sum_{i=0}^{p-2} (g^j)^i = (g^{j(p-1)} - 1)/(g^j - 1)$, whose numerator vanishes and whose denominator does not, since $g^j \neq 1$ when $(p-1) \nmid j$. □

This is the engine of Putnam 2021 A5, applied at $p = 43$ and $p = 47$, and it powers the substitution route in Putnam 2019 A5. It is also the reason the answer in 2021 A5 is a condition on j modulo 42 and modulo 46 rather than anything about 2021 itself. A companion one-liner from the same problem: for odd m the reduced residues pair as $\{n, m-n\}$ and $(m-n)^j \equiv (-1)^j n^j$, so for odd j the whole power sum vanishes modulo m immediately.

- **WHEN** an exponent varies and a fixed modulus is in play **DO** compute the multiplicative order and reduce the exponent modulo it **BECAUSE** a^s depends only on s modulo $\operatorname{ord}_m(a)$, so an infinite family of exponents collapses to finitely many cases.
- **WHEN** you meet $a^n \pm 1$ inside a divisibility question **DO** use $\gcd(a^m - 1, a^n - 1) = a^{\gcd(m,n)} - 1$ and the odd-cofactor rule for the plus sign **BECAUSE** these are the closed forms of the Euclidean recursion on exponents, and they turn a divisibility question about huge numbers into one about their indices.
- **WHEN** a product over all nonzero residues, or a factorial near a prime, appears **DO** invoke Wilson, and if the exponents vary, reindex $k \mapsto p - k$ and multiply the two expressions **BECAUSE** pairing with inverses is the only structural handle on such a product, and $k \mapsto p - k$ is the only other involution available.

4.6 Primitive roots and the structure of the unit group

Theorem 4.6.1. *$(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic of order $p - 1$ for every prime p. A generator is called a primitive root.*

Proof. A polynomial of degree d over the field $\mathbb{Z}/p$ has at most d roots. If the group had exponent $e < p - 1$ then every element would satisfy $x^e = 1$, giving $p - 1 > e$ roots of a degree e polynomial. So the exponent is $p - 1$, and a finite abelian group whose exponent equals its order is cyclic. □

Fixing a primitive root g gives the discrete logarithm: an isomorphism $\mathbb{Z}/(p-1) \to (\mathbb{Z}/p)^\times$, $i \mapsto g^i$. Under it, $x \mapsto x^k$ becomes $i \mapsto ki$, which is linear. That is the bifurcation point of Putnam 2012 B6: cubing is nonlinear on residues and linear on exponents, and it is bijective exactly when $\gcd(k, p-1) = 1$, which is why $p \equiv 2 \pmod 3$ makes the cube map a permutation. It is also the source of the standard sufficiency test used in Putnam 2024 A4: if $1, a, \ldots, a^m$ are pairwise distinct modulo p and $m \geq (p-1)/2$, then $\operatorname{ord}_p(a) \geq m + 1 > (p-1)/2$, and every proper divisor of $p - 1$ is at most $(p-1)/2$, so a is a primitive root.

Theorem 4.6.2 (Structure of $(\mathbb{Z}/m)^\times$)**.** *For an odd prime p, $(\mathbb{Z}/p^e)^\times$ is cyclic of order $p^{e-1}(p-1)$, and a primitive root modulo p^2 is a primitive root modulo p^h for every h. For $p=2$: trivial modulo 2, cyclic of order 2 modulo 4, and $\mathbb{Z}/2\times\mathbb{Z}/2^{a-2}$ for $a\geq 3$, generated by -1 and 3, so $\operatorname{ord}_{2^a}(3)=2^{a-2}$. In general $(\mathbb{Z}/m)^\times$ is cyclic only for $m\in\{1,2,4,p^e,2p^e\}$ with p an odd prime.*

Know the failure cases as well as the successes. Putnam 2021 A5 must split its argument across 43 and 47 precisely because $(\mathbb{Z}/2021)^\times\cong C_{42}\times C_{46}$ is not cyclic, so there is no single multiplier g to run the geometric series argument at the composite modulus. Putnam 2023 B5 needs the odd prime power case to supply a generator whose order $p^{e-1}(p-1)$ is even, which is the parity obstruction the whole problem rests on. Putnam 2012 B6 needs the 2-power case to show $\varphi(d)/\operatorname{ord}_d(3)$ is even for every d divisible by 4 except $d=4$. Putnam 2009 B6 needs that 2 is a primitive root modulo 9 and therefore modulo every 3^h, so every residue coprime to 3 is a power of 2 there.

One further fact you will need for cycle counting.

Proposition 4.6.3. *Under $x\mapsto mx$ on $\mathbb{Z}/n$ with $\gcd(m,n)=1$, the orbit of x has length $\operatorname{ord}_{n/\gcd(x,n)}(m)$.*

Proof. $m^k x\equiv x \pmod n$ says $n\mid x(m^k-1)$, that is $n/\gcd(x,n)$ divides m^k-1 since $x/\gcd(x,n)$ is coprime to it. □

Writing $x=p^j u'$ with u' a unit reduces an orbit computation on $\mathbb{Z}/p^e$ to one on $\mathbb{Z}/p^{e-j}$, a descent on the p-adic filtration. Putnam 2023 B5 counts exactly k cycles of length $\varphi(p^e)$ this way, an odd count of an even length, which is what its parity lemma forbids.

- **WHEN** a power map, an iterated power map, or a kth power condition sits on a prime modulus **DO** take discrete logarithms: work in $\mathbb{Z}/(p-1)$ where the map becomes multiplication by k **BECAUSE** the exponent group is where power maps are linear, and linear maps on $\mathbb{Z}/(p-1)$ are completely described by $\gcd(k,p-1)$.
- **WHEN** you want to argue at a composite modulus using a generator **DO** first check the unit group is actually cyclic; if it is not, split by CRT and argue at each prime power **BECAUSE** cyclicity fails for every modulus outside $\{1,2,4,p^e,2p^e\}$, and an argument that silently assumes a generator at a composite modulus is simply wrong.

4.7 Quadratic residues and reciprocity

Definition 4.7.1. For an odd prime p and $p\nmid a$, the Legendre symbol $\left(\frac{a}{p}\right)$ is $+1$ if a is a nonzero square modulo p and -1 otherwise; it is 0 if $p\mid a$.

Theorem 4.7.2 (Euler's criterion)**.** $a^{(p-1)/2}\equiv\left(\frac{a}{p}\right) \pmod p$. *Consequently the symbol is completely multiplicative in a, exactly $(p-1)/2$ of the nonzero classes are squares, and $\sum_{a=1}^{p-1}\left(\frac{a}{p}\right)=0$.*

Proof. $(\mathbb{Z}/p)^\times$ is cyclic of order $p-1$ generated by g; $a=g^i$ is a square iff i is even, and $a^{(p-1)/2}=(g^{(p-1)/2})^i=(-1)^i$ since $g^{(p-1)/2}$ is the unique element of order 2, namely -1. Multiplicativity and the count follow from the parity of i. □

Theorem 4.7.3 (Reciprocity and its supplements)**.** *For distinct odd primes p,q:*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right)=(-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}},\qquad\left(\frac{-1}{p}\right)=(-1)^{\frac{p-1}{2}},\qquad\left(\frac{2}{p}\right)=(-1)^{\frac{p^2-1}{8}}.$$

Hence -1 is a square modulo p iff $p\equiv 1 \pmod 4$; 2 is a square iff $p\equiv\pm 1 \pmod 8$; -2 is a square iff $p\equiv 1,3 \pmod 8$; -3 is a square iff $p\equiv 1 \pmod 3$; and for $p\equiv 1 \pmod 4$, $\left(\frac{p}{q}\right)=\left(\frac{q}{p}\right)$.

I am not going to reproduce a proof of reciprocity; the shortest honest route is Gauss's lemma plus a lattice point count, and it is a standard theorem you must know by statement. State it as a named theorem in the arena and no marks are lost. What you cannot do is guess the supplements: memorise them, and memorise them in the useful form, namely as congruence conditions on p modulo 4, 8 and 3.

The Putnam uses are direct. Putnam 2022 A3 needs $x^2 - x - 1$ to split over $\mathbb{F}_p$, that is 5 to be a square; since $5 \equiv 1 \pmod 4$, $\left(\frac{5}{p}\right) = \left(\frac{p}{5}\right)$, so it splits exactly when $p \equiv \pm 1 \pmod 5$. The same problem needs the count of roots of a quadratic over $\mathbb{F}_p$: for $a \neq 0$ and p odd, $ax^2 + bx + c = 0$ has 2, 1 or 0 roots according as the discriminant is a nonzero square, zero, or a non-square. Putnam 2009 A4 uses squares modulo 5 to kill $q^2 \equiv 2, 3$, and uses the first supplement to explain why the modulus had to be 5: the excluded class must be closed under negation, which needs -1 to be a square. Putnam 2021 B4 uses solvability of $x^2 \equiv -1$ exactly for $p \equiv 1 \pmod 4$, with $((p-1)/2)!$ an explicit root. Putnam 2025 B5 needs elements of order 3, 4 and 6 in $\mathbb{F}_p^\times$, which exist iff $3 \mid p-1$ and $4 \mid p-1$ respectively, and this is the same content as $\left(\frac{-3}{p}\right)$ and $\left(\frac{-1}{p}\right)$. Putnam 2012 B6 needs all three of $\left(\frac{-1}{p}\right)$, $\left(\frac{2}{p}\right)$, $\left(\frac{-3}{p}\right)$.

Theorem 4.7.4 (Gauss's lemma, general odd modulus)**.** *Let q be odd, $\gcd(b, q) = 1$, and $A = \{a : 1 \leq a \leq (q-1)/2,\ \gcd(a, q) = 1\}$, so $|A| = \varphi(q)/2$. Reduce each ba into $\{-(q-1)/2, \ldots, (q-1)/2\}$ and let n count the negative representatives. Then $b^{\varphi(q)/2} \equiv (-1)^n \pmod q$. For $q = p$ prime this is the Legendre symbol; for composite q it is a congruence only and is* not *the Jacobi symbol.*

Putnam 2015 A5 is that theorem with $b = 2$, converting a count of coprime residues in an interval into the congruence $(-1)^{N_q} \equiv (-2)^{\varphi(q)/2} \pmod q$. The composite case is where the problem bites: for odd q with at least two distinct prime factors, every $\varphi(p_i^{k_i})$ is even, so $\varphi(p_i^{k_i}) \mid \varphi(q)/2$, hence $b^{\varphi(q)/2} \equiv 1$ at every prime power and therefore modulo q. That collapse is the entire "only if" half, and it is why the answer consists of prime powers with $p \equiv 5, 7 \pmod 8$.

The Jacobi symbol $\left(\frac{a}{n}\right) = \prod \left(\frac{a}{p_i}\right)^{e_i}$ for odd $n > 0$ obeys reciprocity and both supplements verbatim, which makes it the right computational tool: you may reduce and flip without ever factoring. Be precise about what it does not say. $\left(\frac{a}{n}\right) = 1$ does not imply a is a square modulo n. Zolotarev's lemma gives the conceptual reason it exists: for odd n and $\gcd(a, n) = 1$, the sign of the permutation $x \mapsto ax$ of $\mathbb{Z}/n$ equals $\left(\frac{a}{n}\right)$. Putnam 2012 B6 is a problem about the sign of a multiplication-type permutation, so Zolotarev is the theory orbiting it; the modulus there is $p - 1$, which is even, so the lemma does not apply directly, but it tells you immediately why an inversion count produces conditions modulo 4 and modulo 8.

- **WHEN** a square, a discriminant, or a condition of the form "$x^2 \equiv a$ is solvable" appears modulo a prime **DO** compute the Legendre symbol using multiplicativity, the two supplements and reciprocity, and read the answer off as a congruence condition on p **BECAUSE** Euler's criterion makes the symbol a homomorphism, so the whole computation is a chain of flips and reductions with no search.
- **WHEN** you must count something over the units in half a period, or count sign changes of ba modulo q **DO** recognise Gauss's lemma and convert the count into the congruence $b^{\varphi(q)/2} \equiv (-1)^n$ **BECAUSE** the negative representatives are exactly the pairs whose sign flips under multiplication by b, and the product over a half system converts that sign count into a power.

4.8 Arithmetic functions, Dirichlet convolution and Mobius inversion

Definition 4.8.1. $f : \mathbb{N} \to \mathbb{C}$ is multiplicative if $f(1) = 1$ and $f(mn) = f(m)f(n)$ whenever $\gcd(m, n) = 1$; completely multiplicative if that holds for all m, n. The Dirichlet convolution is

$(f * g)(n) = \sum_{d|n} f(d)g(n/d)$.

Theorem 4.8.2. *The convolution of two multiplicative functions is multiplicative. Consequently a multiplicative function is determined by its values on prime powers, and evaluating it at $n = \prod p_i^{e_i}$ is a product of r evaluations rather than a sum over all $\tau(n)$ divisors.*

Proof. If $\gcd(m, n) = 1$ every divisor of mn factors uniquely as $d_1 d_2$ with $d_1 \mid m$, $d_2 \mid n$, and the two parts are coprime; so the double sum factors. □

The standard cast: φ with $\varphi(p^k) = p^{k-1}(p-1)$ and $\sum_{d|n} \varphi(d) = n$; $\tau = 1 * 1$; $\sigma = \mathrm{Id} * 1$; the Mobius function $\mu(1) = 1$, $\mu(d) = (-1)^r$ for a product of r distinct primes, $\mu(d) = 0$ if a square divides d.

Theorem 4.8.3 (Mobius inversion)**.** *$\sum_{d|f} \mu(d)$ equals 1 if $f = 1$ and 0 otherwise, that is $\mu * 1 = \varepsilon$. Hence $g = h * 1$ if and only if $h = g * \mu$: if $g(k) = \sum_{d|k} h(d)$ then $h(k) = \sum_{d|k} \mu(d)g(k/d)$.*

Proof. For $f = \prod_{i=1}^{r} p_i^{e_i}$ with $r \geq 1$, only squarefree divisors contribute, so the sum is $\sum_j \binom{r}{j}(-1)^j = (1-1)^r = 0$. Inversion is then associativity of convolution against $\mu * 1 = \varepsilon$. □

Putnam 2015 A3 is the flagship. The double product reduces, after taking $\log_2$, to Pillai's gcd sum $P(n) = \sum_{a=1}^{n} \gcd(a, n)$. Counting the a with $\gcd(a, n) = d$ as $\varphi(n/d)$ gives $P = \mathrm{Id} * \varphi$, hence multiplicative, hence computable at $2015 = 5 \cdot 13 \cdot 31$ in three evaluations of $P(p) = 2p - 1$ instead of a sum over eight divisors. That is the whole reroute: recognise the convolution, invoke multiplicativity, evaluate on primes.

Putnam 2015 A5 uses Mobius inversion in its most reusable form: a count with a coprimality side condition is the Mobius transform of the count without it,

$$\#\{a \leq y : \gcd(a, q) = 1\} = \sum_{d|q} \mu(d) \left\lfloor \frac{y}{d} \right\rfloor,$$

and modulo 2 the signs disappear and only squarefree divisors survive. Putnam 2023 B6 needs both the inversion itself, to express the all-ones vector in the basis of divisor columns, and the aggregate identity

$$\sum_{m=1}^{n} M\Big(\Big\lfloor \frac{n}{m} \Big\rfloor\Big) = \sum_{dm \leq n} \mu(d) = \sum_{f \leq n} \sum_{d|f} \mu(d) = 1,$$

where M is the Mertens sum. Look at what happened in the middle of that display: a sum over pairs was reordered by grouping on the product $f = dm$ instead of on d or on m. That reordering is the single most valuable move in divisor sums.

Theorem 4.8.4 (Reordering divisor sums, and the hyperbola method)**.** *$\sum_{n \leq x} \sum_{d|n} f(d) = \sum_{d \leq x} f(d)\lfloor x/d \rfloor$. More generally, for $F = f * 1$ style sums, $\sum_{n \leq x}(f * g)(n) = \sum_{d \leq y} f(d)G(x/d) + \sum_{e \leq x/y} g(e)F(x/e) - F(y)G(x/y)$ for any $1 \leq y \leq x$.*

Putnam 2015 B6 is precisely this move: the divisor condition $d < \sqrt{2k}$ is rewritten, by setting $k = dm$, as $d < 2m$, so the whole arithmetic sum becomes a double sum over the lattice half plane $\ell \leq m$, and only then can analysis begin. Any condition on d that mentions k is a condition on the pair $(d, k/d)$; write the pair down and the constraint usually becomes linear.

Finally, the summation tool you need whenever an arithmetic sum meets an analytic weight.

Theorem 4.8.5 (Abel summation)**.** *With $A(x) = \sum_{n \leq x} a_n$ and b differentiable,*

$$\sum_{n \leq x} a_n b(n) = A(x)b(x) - \int_1^x A(t)b'(t)\,dt,$$

and in discrete form $\sum_{n=1}^{N} a_n b_n = A(N)b_N - \sum_{n=1}^{N-1} A(n)(b_{n+1} - b_n)$.

Use it when you know the partial sums of the arithmetic part and the smooth part is monotone: Putnam 2015 B6 and Putnam 2016 B2 both need an asymptotic sandwich rather than an exact formula, and Abel summation is how a divisor count with known average order is turned into an estimate. Putnam 2016 B2 also carries an honest warning worth internalising: an exact local formula does not imply an exact global one, because a cumulative count generally splits the last block partway through. Recognising that early redirects you from a doomed search for a closed form to the asymptotic that the problem actually wants.

Related, and quotable: a Dirichlet series with periodic coefficients of mean zero converges for $s > 0$ and factors through ζ. In Putnam 2017 B4 the relevant $L(s) = -\zeta(s)(1 - 2^{1-s})^2$, and since ζ has a simple pole of residue 1 at $s = 1$ while $(1 - 2^{1-s})^2$ has a double zero with leading coefficient $(\log 2)^2$, the target sum $-L'(1)$ is $(\log 2)^2$. That structural reading tells you in advance whether a given periodic weight will give a clean answer or one contaminated by the Euler to Mascheroni constant.

- **WHEN** a sum runs over $a = 1$ to n with a $\gcd(a, n)$ in it, or over the divisors of n **DO** group the terms by the value of the gcd, recognise the result as a Dirichlet convolution, and use multiplicativity to evaluate on prime powers **BECAUSE** grouping by gcd turns a sum of size n into a sum of size $\tau(n)$, and multiplicativity turns that into a product of r terms.
- **WHEN** a count carries a coprimality condition **DO** apply Mobius inversion to remove the condition, leaving a sum of $\mu(d)$ times the unconditioned count at y/d **BECAUSE** $\mu * 1 = \varepsilon$ is exactly an inclusion and exclusion over the prime divisors, so the coprimality condition is the Mobius transform of no condition at all.
- **WHEN** a double sum runs over pairs (d, m) with dm constrained **DO** reorder by grouping on the product, or split at a threshold and use the hyperbola method **BECAUSE** the constraint region is a hyperbola under the map $(d, m) \mapsto dm$, and both factors are small on one side of the symmetry point.

4.9 Diophantine equations

There are exactly four things you do to a Diophantine equation, and you should run them in this order: factor it, bound it, reduce it modulo something, or descend on it. Factoring is first because it is the only one that produces the complete solution set.

Factor it

$ax + by = c$ is solvable iff $\gcd(a, b) \mid c$, and the general solution is one particular solution from the extended Euclidean algorithm plus $t(b/d, -a/d)$.

A bilinear equation is a rectangle waiting to be completed. From $pxy = q(x + y)$, multiply by p and rearrange to $(px - q)(py - q) = q^2$. This goes under the name Simon's Favourite Factoring Trick, and there is nothing tricky in it: it is completing the square transposed into multiplication, and the reason it is natural is that a bilinear form in two variables becomes a product of two linear forms as soon as you clear the constant, exactly as a quadratic becomes a square as soon as you clear the linear term. Putnam 2018 A1 asks for the pairs with $1/a + 1/b = 3/2018$; the completed form turns it into a divisor-pair search over the nine divisors of 2018^2, filtered by the residue condition modulo 3, with the negative divisor pairs checked and rejected on positivity rather than dismissed on sight.

A difference of squares is the same move. Putnam 2011 B2 reaches $8p = (q - a)(q + a)$; both factors are positive and even, so with $q - a = 2s$, $q + a = 2t$ you get $st = 2p$, and the divisor pairs of $2p$ for prime p are $(1, 2p)$ and $(2, p)$, giving the explicit list $q \in \{2p + 1, p + 2\}$.

The consecutive-sums family is worth memorising outright. N is a sum of $k \geq 2$ consecutive positive integers starting at a iff $2N = k(2a + k - 1)$, and setting $m = 2a + k - 1$ this says $2N = km$ with $k < m$ of *opposite parity*. Since the power of 2 is forced into one side, such factorisations biject with divisors of the odd part of N: the number of representations as a sum of one or more consecutive positive integers is the number of odd divisors, so with two or more it is that number minus one, and N has none iff N is a power of two. Putnam 2017 B2 is that theorem plus $\tau(u) = 2$; Putnam 2015 B6 is the same parametrisation with the run centred rather than started, so the condition becomes $\ell \leq m$.

Bound it, reduce it, descend on it

Bounding: if a variable appears with a higher degree on one side it eventually dominates, so only finitely many small cases survive; and an integer strictly greater than another is at least one more, which is how Putnam 2021 A6 upgrades $Q(2) > Q(1) \geq 1$ to $Q(2) \geq 2$, the inequality that "composite" actually requires. Reducing: the residue tables of Section 4. Descending: replaced, as argued in Section 3, by a single valuation comparison. Putnam 2024 A1 is the model.

Parity deserves its own line because it is nearly free. An integer and its square have the same parity, so comparing parities of $AC^2 - BC^2$, which is computable from coordinates, settles the parity of $AC - BC$ without computing either distance; that is the step that removes the last case in Putnam 2010 B2. And a product of m consecutive integers is $m!\binom{k}{m}$, hence divisible by $m!$ and in particular even for $m \geq 2$, which is why every term with $j \geq 3$ drops out modulo 2 in Putnam 2022 B1.

Pell's equation $x^2 - dy^2 = 1$ for nonsquare $d > 0$ has infinitely many solutions, generated by powers of the fundamental unit of $\mathbb{Z}[\sqrt{d}]$, which is found from the continued fraction of $\sqrt{d}$. In the corpus it appears not as an equation to solve but as a recursion to recognise: the convergent denominators of $\sqrt{2} - 1$ satisfy $q_j = 2q_{j-1} + q_{j-2}$ in Putnam 2020 B6.

- **WHEN** an equation is bilinear, or reduces to $xy + \text{linear} = \text{constant}$ **DO** complete the rectangle into $(\alpha x + \beta)(\gamma y + \delta) = \text{constant}$ and enumerate divisor pairs, signs included **BECAUSE** a bilinear form factors once the constant term is corrected, and a product equalling a fixed integer is a finite search rather than an infinite one.
- **WHEN** a problem mentions sums of consecutive integers **DO** write $2N = km$ with k the number of terms and $m = 2a + k - 1$, note the opposite parity, and count divisors of the odd part **BECAUSE** the arithmetic mean of a run of consecutive integers is either an integer or a half-integer, which is exactly the opposite-parity condition in disguise.
- **WHEN** you must prove no solutions exist **DO** try in order: a valuation comparison, a residue table modulo $3, 4, 8, 9$, then a size bound **BECAUSE** these are the three obstructions that survive at every scale; a search of small cases proves nothing.

4.10 Gaussian integers and other quadratic rings

Definition 4.10.1. $\mathbb{Z}[i] = \{x + iy : x, y \in \mathbb{Z}\}$ with norm $N(x + iy) = x^2 + y^2$. More generally $\mathbb{Z}[\sqrt{d}]$ carries $N(a + b\sqrt{d}) = a^2 - db^2$, the product of an element with its conjugate.

Theorem 4.10.2. *N is multiplicative. $\mathbb{Z}[i]$ is Euclidean with respect to N, hence a unique factorisation domain; its units are exactly the four elements of norm 1, namely $\pm 1, \pm i$; any element whose norm is a rational prime is prime. The rational prime 2 ramifies: $2 = -i(1 + i)^2$ with $1 + i$ prime of norm 2, so $v_{1+i}(m) = 2v_2(m)$ for a rational integer m. A rational prime p splits iff $p \equiv 1 \pmod 4$ and is inert iff $p \equiv 3 \pmod 4$.*

The consequences are exactly the sums-of-two-squares theory. A positive integer m is a sum of two integer squares iff every prime $\equiv 3 \pmod 4$ divides it to an even power, and the number

of ordered representations is $r_2(m) = 4(d_1(m) - d_3(m))$ where d_j counts divisors $\equiv j \pmod 4$. Putnam 2021 A1 needs the twelve lattice vectors of norm 25 and uses r_2 as the completeness check on the list. Putnam 2010 B2 needs to know which distances between lattice points admit only axis-aligned realisations, which is why the argument works for $AB = 1$ or 2 and fails at 5. Putnam 2019 B1 is the cleanest: $N(z) = 2^k$ forces $z = i^a(1+i)^k$, so there are exactly four Gaussian integers of each norm 2^k, and the whole set P_n is described in one line instead of by an induction on coordinates modulo 4. Putnam 2011 B4 turns on the ramification exactly: working with $T + iW$ directly yields only 2^{1005}, while taking a square root first and then norms yields 2^{2010}, and the exponent in the problem is calibrated to detect precisely that factor of two.

The valuation machinery transfers verbatim: for a prime π in a unique factorisation domain, $v_\pi(zw) = v_\pi(z) + v_\pi(w)$ and $v_\pi(z+w) \geq \min$. In $\mathbb{Z}[i]$ the residue field at $1+i$ is $\mathbb{F}_2$, so every unit reduces to 1 and every element of positive valuation to 0; that parity dichotomy is what drives the descent in Putnam 2019 B1.

For real quadratic rings the useful facts are smaller but just as sharp. The conjugate surd $(a+\sqrt{d})(a-\sqrt{d}) = a^2 - d$ is a rational integer, which is how Putnam 2007 B3 clears the irrational coefficient $3+\sqrt{5}$ to the rational integer 4. If d is not a perfect square, an element $p + q\sqrt{d}$ of $\mathbb{Z}[\sqrt{d}]$ that happens to be rational must have $q = 0$; that is what forces the ratio a_{kn}/a_n in Putnam 2015 A2 to be an ordinary integer and delivers the divisibility. The Lucas sequences $V_n = \alpha^n + \beta^n$ for the roots of $x^2 - Px + Q$ live in exactly this ring, and $V_n \mid V_{kn}$ for odd k; the companion sequence U_n satisfies the stronger $\gcd(U_m, U_n) = U_{\gcd(m,n)}$ for all k, of which the Fibonacci law $\gcd(F_a, F_b) = F_{\gcd(a,b)}$ is the case $P = 1, Q = -1$. Know which of the two families you are in: mixing them up is the standard error, and Putnam 2015 A2 exists to punish it.

Finally, value sets of norm forms are closed under multiplication, since the norm is. The examples worth carrying are $a^2 + b^2$ (Gaussian), $a^2 + ab + b^2$ (Eisenstein, after $b \mapsto -b$), and $a^3 + b^3 + c^3 - 3abc$, which is a 3×3 circulant determinant. In Putnam 2019 A1 the answer is the set of nonnegative integers not congruent to 3 or 6 modulo 9, and multiplicative closure of that set is the consistency check that tells you the answer is right before you write the proof.

- **WHEN** a sum of two squares, a norm $x^2 + y^2$, or a lattice point at a prescribed distance appears **DO** move into $\mathbb{Z}[i]$, factor the norm there, and use unique factorisation and the four units **BECAUSE** $x^2 + y^2 = N(x + iy)$ is multiplicative, so a multiplicative question about the norm is a factorisation question in a ring where factorisation is unique.
- **WHEN** an integer sequence is $\alpha^n + \beta^n$ or $(\alpha^n - \beta^n)/(\alpha - \beta)$ for conjugate quadratic surds **DO** prove divisibility statements inside $\mathbb{Z}[\sqrt{d}]$ and then use "rational plus in the ring implies integer" to descend **BECAUSE** the ring makes the algebra linear; the descent step is the only place the irrationality of $\sqrt{d}$ is spent.

4.11 Polynomial congruences, Hensel lifting and finite fields

Roots modulo p and the ideal $(x^p - x)$

$\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ is a field, so a nonzero polynomial of degree d over it has at most d roots; that single fact proved cyclicity in Section 6 and bounds root counts everywhere. By Fermat, $x^p - x = \prod_{a \in \mathbb{F}_p}(x - a)$, so it vanishes as a function while being a nonzero polynomial of degree p: the evaluation map $\mathbb{F}_p[x] \to \{\text{functions } \mathbb{F}_p \to \mathbb{F}_p\}$ is surjective with kernel exactly $(x^p - x)$. Its formal derivative is -1. Putnam 2011 B6 exploits precisely that asymmetry: adding $x^p - x$ changes no value but shifts every derivative by -1, converting simple roots into double roots at a cost of one unit of degree.

The other polynomial identity you need is the first order Taylor congruence.

Lemma 4.11.1. *For $F \in \mathbb{Z}[x]$ and $a \in \mathbb{Z}$, $F(x) \equiv F(a) + F'(a)(x-a) \pmod{(x-a)^2}$. In particular $x^n - 1 \equiv n(x-1) \pmod{(x-1)^2}$, and $x^n + 1 \equiv n(x+1) \pmod{(x+1)^2}$ for odd n.*

Proof. Expand F in powers of $(x-a)$ by repeated synthetic division; the terms of degree at least two are divisible by $(x-a)^2$, and the first two coefficients are $F(a)$ and $F'(a)$. □

This is the identity that converts an exponent into a coefficient, which is the whole construction in Putnam 2009 B6. That problem also needs the bridge from polynomials to integers: if $F = \sum d_j x^j$ has integer coefficients with $0 \le d_j < b$, then the base b digits of $F(b)$ are exactly the d_j and no carrying occurs, so remainder identities in $\mathbb{Z}[x]$ specialise to remainder identities for integers. That faithfulness is exactly where a hypothesis like $x > n$ is spent. The same specialisation principle names the family Putnam 2021 A6 lives in: Cohn's criterion, extended by Brillhart, Filaseta and Odlyzko, says that if $P(b)$ is prime for some base $b \ge 2$ with digits as coefficients then P is irreducible over $\mathbb{Z}$; recognising the family tells you the proof must be root location plus a shift, not a congruence.

Hensel lifting

Theorem 4.11.2 (Hensel, simple root)**.** *Let $P \in \mathbb{Z}[x]$, $P(a) \equiv 0 \pmod{p^k}$ and $p \nmid P'(a)$. Then there is a unique b modulo p^{k+1} with $b \equiv a \pmod{p^k}$ and $P(b) \equiv 0 \pmod{p^{k+1}}$. Iterating gives a unique root in the p-adic integers congruent to a modulo p^k.*

Proof. Write $b = a + tp^k$. By the Taylor congruence, $P(b) \equiv P(a) + tp^k P'(a) \pmod{p^{2k}}$, and $2k \ge k+1$. So $P(b) \equiv 0 \pmod{p^{k+1}}$ says $P(a)/p^k + tP'(a) \equiv 0 \pmod{p}$, a linear congruence in t whose coefficient $P'(a)$ is a unit, hence uniquely solvable. □

Read the proof again and notice its shape: linearise on the kernel of $\mathbb{Z}/p^{k+1} \to \mathbb{Z}/p^k$, which has order p, and demand the slope be a unit. Because the slope condition does not mention k, one induction closes all levels at once. Putnam 2008 B4 is this argument with the root replaced by an arbitrary value: the hypothesis $p \nmid h'$ is Hensel's simple root condition, and the general theorem the problem instantiates is Kempner's criterion, that for $k \ge 2$ a polynomial h permutes $\mathbb{Z}/p^k\mathbb{Z}$ iff it permutes $\mathbb{Z}/p\mathbb{Z}$ and $h'(a) \not\equiv 0 \pmod{p}$ for every integer a. Cite the criterion by name in a remark, then prove the special case; pretending the citation is the proof loses marks.

Finite fields

For every prime power $q = p^n$ there is a field $\mathbb{F}_q$ with q elements, unique up to isomorphism, constructed as $\mathbb{F}_p[x]/(f)$ for any irreducible f of degree n; it has characteristic p, is an n-dimensional $\mathbb{F}_p$ vector space, and $\mathbb{F}_q^\times$ is cyclic of order $q-1$ by the same at-most-d-roots argument as before. Every element satisfies $z^q = z$, and the Frobenius $z \mapsto z^p$ is a field automorphism.

The Putnam use is almost always counting or a dichotomy at $q = p$. Putnam 2022 A3 needs the root count of a quadratic over $\mathbb{F}_p$ and thus the 0 or 2 dichotomy. Putnam 2024 A4 needs $\mathbb{F}_p^\times$ cyclic, that the inverse of a primitive root is a primitive root, and that $a^{p-1-k} = a^{-k}$, so the complement of a long geometric progression is $\{0, c, c^2, c^3\}$ with $c = a^{-1}$. Putnam 2011 B6 needs $z^p = z$ for all z and $z^{p-1} = 1$ for $z \ne 0$ simultaneously, one to kill a prefactor and one to evaluate a derivative. Putnam 2017 B6 needs the translation action of $\mathbb{F}_p$ on itself: a linear form $L(x) = \sum c_i x_i$ satisfies $L(x + t\mathbf{1}) = L(x) + t\sum c_i$, so L is translation invariant exactly when $\sum c_i \equiv 0$, and if $\sum c_i$ is a unit then L is a bijection along every translation orbit. When the coefficients are all positive integers summing to exactly p, every proper nonempty block sum lies strictly between 0 and p, so no proper degeneracy can occur; that single design condition does two jobs at once.

- **WHEN** a congruence must be solved to a higher power of p and you already have a solution modulo p **DO** check the derivative is a unit modulo p and lift by Hensel, uniquely, one level at a time **BECAUSE** the kernel of $\mathbb{Z}/p^{k+1} \to \mathbb{Z}/p^k$ has order p, on which the polynomial is affine with slope $P'(a)$, so a unit slope makes the lift a bijection.

- **WHEN** a problem counts solutions of a polynomial equation modulo a prime **DO** work in $\mathbb{F}_p$, use the degree bound on root counts and the cyclicity of $\mathbb{F}_p^\times$, and only then translate back **BECAUSE** being a field is what licenses both facts, and neither survives at a composite modulus.

4.12 Distribution of primes: what you may quote

You will never be asked to prove these on the Putnam, and you may cite all of them by name. Euclid: there are infinitely many primes. Bertrand's postulate: for every $n \geq 1$ there is a prime p with $n < p \leq 2n$. Dirichlet: if $\gcd(a, d) = 1$ there are infinitely many primes $\equiv a \pmod{d}$. The prime number theorem: $\pi(x) \sim x/\log x$. Treat the last as a citable fact and not a tool; a Putnam solution that needs the prime number theorem quantitatively is almost certainly the wrong solution.

What you do construct by hand is gaps. For any $t \geq 1$ the t integers $(t+1)!+2, \ldots, (t+1)!+(t+1)$ are all composite, since k divides both $(t+1)!$ and k for $2 \leq k \leq t+1$, and the sum exceeds k. Putnam 2006 A2 needs exactly that: one construction, parametrised only by a bound, witnessing compositeness simultaneously for a whole finite set of shifts. The same problem contains a negative lesson worth carrying: the set $\{p-1 : p \text{ prime}\}$ is infinite but has no periodic or arithmetic structure, so a direct search for periodicity in a set built from it is not viable, and the indirect argument is necessary rather than merely convenient.

Two practical notes. To prove a specific number is prime, trial divide by primes up to its square root and say so. To prove a number is composite, exhibit a divisor strictly between 1 and it; the divisor need not be prime, and in Putnam 2010 A4 the exhibited divisor $10001 = 73 \cdot 137$ is itself composite. Also useful: $n \mid 2^n$ iff n is a power of two, since an odd prime factor of n would have to divide 2^n.

4.13 Continued fractions, approximation, Farey and Stern to Brocot

Theorem 4.13.1 (Dirichlet approximation)**.** *For real α and integer $N \geq 1$ there are integers p, q with $1 \leq q \leq N$ and $|q\alpha - p| < 1/N$, hence $|\alpha - p/q| < 1/(qN)$.*

Proof. Place the $N+1$ fractional parts $\{0\}, \{\alpha\}, \ldots, \{N\alpha\}$ on the circle $[0,1)$ of circumference 1; two of them lie within $1/(N+1)$ of each other, and their difference is $q\alpha - p$ for some $1 \leq q \leq N$. □

Pigeonhole on the circle is the single most reusable mechanism in this section, and it generalises the moment you see it: for $N+1$ points on a circle of circumference L, two are within $L/(N+1)$, and when the points are generated by a recursion $s_i = s_{i-1} + x_i$ the same bound applies to the gaps between the sorted fractional parts. Putnam 2006 B2 is exactly that, and the reason partial sums rather than arbitrary subsets are the right auxiliary family is that only partial sums differ by a contiguous block. Putnam 2011 B1 is a miniature of the same theme, trapping a constructed real inside a window by integer choices.

Theorem 4.13.2 (Subgroups of $\mathbb{R}$, and Kronecker)**.** *A nontrivial subgroup of $(\mathbb{R}, +)$ is either $c\mathbb{Z}$ for some $c > 0$ or dense. Hence $\mathbb{Z} + \alpha\mathbb{Z}$ is dense iff α is irrational, and the fractional parts of $n\alpha$ are dense in $[0,1)$.*

Proof. Let c be the infimum of the positive elements. If $c > 0$ then c is attained (two elements within c of each other differ by a smaller positive element) and the group is $c\mathbb{Z}$ by the division algorithm. If $c = 0$ the group has arbitrarily small positive elements and their multiples fill $\mathbb{R}$. □

Putnam 2023 A4 is that dichotomy applied to a group $2\mathbb{Z} + 2\varphi\mathbb{Z}$ coming from icosahedral coordinates: not cyclic, therefore dense, and that is the only analytic step in the solution. Putnam 2016 B5 needs the sharper form that $\{a\theta - b : a, b > N\}$ is dense for irrational θ, plus the irrationality

of $\log 3/\log 2$, which is unique factorisation in one line: $\log 3/\log 2 = p/q$ would give $3^q = 2^p$. Weyl upgrades density to equidistribution, but be warned, as Putnam 2020 B6 warns: equidistribution gives $\sum_{k\le n}(-1)^{\lfloor k\theta\rfloor} = o(n)$ and nothing about signs, so no averaging or discrepancy bound can prove a nonnegativity statement. Knowing that in advance saves the session.

Refinement of the pigeonhole bound: the three distance theorem says the points $\{\theta\},\dots,\{N\theta\}$ cut the circle into arcs of at most three lengths, the largest being the sum of the other two. Its proof is the continued fraction recursion, the same descent as the Euclidean algorithm. It is the general shape of the sets appearing in Putnam 2012 A4 and Putnam 2006 B2.

Continued fractions themselves: $\alpha = [a_0; a_1, a_2, \dots]$ with the Gauss map $T(x) = \{1/x\}$ stepping the expansion. A purely periodic expansion means a fixed point structure, and $\sqrt{2}-1 = 1/(2+(\sqrt{2}-1)) = [0;2,2,2,\dots]$ is a genuine fixed point of T. That self-reference is why the run length coding in Putnam 2020 B6 reproduces the original sequence rather than an infinite family of relatives, and the convergent denominators obey the Pell recursion $q_j = 2q_{j-1} + q_{j-2}$, so $q_j \equiv j \pmod 2$.

Farey and mediants are the combinatorial face of the same theory.

Theorem 4.13.3 (Unimodular pairs, mediants, Farey)**.** *If $m/q < m'/q'$ in lowest terms satisfy $m'q - mq' = 1$ then $\gcd(q,q') = 1$, the gap is $1/(qq')$, and the mediant $(m+m')/(q+q')$ is automatically in lowest terms, lies strictly between them, and forms a unimodular pair with each endpoint. Any u/v in lowest terms strictly between them has $v \ge q+q'$, with equality only for the mediant. The Farey sequence F_N consists of the rationals in $[0,1]$ of denominator at most N; consecutive elements are exactly the unimodular pairs with $q, q' \le N < q+q'$, and $|F_N| = 1 + \sum_{k\le N}\varphi(k)$, of order $3N^2/\pi^2$.*

Iterating the mediant from the root pair $(0/1, 1/1)$ generates the tree of Stern and Brocot, which produces every rational in $(0,1)$ exactly once, and the path to a given rational is its continued fraction in run length form. Putnam 2014 B6 is built entirely from this: unimodularity is exactly the condition making a chord slope an integer, and the statistic $v-(q+q')$ is a nonnegative integer that strictly decreases on descent, which turns an informal repetition into a finite induction.

Floor sequences are the last member of the family. For irrational $0<\theta<1$ the level sets of $\lfloor k\theta\rfloor$ have run lengths $c_i = \lfloor (i+1)/\theta\rfloor - \lfloor i/\theta\rfloor \in \{\lfloor 1/\theta\rfloor, \lfloor 1/\theta\rfloor + 1\}$, and the long runs are indexed by the Beatty sequence $\lfloor j\beta\rfloor$ with $\beta = 1/\theta$. Rayleigh's theorem completes the picture: for irrational $\beta, \beta' > 1$ with $1/\beta + 1/\beta' = 1$, the sequences $\lfloor j\beta\rfloor$ and $\lfloor j\beta'\rfloor$ partition the positive integers. For rational $x = p/q$ in lowest terms, $\lfloor (n+q)x\rfloor = \lfloor nx\rfloor + p$, so the parity word of $\lfloor nx\rfloor$ has period q if p is even and $2q$ if p is odd; Putnam 2015 A4 is the rigidity of increments, $\lfloor (n+1)x\rfloor - \lfloor nx\rfloor \in \{0,1\}$, pushed to its limit. Keep Hermite's identity $\sum_{j=0}^{d-1}\lfloor y + j/d\rfloor = \lfloor dy\rfloor$ and the pair $\lfloor \lfloor y\rfloor/d\rfloor = \lfloor y/d\rfloor$ in the same drawer.

- **WHEN** a real irrational multiplier generates a sequence and you need two of its values close together, or a value inside a small window **DO** pigeonhole $N+1$ fractional parts on the circle of circumference 1 **BECAUSE** $N+1$ points in N arcs forces a collision, and the difference of two fractional parts of multiples is again of that form.

- **WHEN** a set of reals is closed under addition and subtraction **DO** decide cyclic against dense by looking at the infimum of its positive elements **BECAUSE** those are the only two possibilities, and the dichotomy converts an algebraic hypothesis directly into a topological conclusion.

- **WHEN** rationals appear ordered by size with denominators controlled **DO** think Farey and mediants: the first new rational between two neighbours is the mediant, and its denominator is the sum **BECAUSE** unimodularity is preserved by the mediant, so the whole interval structure is generated by one operation with an integer monovariant.

4.14 The recursion section: arithmetic recursions in general form

Every recursion the corpus threw up in this field is collected here, stated in general form with its solution method and its recognition cue. This is the section to reread before a paper.

Descent and well founded induction

Proposition 4.14.1 (Induction on a non-standard measure)**.** *Strong induction does not require the measure to be the size of the objects. Any nonnegative-integer-valued statistic that strictly decreases at each step licenses a valid induction, even while the objects themselves grow without bound.*

Putnam 2009 B1 is the canonical case: to write every positive rational as a quotient of products of prime factorials you induct on the *largest prime factor* of numerator and denominator, using $p = p!/(p-1)!$ with $(p-1)!$ having all prime factors below p. Induction on $\max(a, b)$ fails outright because $(p-1)!$ dwarfs the original numbers. The protocol is: (1) name the statistic, (2) prove strict decrease, (3) check the floor case, (4) invoke well ordering to guarantee termination.

Related descents: Putnam 2013 A2 assumes a property fails at the least member of a family and constructs a strictly smaller member; Putnam 2014 B6 uses $v - (q + q')$ as the measure on a mediant subdivision; Putnam 2009 A4 uses the continued fraction descent "translate into $0 < r < q$, then invert to get denominator $r < q$" as the general template for reachability under a family of affine and inversion maps. And, as argued twice already, the descent $(a, b, c) \mapsto (a/p, b/p, c/p)$ on primitive solutions has closed form "compare valuations once" (Putnam 2024 A1).

Recursions in the exponent

Proposition 4.14.2. *For* $\gcd(g, m) = 1$ *the map* $r \mapsto gr$ *on* $\mathbb{Z}/m$ *is a bijection of a finite set, so the sequence* $r_c = g^c r_0$ *is* purely *periodic with period* $\mathrm{ord}_m(g)$*. If* g *shares a factor with* m*, strip it: the sequence is eventually periodic with pre-period given by that common factor, and the tail lives in the unit group.*

Iterating a power map is a recursion in the exponent: $x_{n+1} = x_n^k$ becomes $i_{n+1} = ki_n \pmod{p-1}$ under the discrete logarithm, so $i_n = k^n i_0$ and the orbit of an element of order d closes after $\mathrm{ord}_d(k)$ steps (Putnam 2012 B6). Every question about iterated power maps modulo p reduces to multiplicative orders in the exponent group. The Euclidean recursion in the exponent gives $\gcd(a^m - 1, a^n - 1) = a^{\gcd(m,n)} - 1$, which mirrors the recursion $\gcd(m, n)$ itself (Putnam 2018 B3). Lifting the exponent is the recursion that lifts orders through powers of p: $v_p(g^{dn} - 1) = v_p(g^d - 1) + v_p(n)$ for odd p (Putnam 2023 B2). And stacking k shifted Mersenne-type divisibility conditions produces a k-deep power tower, because each condition is the identical "is this a power of two" test applied to the exponent produced by the previous one (Putnam 2018 B3).

Linear recursions modulo m

Proposition 4.14.3. *For* $u_{n+1} = c_1 u_n + \cdots + c_k u_{n-k+1}$ *with integer coefficients, the state vector evolves by a fixed matrix* M *over* $\mathbb{Z}/m$*. The sequence is always eventually periodic, and purely periodic whenever* $\det M$ *is a unit modulo* m*.*

For Fibonacci, $T(u, v) = (v, u + v)$ has $\det = -1$, a unit modulo every m, so the sequence is purely periodic modulo every m: the Pisano period $\pi(m)$, with $\pi(2) = 3$, $\pi(4) = 6$, $\pi(8) = 12$, $\pi(10) = 60$ and $\pi(m) \leq 6m$. Putnam 2018 B4 is a real-analysis-looking problem whose whole content is that x_n depends on F_n modulo $d = 4F_N$, hence is periodic. Putnam 2021 B4 needs only $\pi(4) = 6$ with cycle $0, 1, 1, 2, 3, 1$, from which $F_m \equiv 1$ or $2 \pmod 4$ for odd m.

The first order inhomogeneous case is worth a line of its own because it appears in disguise constantly. $u_{m+1} = bu_m + c$ with $b \neq 1$ has fixed point $p^* = c/(1-b)$, and translating by p^* turns the recursion into pure multiplication: $u_m = b^m(u_0 - p^*) + p^*$. Repunits are the case $b = 10$, $c = 1$,

$u_0 = 0$, giving $R_m = (10^m - 1)/9$ (Putnam 2020 A1, Putnam 2007 A4), and the same template covers any base.

Digit and floor recursions

Digit characters: $u_{bk+r} = \chi(r)u_k$ defines a sequence whose value at n is a product over the base b digits of n. Thue and Morse is $b = 2$, $\chi = (1, -1)$, giving $(-1)^{d(k)}$ with d the binary digit sum (Putnam 2020 B1); Rudin and Shapiro is $c(2n) = c(n)$, $c(2n+1) = (-1)^n c(n)$, giving (-1) to the number of adjacent pairs of ones (Putnam 2013 B1). The structural consequence: on an aligned block of length b^m the sum of u against any function becomes an m-fold operator product, which is what makes finite difference identities available.

Floor recursions: $\pi(n) = \pi(n-m) + 1$ has general solution $\lfloor n/m \rfloor$ plus an arbitrary function of the residue, so any statistic obeying such a recursion should be linearised as $\pi(n) = (n + c_{n \bmod m})/m$ before being compared at several arguments; that reduces Putnam 2022 A5 from forty nine residue cases to one interval bound. The telescoping identity $a_1 + \cdots + a_k = \lfloor mk/n \rfloor$ for $a_k = \lfloor mk/n \rfloor - \lfloor m(k-1)/n \rfloor$ is the closed form of $s_k = s_{k-1} + a_k$ and gives periodicity, letter counts and shift identities all at once (Putnam 2018 A4). The Gauss map $\theta \mapsto \{1/\theta\}$ is the recursion on the parameter behind run length codings, so the rth coding has parameter $T^r(\theta)$ and a fixed point of T gives a self-referential induction (Putnam 2020 B6).

Multiplicativity, convolution and fibres as recursions

Multiplicativity is a recursion on the number of distinct prime factors: peel off one prime power, multiply, base case $f(1) = 1$ (Putnam 2015 A3). Mobius inversion is the general poset inversion: a family $v_k = \sum_{j \le k} e_j$ is a basis with transition determinant 1, and the inverse transition is the poset Mobius function (Putnam 2023 B6). Fibre recursions: the fibre over $2N$ of the map $(k, n) \mapsto k2^n$ is the fibre over N plus one element at the doubled end, so fibre sums satisfy a one-step recursion whose solution is the geometric collapse $1 - \sum_{i=1}^{e} 2^{-i} = 2^{-e}$ (Putnam 2016 B6). Divisibility chains: if F satisfies $F(u) \mid F(v)$ whenever $u \mid v$, iterating F from any seed gives a nested chain of divisor sets whose closure is their union (Putnam 2025 B3). Hensel is the level to level recursion $B_k \Rightarrow B_{k+1}$ whose invariant carries no k, so one induction closes all levels (Putnam 2008 B4). And composition raises congruence level: if $g \equiv x + p^e h$ coefficientwise then $g^{\circ j} \equiv x + jp^e h \pmod{p^{2e}}$, so $g^{\circ p} \equiv x \pmod{p^{e+1}}$, which is why the orbit of 0 under a quadratic map has period exactly 2^{j-1} modulo 2^j in Putnam 2025 A6.

- **WHEN** a recursion involves a modulus and an invertible step map **DO** conclude pure periodicity at once and compute the period as an order or a Pisano period **BECAUSE** an injective self-map of a finite set is a bijection, and a bijection has no pre-period.
- **WHEN** a substitution makes the objects bigger but is supposed to terminate **DO** find the auxiliary nonnegative integer statistic it decreases and induct on that instead of on size **BECAUSE** well ordering only needs some strictly decreasing nonnegative measure, and size is rarely the right one.
- **WHEN** an iteration is claimed to gain exactly one power of p per step **DO** prove the one-step statement with LTE or the coefficientwise composition congruence, then iterate **BECAUSE** both give exact equalities, so the induction closes without an accumulating error term.

4.15 The exotica: learn these last

Everything below appeared once or twice in twenty years. Recognise the name, know where to look, and spend no drill time here until Sections 2 to 8 are automatic.

Ostrowski numeration, the representation $n = \sum d_j q_j$ in convergent denominators of an irrational, is the machinery of the published continued fraction solution to Putnam 2020 B6, and must be installed before that solution can even be read. Zeckendorf representation, the unique Fibonacci sum with no two consecutive indices, appears in Putnam 2020 A5; the general lesson, flagged in Putnam 2014 B1, is that uniqueness of representation can be governed by an adjacency constraint rather than by a digit bound, and you should know both mechanisms exist. The factorial number system $N = \sum c_i\, i!$ with $0 \le c_i \le i$ appears in Putnam 2007 B6. Kurepa's conjecture, that $0! + 1! + \cdots + (p-1)!$ is never divisible by an odd prime p, is the reason Putnam 2011 B6 must be solved by a root count that never inspects a value: no solution can name a good residue, because naming one would settle an open problem. Zolotarev's lemma (Putnam 2012 B6) and Kempner's criterion (Putnam 2008 B4) are single-appearance named theorems worth citing rather than reproving. Polya's basis theorem for integer-valued polynomials, and Bhargava's P-orderings which generalise $n!$ to arbitrary subsets of a Dedekind domain, are named in the remark to Putnam 2016 A1 and nowhere else. And p-adic interpolation of iterates, the Bell and Poonen circle of ideas, tells you where Putnam 2025 A6 lives and why its answer has the shape it does; the exam proof is an elementary induction and you should write that one.

Chapter 5

Sequences, recurrences and generating functions

5.1 What this chapter is for

A-level gave you four things about sequences and let you believe they were the subject. You can sum an arithmetic series, sum a geometric series, run the method of differences on a fraction that has been pre-split for you, and prove a divisibility statement by ordinary induction where the inductive step is one line of algebra. Further Maths adds Maclaurin series, treated as an analytic object that you differentiate and evaluate, and first order recurrence relations, treated as a machine for producing the first four terms.

None of that is the subject. The subject is this: a recursion is a linear or nonlinear operator acting on the space of sequences, and solving a recursion means diagonalising that operator or conjugating it to something you already know. Once you see it that way, an enormous fraction of the Putnam collapses. Across the twenty papers analysed for this archive, recursion is the single most frequent structural feature: over three hundred separate pieces of required knowledge carry an explicit recursive generalisation, spread across every field on the paper, from number theory to real analysis to games. That is why this is the longest chapter in the book, and why it is the one that repays being read twice.

At the end of this chapter you will be able to do the following without hesitating. Solve any constant coefficient linear recurrence of any order, including repeated and complex roots, and including inhomogeneous forcing with resonance. Solve any first order recurrence, constant or variable coefficient, by the summation factor. Recognise the six substitutions that turn a nonlinear recursion into a linear one, and know which cue triggers which. Read a recursion as an iterated map, locate its fixed points, classify them by the multiplier, and predict the orbit. Prove that a recursion modulo m is eventually periodic, and say exactly when it is purely periodic. Build a transfer matrix from a deletion map. Manipulate generating functions as algebra rather than analysis, including the exponential and probability variants, partial fractions, roots of unity filters, and the snake oil method. Extract the asymptotics of a recursively defined sequence with no closed form. And run induction in all six of its shapes, above all the shape nobody teaches you at school: strengthening the statement so that the induction goes through.

One standing instruction governs the whole chapter. Every time you meet a recursion, in any field, on any paper, you do not solve it and move on. You name its general form, you name the solution method for that form, you name the invariant it preserves, and you file it. That is the difference between a solved problem and an installed circuit.

5.2 The shift operator and the complete theory of constant coefficient linear recurrences

The definition and why it is the right one

Definition 5.2.1 (The sequence space and the shift)**.** Let $\mathcal{S}$ be the set of all sequences $x = (x_0, x_1, x_2, \dots)$ with entries in a field K; unless stated otherwise take $K = \mathbb{C}$. Under termwise addition and scalar multiplication $\mathcal{S}$ is a vector space. Define the *shift* $E : \mathcal{S} \to \mathcal{S}$ by $(Ex)_n = x_{n+1}$, and the *forward difference* $\Delta = E - I$, so $(\Delta x)_n = x_{n+1} - x_n$.

E is linear, and any two polynomials in E commute, because polynomials in a single operator always commute. That one sentence is the whole reason this formalism pays: you may factorise, reorder and compose annihilators at will.

Definition 5.2.2 (Linear recurrence, characteristic polynomial)**.** A sequence x satisfies a *constant coefficient linear recurrence of order k* if

$$x_{n+k} = c_1 x_{n+k-1} + c_2 x_{n+k-2} + \cdots + c_k x_n \qquad (n \geq 0), \quad c_k \neq 0.$$

Equivalently $q(E)x = 0$ where $q(t) = t^k - c_1 t^{k-1} - \cdots - c_k$. The polynomial q is the *characteristic polynomial* and $q(t) = 0$ is the *characteristic equation*.

Go back to the definition and the point of interest falls out. A recurrence is not a rule for generating numbers; it is the statement that a fixed polynomial in E kills your sequence. The set of sequences killed by $q(E)$ is a kernel, hence a vector space, hence has a basis, hence has a closed form. Everything below is the computation of that basis.

The structure theorem

Theorem 5.2.3 (Structure of solutions)**.** *Let $q(t) = \prod_{i=1}^{r} (t - \lambda_i)^{m_i}$ with the λ_i distinct and nonzero, and $\sum_i m_i = k$. Then $\ker q(E)$ has dimension k, with basis the k sequences*

$$\left(n^j \lambda_i^n\right)_{n \geq 0}, \qquad 1 \leq i \leq r, \quad 0 \leq j < m_i.$$

Proof. First take $q(t) = (t - \lambda)^m$. Write $x_n = \lambda^n y_n$, legitimate since $\lambda \neq 0$. Then

$$\big((E - \lambda I)x\big)_n = x_{n+1} - \lambda x_n = \lambda^{n+1}(y_{n+1} - y_n) = \lambda^{n+1}(\Delta y)_n,$$

so by induction $\big((E - \lambda I)^m x\big)_n = \lambda^{n+m}(\Delta^m y)_n$. Hence $(E - \lambda I)^m x = 0$ if and only if $\Delta^m y = 0$. Now Δ lowers the degree of a polynomial sequence by exactly one and annihilates exactly the constants, so $\Delta^m y = 0$ if and only if y_n is a polynomial in n of degree less than m. That is the claim, and the dimension is m.

For the general case put $q_i = q/(t - \lambda_i)^{m_i}$. The q_i have no common root, so there are polynomials a_i with $\sum_i a_i q_i = 1$. Given $x \in \ker q(E)$ set $x^{(i)} = a_i(E) q_i(E) x$. Then $\sum_i x^{(i)} = x$, and $(E - \lambda_i I)^{m_i} x^{(i)} = a_i(E) q(E) x = 0$, so each $x^{(i)}$ lies in the corresponding kernel. The sum is direct: if $\sum_i z^{(i)} = 0$ with $z^{(i)} \in \ker(E - \lambda_i I)^{m_i}$, apply $q_j(E)$, which annihilates every term but the jth, and which is injective on $\ker(E - \lambda_j I)^{m_j}$ because $q_j(\lambda_j) \neq 0$. Dimensions therefore add to k. □

Corollary 5.2.4 (The three cases you will actually meet)**.** *For $k = 2$, with $x_{n+2} = p x_{n+1} + q x_n$ and characteristic equation $t^2 - pt - q = 0$: distinct roots $\lambda_1 \neq \lambda_2$ give $x_n = A\lambda_1^n + B\lambda_2^n$; a repeated root λ gives $x_n = (A + Bn)\lambda^n$; a conjugate pair $re^{\pm i\theta}$ gives the real form $x_n = r^n\big(A \cos n\theta + B \sin n\theta\big)$. In every case A and B are fixed by any two consecutive terms, through a 2×2 linear system whose matrix is invertible precisely because the basis sequences are independent.*

This is the machinery behind Putnam 2007 B3, where the derived recurrence $x_{n+2} = 6x_{n+1} - 4x_n$ has roots $3 \pm \sqrt{5}$; behind Putnam 2015 A2 and Putnam 2015 B3, where the closed form of an order two recurrence is what lets a divisibility or an eigenvalue statement be read off; and behind Putnam 2012 A1, where the growth rate of a Fibonacci-dominated sequence is the dominant root.

Proposition 5.2.5 (Window principle, Putnam 2010 B6)**.** *A solution of an order k constant coefficient recurrence valid for all $n \geq N$ is determined by any k consecutive terms with indices $\geq N$. Consequently, if two sequences taking values in the same vector space satisfy the same recurrence from index N on and agree at k consecutive indices, they agree at every index from that window onward.*

Proof. The recurrence expresses each new term as a fixed linear combination of the previous k, so the window $(x_n, \ldots, x_{n-k+1})$ determines the window one step later. Induct. □

This is the discrete twin of uniqueness for an initial value problem, and it is why Putnam 2010 B6 can convert finitely many verified instances into all instances. It is also the licence for the most economical proof style in the whole subject: if you want to prove two explicitly defined sequences are equal, do not compare closed forms. Show they satisfy the same recurrence, check k initial terms, stop. That is exactly the move in Putnam 2017 A2, where an auxiliary sequence P is shown to satisfy the defining relation of Q and then matched on two terms.

The companion matrix and everything it hands you

Definition 5.2.6 (Companion form)**.** For $x_{n+1} = c_1 x_n + \cdots + c_k x_{n-k+1}$ put $v_n = (x_n, x_{n-1}, \ldots, x_{n-k+1})^T$. Then $v_{n+1} = C v_n$ where C has first row $(c_1, \ldots, c_k)$ and the identity in rows 2 to k shifted left. The characteristic polynomial of C is exactly q.

Three consequences, all of them Putnam currency.

Proposition 5.2.7 (Cassini type invariants, Putnam 2017 A2 and 2021 B4)**.** *Let $u_n = a u_{n-1} + b u_{n-2}$ and let M_n be the matrix with rows (u_{n-1}, u_n) and (u_{n-2}, u_{n-1}). Then $M_{n+1} = T M_n$ with T the companion matrix having rows (a, b) and $(1, 0)$, so*

$$u_{n-1}^2 - u_n u_{n-2} = \det M_n = (\det T)^{n-1} \det M_1 = (-b)^{n-1}\big(u_0^2 - u_1 u_{-1}\big).$$

When $\det T = 1$ the quantity $u_{n-1}^2 - u_n u_{n-2}$ is the same constant for every n.

Proof. Multiply out: the top row of TM_n is $(a u_{n-1} + b u_{n-2},\ a u_n + b u_{n-1}) = (u_n, u_{n+1})$ and the bottom row is (u_{n-1}, u_n), which is M_{n+1}. Now take determinants and use multiplicativity. □

For Fibonacci, $a = b = 1$ and $\det T = -1$, and the identity is Cassini's, $F_{n+1}F_{n-1} - F_n^2 = (-1)^n$. In Putnam 2017 A2 the same one line determinant computation replaces an algebraic identity that otherwise looks unmotivated. That is the point: an identity produced by a determinant is not an identity you had to guess.

Proposition 5.2.8 (Powers of a matrix satisfy a linear recurrence)**.** *By Cayley and Hamilton, if $P(t) = \det(tI - A)$ has degree n then $P(A) = 0$, so the matrix powers satisfy $A^{m+1} = c_1 A^m + \cdots + c_n A^{m+1-n}$ for every $m \geq n - 1$. Any monic annihilating polynomial of degree d gives such a recurrence of order d.*

Putnam 2010 B6 and Putnam 2015 B3 both use this; for a 2×2 matrix it reads $M^2 = (\operatorname{tr} M)M - (\det M)I$, which you should be able to write without thinking.

Proposition 5.2.9 (Power sums of roots, Putnam 2022 A6 and 2019 A1)**.** *If $f(t) = t^N - e_1 t^{N-1} + e_2 t^{N-2} - \cdots + (-1)^N e_N$ has roots $r_1, \ldots, r_N$ with multiplicity, then $s_j = \sum_i r_i^j$ satisfies, for $j > N$,*

$$s_j = e_1 s_{j-1} - e_2 s_{j-2} + \cdots + (-1)^{N-1} e_N s_{j-N},$$

a linear recurrence whose characteristic polynomial is f itself. For $j \leq N$ the Newton identities $j e_j = \sum_{i=1}^{j} (-1)^{i-1} e_{j-i} s_i$ start the sequence off.

Proof. Each root satisfies $r_i^j = e_1 r_i^{j-1} - \cdots$; sum over i. □

Read that in reverse and it is a decision rule: whenever a problem hands you all the power sums of an unknown multiset, pass to the elementary symmetric functions, because those are the coefficients of a polynomial and polynomials are objects you can factor. Putnam 2022 A6 and Putnam 2015 A6 both turn on this.

Proposition 5.2.10 (Annihilating polynomial from a finite list of values, Putnam 2007 B5)**.** *If a quantity u is known to equal one of $v_0, \dots, v_{k-1}$, then $\prod_j (u - v_j) = 0$ identically, so*

$$u^k = \sum_{i=0}^{k-1} (-1)^{k-i-1} e_{k-i}(v_0, \dots, v_{k-1})\, u^i.$$

This is a linear recurrence of order k on the powers $u^0, u^1, u^2, \dots$, and it never requires knowing which of the k cases actually occurs. That is exactly why it works on Putnam 2007 B5, where the case is unknowable but irrelevant.

Inhomogeneous recurrences and resonance

Theorem 5.2.11 (Particular plus homogeneous)**.** *The general solution of $q(E)x = f$ is one particular solution plus the general solution of $q(E)x = 0$. If $f_n = \mu^n p(n)$ with $\deg p = e$, there is a particular solution of the form $\mu^n n^s \tilde{p}(n)$ with $\deg \tilde{p} = e$, where s is the multiplicity of μ as a root of q (with $s = 0$ if $q(\mu) \neq 0$).*

Proof. Apply Theorem 5.2.3 to the enlarged annihilator $q(t)(t - \mu)^{e+1}$, which kills both sides. □

Two special cases are worth banking as reflexes.

Constant forcing: $q(E)x = c \cdot \mathbf{1}$. Since $q(E)$ acts on the constant sequence by multiplication by $q(1)$, the particular solution is $c/q(1)$ whenever $q(1) \neq 0$, and $x - c/q(1)$ is homogeneous. Equivalently, for $u_m = \sum_i a_i u_{m-i} + c$ put $u_m = v_m + K$ with $K = c/(1 - \sum_i a_i)$, valid whenever $\sum_i a_i \neq 1$. This is exactly the normalisation in Putnam 2015 B5, where $u_m = u_{m-1} + u_{m-3} + 1$ becomes $v_m = v_{m-1} + v_{m-3}$ with $K = 1/(1 - 2) = -1$, and the homogeneous cubic $t^3 = t^2 + 1$ has the supergolden root $\psi \approx 1.4656$ that governs the growth. If instead $q(1) = 0$ the particular solution is a polynomial in n of degree the multiplicity of the root 1, which is the discrete version of resonance.

Geometric forcing: $u_n = \alpha u_{n-1} + \beta^n$, $u_0 = 0$, has solution

$$u_n = \frac{\beta(\beta^n - \alpha^n)}{\beta - \alpha} \quad (\alpha \neq \beta), \qquad u_n = n\beta^n \quad (\alpha = \beta).$$

Putnam 2022 B5 is precisely this, with the apparent singularity of a closed form at one parameter value being nothing more than the resonance $\alpha = \beta$. Recognising that in advance is what removes the case split from the write-up.

Reading the operator the examiners handed you

Putnam 2015 B5 deserves a second look because it shows the highest form of this circuit. The problem hands you a linear relation whose operator factorises as

$$t^5 - t^4 - t^3 + 1 = (t - 1)(t + 1)(t^3 - t^2 - 1).$$

Each factor has a meaning: $(t - 1)$ is the constant inhomogeneity, $(t + 1)$ is a doubling under reversal, and the cubic is the intrinsic recursion of the counting problem. Factorising the operator before doing any counting predicts the recursion $u_m = u_{m-1} + u_{m-3} + 1$ in advance. That is what "read the cue" means at Putnam level: the examiners wrote the answer into the coefficients of the relation, and the factorisation reads it back out.

One more transform, cheap and frequently decisive.

Proposition 5.2.12 (Subsequences, Putnam 2016 A2)**.** *If r_1, r_2 are the roots of $t^2 - pt + q$ then r_1^τ, r_2^τ are the roots of a quadratic obtained from the power sums of r_1, r_2; for $\tau = 2$ it is $t^2 - (p^2 - 2q)t + q^2$ and for $\tau = 3$ it is $t^2 - (p^3 - 3pq)t + q^3$. So the every τth term subsequence of a linear recurrence is again a linear recurrence, of the same order, with explicitly computable coefficients.*

The Fibonacci case is worth memorising: $t^2 - t - 1$ has $p = 1$, $q = -1$, so the squared roots satisfy $t^2 - 3t + 1 = 0$, which is why $\varphi^2 = (3 + \sqrt{5})/2$ appears wherever alternate Fibonacci numbers do, as in Putnam 2016 A2 and Putnam 2019 B5.

- **WHEN** a sequence is defined by a fixed linear rule from a fixed number of previous terms **DO** write the characteristic polynomial, factor it, and write $x_n = \sum_i p_i(n)\lambda_i^n$ with $\deg p_i$ below the multiplicity of λ_i, fitting the p_i from the first k terms **BECAUSE** the recurrence says $q(E)x = 0$, and $\ker q(E)$ has exactly that basis by Theorem 5.2.3

- **WHEN** you must prove two explicitly defined sequences are equal **DO** show both satisfy the same order k recurrence and check k consecutive terms **BECAUSE** a solution is determined by any window of k consecutive terms, Proposition 5.2.5

- **WHEN** a quadratic invariant such as $u_{n-1}^2 - u_n u_{n-2}$ is asserted **DO** write the companion matrix, form the 2×2 matrix of consecutive terms and take determinants **BECAUSE** det is multiplicative and the state matrix is multiplied by the companion matrix each step, so the invariant is $(\det T)^{n-1}$ times its initial value

- **WHEN** the problem hands you a relation among shifted values with numerical coefficients **DO** factorise that polynomial in the shift and read off one meaning per factor **BECAUSE** polynomials in E commute, so the relation is the composite of independent annihilators, each of which you can solve separately

- **WHEN** a recurrence has an added constant or an added geometric term **DO** find one particular solution by the resonance rule and add the homogeneous general solution **BECAUSE** the solution set of a linear equation is a coset of the kernel

- **WHEN** all the power sums of an unknown multiset are given **DO** convert to elementary symmetric functions by Newton's identities and study the polynomial **BECAUSE** power sums of roots satisfy the linear recurrence whose characteristic polynomial is that polynomial, so the coefficients are the real data

5.3 First order recursions: the whole family solved

Order one is where most Putnam recursions actually live, and there is no reason ever to be stuck on one. There are exactly four shapes.

Shape 1: pure geometric

$c_{m+1} = rc_m$ has solution $c_m = r^m c_0$. Trivial, but say aloud what it buys: the geometric series $\sum r^m = 1/(1-r)$ for $|r| < 1$ is nothing but this recursion summed, which is why Putnam 2013 A3 can expand every $1/(1 - x^{k+1})$ as a series and why Putnam 2015 B4 evaluates three separate geometric sums with ratios $1/15$, $2/5$ and $2/3$. The multivariable version, one geometric factor per extreme ray of a simplicial cone, is the lattice point count of Putnam 2015 B4 in full.

Shape 2: affine, and the fixed point

Proposition 5.3.1 (Affine recursions). *For $u_{k+1} = au_k + b$ with $a \neq 1$, the map has the unique fixed point $c = b/(1-a)$, and substituting $w_k = u_k - c$ gives $w_{k+1} = aw_k$. Hence*

$$u_k = a^k(u_0 - c) + c, \qquad c = \frac{b}{1-a}.$$

If $|a| < 1$ every orbit converges geometrically to c; if $a = 1$ the recursion is arithmetic and $u_k = u_0 + kb$.

Proof. $u_{k+1} - c = au_k + b - c = a(u_k - c) + (ac + b - c) = a(u_k - c)$ since $c(1-a) = b$. □

Do not memorise the closed form. Memorise the move: subtract the fixed point. The fixed point is where the recursion says nothing changes, so measuring from it removes the constant, and what is left is pure scaling. This single proposition is what solves Putnam 2024 B4 (the mean of a Markov chain with affine drift, $e_{k+1} = \lambda e_k + c$), Putnam 2022 B6 (an affine contraction turning an infinite chain of inequalities into one global bound by continuity), Putnam 2015 B4, Putnam 2020 A1 and Putnam 2007 A4 (repunits: $R_{k+1} = 10R_k + 1$, fixed point $-1/9$, hence $R_k = (10^k - 1)/9$), Putnam 2019 A3, and the geometric sum $S_{m+1} = wS_m + 1$ everywhere it appears.

A useful corollary of the contraction case: if f is continuous and satisfies $f(u) \leq f(T(u))$ for an affine contraction T, then iterating gives $f(u) \leq f(T^k(u))$ for every k, and letting $k \to \infty$ gives $f(u) \leq f(c)$ at the fixed point. An infinite chain of inequalities has been collapsed into one bound, and continuity did the collapsing. That is the whole boundedness step of Putnam 2022 B6.

Shape 3: variable coefficient, and the summation factor

Theorem 5.3.2 (Summation factor). *Let $f(k) = \lambda_k f(k-1) + R(k)$ for $k > k_0$, with all $\lambda_k \neq 0$. Put $P_k = \prod_{j=k_0+1}^{k} \lambda_j$. Then*

$$f(k) = P_k\Big(f(k_0) + \sum_{j=k_0+1}^{k} \frac{R(j)}{P_j}\Big).$$

Proof. Divide the recursion by P_k and use $P_k = \lambda_k P_{k-1}$: it becomes $f(k)/P_k = f(k-1)/P_{k-1} + R(k)/P_k$, which telescopes. □

The quantity $1/P_k$ is the summation factor, and it is exactly the discrete integrating factor from first order linear ODEs. Two instances. In Putnam 2017 B6 the multiplier is $-k$, so the summation factor is $k!$, and after normalising the recursion has constant multiplier -1 with binomial inhomogeneity, at which point Pascal's rule supplies the particular solution instantly: since $\binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n}{k}$, the sequence $g(k) = \binom{n-1}{k}$ solves $g(k) + g(k-1) = \binom{n}{k}$, so the general solution is $\binom{n-1}{k} + C(-1)^k$. Bank that pair. In Putnam 2020 A4 the purely multiplicative case is used: $d_N = r(N)d_{N-1}$ has solution $d_N = d_1 \prod_{i=2}^{N} r(i)$, and with $r(N) = -1/(N+1)$ the product is a reciprocal factorial with alternating sign, so summing the increments produces a truncated exponential series and the constant $1/e$ appears without ever being guessed.

Shape 4: polynomial coefficients, and why you normalise first

A recursion whose coefficients are polynomials in the index, such as

$$a_N = a_{N-1} + \frac{a_{N-2}}{N-1},$$

has no closed form in the sense of Theorem 5.2.3. The move is to normalise: guess a substitution $a_N = \phi(N)u_N$ that makes the coefficients collapse, then look for a first order recursion on the increments. In Putnam 2020 A4 the substitution $a_N = (N+1)u_N$ turns the relation into $(N+1)u_N = Nu_{N-1} + u_{N-2}$, and subtracting $(N+1)u_{N-1}$ from both sides gives $(N+1)d_N = -d_{N-1}$

for $d_N = u_N - u_{N-1}$. That is Shape 3, solved by the product, and summing the increments gives the closed form. Notice what was not done: nobody guessed the answer and verified it by induction. The derivation produced it.

Two further recognisers for polynomial coefficient recursions.

First, they are exactly the recursions whose generating functions satisfy linear differential equations. If $y = \sum_k b_k x^k$ satisfies $y' = Q(x)y$ with $Q = \sum_i q_i x^i$ a polynomial, then $(k+1)b_{k+1} = \sum_i q_i b_{k-i}$, and in terms of the normalised coefficients $c_k = k!\, b_k$ this reads

$$c_{k+1} = \sum_{i \geq 0} q_i \, k(k-1)\cdots(k-i+1)\, c_{k-i},$$

a linear recursion with falling factorial coefficients. This is the engine of Putnam 2022 B1, where $y = e^{P(x)}$ satisfies $y' = P'y$ and the whole arithmetic content of the problem lives in the c_k, not the b_k. The same shape with $Q = 1 + x$ produces the involution numbers and with $Q = e^x$ the Bell numbers.

Second, a running sum in the recursion is removed by differencing. If $c_N a_N = b_N + \lambda \sum_{m<N} a_m$, write the same identity at $N-1$ and subtract: the two running sums differ by exactly one term, and you are left with a short recursion. In Putnam 2020 A4, $(N-1)w(N) = w(N-1) + 2\sum_{m=1}^{N-2} w(m)$ becomes $w(N) = w(N-1) + w(N-2)/(N-1)$. In Putnam 2017 A5, $nu_n = \sum_{k<n} f(u_k)$ becomes $nu_n - (n-1)u_{n-1} = f(u_{n-1})$, and applied to probability generating functions it gives $P_n = \frac{X+n-1}{n} P_{n-1}$, a product. Bank the general solution of that last shape too: $u_n = \frac{n+\alpha-1}{n} u_{n-1}$ has solution $u_n = u_0 \binom{n+\alpha-1}{n}$, a rising factorial over $n!$.

Finally, the polynomially forced additive shape. If $u_n = u_{n-1} + q(n)$ with $\deg q = d$, then $u_n = u_{n_0} + \sum_{m>n_0}^{n} q(m)$, which is a polynomial in n of degree $d+1$. Solve it either by Faulhaber summation or, faster, by positing an undetermined polynomial of degree $d+1$ and matching first differences. Every peeling induction with a polynomial per step cost is this shape; Putnam 2007 A6 is one, and Putnam 2023 A1 and Putnam 2024 B5 are the sum of squares and general partial sum instances. The underlying identity is that power sums come from telescoping $(m+1)^{k+1} - m^{k+1}$, which is the recursion behind Faulhaber's formulas.

- **WHEN** a first order recursion has a constant added **DO** find the fixed point $b/(1-a)$ and subtract it **BECAUSE** the fixed point is the value the map leaves alone, so measuring from it makes the map pure multiplication, Proposition 5.3.1
- **WHEN** a first order recursion has index dependent multipliers **DO** divide by the running product P_k of the multipliers and telescope **BECAUSE** that is the discrete integrating factor, Theorem 5.3.2
- **WHEN** a recursion involves the sum of all earlier terms **DO** write the same identity one index down and subtract **BECAUSE** consecutive running sums differ by one term, so subtraction converts an unbounded memory into a bounded one
- **WHEN** the coefficients of a recursion are polynomials in the index **DO** look for a normalising substitution $a_N = \phi(N)u_N$ that clears them, then pass to increments **BECAUSE** polynomial coefficient recursions have no exponential closed form, but their increments frequently satisfy Shape 3, which does
- **WHEN** a recursion is multiplicative, $x_n = cx_{n-1}^k$ **DO** take logarithms of absolute values and handle the sign separately **BECAUSE** logs turn products into sums and powers into multiplication, converting the recursion into a linear one in $\log|x_n|$

The last circuit deserves its statement in full, since Putnam 2018 A2 turns on it. For $x_n = cx_{n-1}^k$ with $k \neq 1$, unrolling the linear recursion in $\log|x_n|$ gives

$$\log|x_n| = \log|c|\,\frac{k^{n-1}-1}{k-1} + k^{n-1}\log|x_1|.$$

For even k the sign is absorbed into $\operatorname{sgn}(c)$ after one step, because $x_{n-1}^k \geq 0$; for odd k the sign itself recurses as $\operatorname{sgn}(x_n) = \operatorname{sgn}(c)\operatorname{sgn}(x_{n-1})$. In Putnam 2018 A2 that observation alone collapses $\det M_n = -(\det M_{n-1})^2$ to the constant -1. The same logarithm move solves the random multiplicative recursion $x_{n+1} = x_n U_{n+1}$ of Putnam 2020 B3, which becomes an additive random walk.

5.4 Telescoping, the method of differences, and creative telescoping

The definition, stated so that the search becomes systematic

Definition 5.4.1 (Discrete antiderivative)**.** G is a *discrete antiderivative* (an antidifference) of the sequence t if $t_k = G(k) - G(k-1)$ for all k in the range of interest. Then

$$\sum_{k=a}^{b} t_k = G(b) - G(a-1).$$

That identity is the discrete Fundamental Theorem of Calculus, and the sum $\sum t_k$ is nothing but the solution of the first order recursion $G(k) = G(k-1) + t_k$. Every telescoping sum you will ever meet is that recursion, and the solution method is always the same: identify G, then evaluate at the two endpoints. A-level taught you to do this once the split had been handed to you. The Putnam hands you nothing; you must construct G.

Four constructions cover almost everything.

Consecutive factors in the denominator. If t_k has the shape $\frac{g(k)-g(k-1)}{g(k-1)g(k)}$ then $t_k = \frac{1}{g(k-1)} - \frac{1}{g(k)}$, so $G(k) = -1/g(k)$. Putnam 2019 B2 is exactly this with $g(k) = \cos^2(k\pi/2n)$, and without it the problem becomes an asymptotic estimate with error control, which costs vastly more ink. The baby case $\frac{1}{m(m-1)} = \frac{1}{m-1} - \frac{1}{m}$ is what makes the answer to Putnam 2016 B6 exactly 1 rather than some transcendental constant.

Prefix sums. Set $s_0 = 0$ and $s_i = s_{i-1} + x_i$. Then every contiguous block sum is $s_j - s_i$. This converts a question about the existence of a block with a numeric property into a question about a pair of partial sums, at which point pigeonhole on residues (discrete) or on fractional parts (continuous) finishes it. That is the backbone of Putnam 2006 B2, and it generalises verbatim to any associative combination: sums, products, exclusive ors.

Step n into n steps of size one. For any function,

$$f(x+n) - f(x) = \sum_{i=0}^{n-1} \big(f(x+i+1) - f(x+i)\big).$$

So any hypothesis about a step n gap can be reconciled against a hypothesis about step one gaps. That is the pivot of the entire necessity proof in Putnam 2010 A2, and it generalises to reconciling a step n relation against a step k relation by telescoping in blocks of k. The same identity in the shape "$\sum_n (g(n) - g(n+m))$ collapses to a fixed head of m terms minus the tail" is what makes mean zero periodic weights tractable in Putnam 2017 B4.

The invertible wrapper. This is the one that looks like an inspiration and is not. If a recursion becomes, after applying an invertible continuous h to both sides,

$$h(x_{n+1}) = h(x_n) - x_n,$$

then $x_n = h(x_n) - h(x_{n+1})$ exactly, so $\sum_{n\geq 0} x_n = h(x_0) - \lim_N h(x_N)$. In Putnam 2016 B1 the recursion is $x_{n+1} = \ln(e^{x_n} - x_n)$; exponentiating gives $e^{x_{n+1}} = e^{x_n} - x_n$, which is the displayed shape with $h = \exp$, and the sum is e^{x_0} minus the limit. There is nothing to guess: the cue is a logarithm wrapped around an expression, and the reaction is to strip the wrapper by applying its inverse to both sides. The same works with $h = \sinh$, $h = \tan$, or any cancellable wrapper.

Products telescope too

Proposition 5.4.2 (Telescoping certificate for a product). *To evaluate $\prod_k f(y_k)$ along an orbit $y_{k+1} = \sigma(y_k)$, look for functions P, Q of one variable with*

$$f(t) = \frac{P(\sigma(t))}{P(t)} \cdot \frac{Q(t)}{Q(\sigma(t))}.$$

Then the partial product up to index n collapses to $\frac{P(y_{n+1})}{P(y_0)} \cdot \frac{Q(y_0)}{Q(y_{n+1})}$, however complicated f looks termwise.

This is the mechanism in Putnam 2014 A3, where $\sigma(t) = t^2$ and infinitely many factors become two boundary evaluations. The same shape, with the auxiliary partial product built from the recursively defined sequence itself, converts the sum in Putnam 2011 A2 into a two term closed form: there, solving $b_n = b_{n-1}a_n - k$ for a_n exposes $a_n = (b_n + k)/b_{n-1}$ as a ratio of shifted terms, and the auxiliary product $T_n = \prod_i b_i/(b_i + k)$ makes the nth term a constant multiple of $T_{n-1} - T_n$. Putnam 2014 A2 is the plainest instance of all: a product of consecutive integer reciprocal differences splits into two factorial telescopes.

For trigonometric sums the standard producer of a telescoping form is the product to sum identity: multiplying a sum of sines or cosines in arithmetic progression by $2\sin(\frac{d}{2})$ where d is the common difference converts each term into a difference of two sines. That is how Putnam 2020 A6 evaluates its sums, and it is the derivation of the Dirichlet kernel.

Summation by parts, and when telescoping is guaranteed

Proposition 5.4.3 (Abel summation). *With $C_m = \sum_{\ell \le m} c_\ell$ and tails $T_\ell = \sum_{m \ge \ell} a_m$ one has $T_\ell - T_{\ell+1} = a_\ell$, and*

$$\sum_{\ell \ge 1} c_\ell T_\ell = \sum_{m \ge 1} a_m C_m$$

whenever both sides converge absolutely.

This is the discrete integration by parts and it is the tool behind Putnam 2015 B6 and behind the polynomial identity in Putnam 2018 B2: multiplying a power series by $(z - 1)$ replaces the coefficient sequence by its own difference sequence plus boundary terms, so a coefficient sequence in arithmetic progression collapses to a geometric series.

Finally, the honest answer to "how do I know a telescoping form exists". For hypergeometric summands, those whose consecutive term ratio is a rational function of the index, Gosper's algorithm decides in finitely many steps whether a closed form antidifference exists, and Zeilberger's creative telescoping extends this to sums depending on a parameter, producing a linear recursion in that parameter together with a certificate. This is a standard body of theory you must know by name; the reference is the book $A = B$ by Petkovsek, Wilf and Zeilberger. You will not run the algorithm in the arena. What you take from it is the decision rule: on a finite hypergeometric sum, searching for G is the first move, because the search is a systematic procedure and not a lottery. That is what separates the ideal route from the estimation route in Putnam 2019 B2, and it is the honest reason such sums close in Putnam 2020 A2, where a Pascal rule application turns the sum into the self referential relation $S_k = 2S_{k-1} + \frac{1}{2}S_k$, hence $S_k = 4S_{k-1}$ and $S_k = 4^k$.

- **WHEN** a finite sum has terms that are ratios with consecutive shifted factors in the denominator **DO** write $t_k = G(k) - G(k-1)$ with G built from one of those factors and evaluate at the endpoints **BECAUSE** the sum is the solved form of $G(k) = G(k-1) + t_k$, so only the boundary survives
- **WHEN** a question asks for the existence of a contiguous block with a numeric property **DO** form the prefix sums and apply pigeonhole to them **BECAUSE** block sums are differences of prefix sums, so a statement about blocks becomes a statement about a pair of points

- **WHEN** a hypothesis is stated for a step of size n and you want it for step one **DO** telescope the step n difference into n consecutive step one differences **BECAUSE** the identity is exact for any function, so the two hypotheses can be compared directly
- **WHEN** a recursion has a function wrapped around it, such as a logarithm or an arctangent **DO** apply the inverse function to both sides and look for the form $h(x_{n+1}) = h(x_n) - x_n$ **BECAUSE** if that form appears the sum of x_n telescopes exactly, converting an unsolvable recursion into an exactly summable one
- **WHEN** an infinite product runs along the orbit of a map **DO** seek P, Q with $f(t) = \frac{P(\sigma t)}{P(t)} \frac{Q(t)}{Q(\sigma t)}$ **BECAUSE** that factorisation makes the partial product collapse to two boundary evaluations
- **WHEN** a finite sum's consecutive term ratio is a rational function of the index **DO** search for an antidifference before attempting any estimate **BECAUSE** Gosper's algorithm guarantees the search terminates, so it is a procedure and not a gamble

5.5 Induction, in every form the Putnam needs

School teaches one form of induction and gives you problems built so that it works. The Putnam uses six, and the difference between a solved and an unsolved problem is usually the choice of which.

The six forms

Ordinary induction needs no comment beyond this: the recursion preserves a property, so verify the base and verify the step. Putnam 2016 B1 uses it to establish positivity of every term, which is exactly the hypothesis that turns the telescoping identity into a strict monotonicity statement. Putnam 2015 A2 uses it to establish that an integer defined recursion stays a sequence of positive integers, without which "the ratio is rational" would be meaningless.

Strong induction assumes the statement for all smaller indices. Use it whenever the recursion reaches back more than one step, or reaches back to an index you can only bound. Putnam 2015 B2 needs a refined version: the step at index q needs the statement for all $q' < q$ plus a fact resolved at index q itself, which is legitimate only because index q is provably settled at an earlier stage of the same construction. Say this explicitly when you write it up: separate "the index that must already be resolved" from "the index currently being proved", and verify the ordering. That is the general licence for any process that references its own earlier output at a computable, provably earlier index.

Downward induction runs from a large case to a small one. Putnam 2008 A1 is the model: a three term cyclic identity is promoted to an n term cyclic identity by applying the base identity to the last two edges of the cycle, replacing them with one shortcut chord, so $S(n) = S(n-1)$; descend to $S(2)$, which is antisymmetry. The general shape is the shortcut chord reduction, and it promotes any local three term relation to a global statement along an arbitrarily long cycle or path.

Two variable induction inducts on a pair, usually on the sum of the two parameters. Putnam 2006 B4 does this with a cube slicing recursion: the generic branch gives $f(n,k) \le 2f(n-1,k-1)$ and the degenerate branch gives $f(n,k) \le f(n-1,k)$, and induction on $n+k$ yields $f(n,k) \le 2^k$. The additive sibling $g(n,d) \le g(n-1,d) + g(n-1,d-1)$ with $g(n,0) = 1$ solves to a partial sum of binomial coefficients, which is the Sauer and Shelah bound. Putnam 2024 B5 is a two parameter recurrence solved by inducting on one parameter after guessing the closed form from the first two or three cases.

Structural induction runs along the covering relations of a partially ordered set rather than along $\mathbb{N}$. Putnam 2013 B3 uses it: the hypothesis says the poset is graded by cardinality and every nonempty element covers one exactly one size smaller, which supports strong induction on rank

with the base case obtained by iterating the covering property down to the minimum. The same induction applies to the flats of a matroid or a modular chain in a subgroup lattice.

Well founded induction on a designed measure is the form nobody teaches and everybody needs. Induction does not require the measure to be the size of the object. Any nonnegative integer valued quantity that strictly decreases at each step licenses a valid strong induction, even while the objects themselves grow arbitrarily large. In Putnam 2009 B1 that measure is the largest prime factor of numerator or denominator; the numbers explode while the measure descends. In Putnam 2014 B6 the measure is the difference between the target's denominator and the current cell's denominator in a Stern and Brocot descent. The protocol is three steps: identify the quantity, prove strict decrease, check the floor case.

Strengthening the hypothesis, which is the actual skill

Here is the pattern that decides more Putnam problems than any other in this chapter. You attempt an induction, the step fails, and the reason it fails is that the inductive hypothesis is too weak to feed the step. The response is never to abandon induction. The response is to prove something stronger.

Three canonical strengthenings.

Generalise the parameter. Whenever an extremal argument lets you discard part of a configuration without loss, the reduced instance is a smaller instance of the same problem, so the claim must be restated for every relevant size before the reduction can serve as an inductive step. Putnam 2018 A3 is this exactly.

Carry the whole chain of previous inequalities, not just the last one. In Putnam 2018 B4 the recursion is $x_{n+1} = 2x_n x_{n-1} - x_{n-2}$ and the aim is to show the orbit escapes zero. The induction hypothesis that works is the full chain $1 = |x_0| \le |x_1| \le \cdots \le |x_n|$; then the reverse triangle inequality gives

$$|x_{n+1}| \ge 2|x_n||x_{n-1}| - |x_{n-2}| \ge |x_n|\big(2|x_{n-1}| - 1\big) \ge |x_n|,$$

and the chain extends. A hypothesis about one term would not have closed.

Replace an inequality by a comparison sequence. If $a_{n+1} = f(a_n)$ with f strictly increasing, then a sequence c_n with $f(c_n) \ge c_{n+1}$ is a *lower solution*, and $a_0 > c_0$ forces $a_n > c_n$ for all n. Proof: monotonicity of f transfers the inequality forward, then the lower solution condition closes it. This is Putnam 2012 B4, where $c_n = \log(n+1)$ and the tangent line inequality supplies $f(c_n) \ge c_{n+1}$. The strengthening here is that you invent the comparison sequence rather than trying to prove the crude bound directly.

Related and equally reusable: the Fibonacci domination lemma of Putnam 2012 A1. If positive reals satisfy $x_{i+2} \ge x_i + x_{i+1}$ with $x_1, x_2 \ge t$, then $x_i \ge F_i t$ by a two step induction. The general form is $x_{n+2} \ge \alpha x_n + \beta x_{n+1}$ with $\alpha, \beta > 0$: the same two lines give growth at least like the dominant root of $t^2 = \beta t + \alpha$. In the Fibonacci case that root is φ, and $F_{12} = 144$ is the predictable output of exponential growth, not a number to memorise.

Bootstrapping: induction on quality rather than on index

A bootstrap is an induction whose index is the strength of the conclusion. You have a crude bound; you feed it into the equation; out comes a better bound; repeat.

Regularity bootstrap: if $f' = \Phi(f)$ with Φ smooth, then $f \in C^k$ implies $\Phi \circ f \in C^k$ implies $f' \in C^k$ implies $f \in C^{k+1}$, so $f \in C^\infty$. That is Putnam 2010 B5 and Putnam 2009 B5.

Exponent bootstrap: given $y' = F(x, y)$ and a crude bound $|y| \le C_0 x^{\alpha_0}$, substituting gives a bound on y' and hence a new exponent $\alpha_1 = T(\alpha_0)$. The true exponent is the attracting fixed point of T. In Putnam 2009 B5, $T(1) = 1/3$ and $T(1/3) = 1/3$, so one iteration suffices.

The most demanding instance in the archive is Putnam 2025 B6, where the bootstrap is stated with "for all sufficiently large n" on both sides so that it iterates without tracking constants: from $g(n) \ge n^s$ one derives $g(n) \ge n^t$ for every $t < \varphi(s) = rs^2 + 1$, and the orbit $s_{j+1} = \varphi(s_j) - \varepsilon$ is

unbounded exactly when $\varphi(s) - s$ is bounded below by a positive constant, which for $\varphi(s) = rs^2 + 1$ means $r > 1/4$. The exponent map is a quadratic iteration, and the answer $r = 1/4$ is the threshold where its fixed points appear. Read that again: the answer to the problem is the bifurcation value of an iterated quadratic, located in the first minute by a power law ansatz.

- **WHEN** an inductive step fails because the hypothesis does not supply enough **DO** strengthen the statement, most often by generalising a parameter, by carrying the entire chain of earlier inequalities, or by inventing an explicit comparison sequence **BECAUSE** the induction is a machine that consumes its own output, so a weak output starves the next step; a stronger statement can be easier to prove
- **WHEN** the objects grow but some auxiliary quantity shrinks **DO** induct on the auxiliary quantity, not on the objects **BECAUSE** well founded induction needs only a strictly decreasing nonnegative integer measure, and the objects themselves are irrelevant to that
- **WHEN** a relation is given for three terms and you want it for n **DO** apply the three term relation to the last two members of the n cycle to shortcut them into one, giving $S(n) = S(n-1)$, and descend **BECAUSE** a local relation propagates along a cycle by repeated shortcutting; the base case is the degenerate two term statement
- **WHEN** a bound is available but too crude **DO** substitute the crude bound back into the defining relation and read off the improved bound, then iterate **BECAUSE** the map from old exponent to new exponent is an iteration whose fixed point is the truth, so the bootstrap converges to the right answer
- **WHEN** a monotone recursion must be compared against an explicit function **DO** find a lower solution c_n with $f(c_n) \geq c_{n+1}$ and check one initial inequality **BECAUSE** monotonicity of f transfers a strict inequality forward, so one comparison at the start propagates forever
- **WHEN** the recursion refers to an earlier index that is not $n-1$ **DO** use strong induction, stating separately which index must already be resolved and proving it precedes the current one **BECAUSE** strong induction is valid for any reference to a provably earlier index; the proof obligation is the ordering, and it must be discharged in writing

5.6 Iteration of maps: fixed points, basins, periodicity

The three facts that classify an orbit

A first order recursion $a_{n+1} = \phi(a_n)$ is the orbit of a point under a map. Three facts do most of the work.

Proposition 5.6.1 (The limit is a fixed point)**.** *If ϕ is continuous and $a_n \to L$ then $L = \phi(L)$.*

Proof. Pass to the limit in $a_{n+1} = \phi(a_n)$, using continuity on the right and the fact that a_{n+1} has the same limit as a_n. □

An infinite process has been reduced to one algebraic equation. This is how the limit is identified in Putnam 2016 B1. It also tells you what to compute first on any convergence problem: solve $\phi(L) = L$ before anything else, because the candidate limits are the only possible answers.

Definition 5.6.2 (Multiplier)**.** If ϕ is differentiable and $\phi(p) = p$, the *multiplier* is $\phi'(p)$. The fixed point is *attracting* if $|\phi'(p)| < 1$, *repelling* if $|\phi'(p)| > 1$, and *neutral* if $|\phi'(p)| = 1$.

Proposition 5.6.3. *If $|\phi'(p)| < 1$ then there is a neighbourhood of p from which every orbit converges to p, geometrically at rate about $|\phi'(p)|$. If $|\phi'(p)| > 1$ then p is repelling for ϕ and attracting for ϕ^{-1}, with rate $1/\phi'(p)$.*

Proof. Near p, $|\phi(x) - p| = |\phi'(p)||x - p|(1 + o(1))$; choose the neighbourhood small enough that the factor is bounded by some $\rho < 1$ and iterate. □

In Putnam 2012 A3 the map is $\phi(t) = t^2/(2 - t^2)$ with $\phi'(0) = 0$ (attracting) and $\phi'(1) = 4$ (repelling), which is precisely why orbits starting in $(-1, 1)$ run to 0 and not to 1; that is the geometric fact underlying the whole uniqueness argument. In Putnam 2021 B6 the map is $\phi(u) = 3u^2 - 2u^3$ with fixed points $0, \frac{1}{2}, 1$ and multipliers $0, \frac{3}{2}, 0$; the constant $2/3$ in the problem statement is exactly $1/\phi'(\frac{1}{2})$, the reciprocal multiplier at the repelling fixed point. Read the constants in the problem as multipliers and the required lemma announces itself before any computation.

Proposition 5.6.4 (Monotone convergence and cobwebs)**.** *If ϕ is increasing on an interval mapped to itself, every orbit is monotone: if $a_1 > a_0$ then $a_{n+1} > a_n$ for all n, so the orbit converges to a fixed point or diverges. If ϕ is decreasing, $\phi \circ \phi$ is increasing, so the even and odd subsequences are each monotone and the orbit converges to a fixed point or to a two cycle.*

That is the entire content of cobweb diagrams, stated so that you can write it down rather than draw it. Together with the fact that an increasing sequence either converges or tends to infinity, and that a convergent sequence has increments tending to zero, it settles the qualitative behaviour of most first order recursions on sight. Putnam 2006 B6 uses exactly that pair to get $a_n \to \infty$ in three lines, which is the hypothesis every later asymptotic step needs.

For the contraction principle at the level the Putnam uses it: if $|\phi(x) - \phi(y)| \le \rho|x - y|$ with $\rho < 1$ on a complete interval mapped to itself, there is a unique fixed point and every orbit converges to it geometrically. You will almost never need the abstract Banach statement; the affine case of Proposition 5.3.1 with $|a| < 1$ is what actually appears, as in Putnam 2022 B6.

Neutral fixed points, where the interesting asymptotics live

When $\phi'(p) = 1$ the classification above says nothing and the convergence is not geometric.

Theorem 5.6.5 (Neutral fixed point asymptotics, Putnam 2020 A3)**.** *Let ϕ be smooth with $\phi(0) = 0$, $\phi'(0) = 1$ and $\phi(x) = x - cx^{p+1} + O(x^{p+2})$ with $c > 0$, $p \ge 1$. For small $x_0 > 0$ the orbit $x_{n+1} = \phi(x_n)$ is positive, strictly decreasing, tends to 0, and satisfies $x_n \sim (cpn)^{-1/p}$.*

Proof. Put $b_n = x_n^{-p}$. Then

$$b_{n+1} = x_n^{-p}\big(1 - cx_n^p + O(x_n^{p+1})\big)^{-p} = x_n^{-p}\big(1 + cpx_n^p + O(x_n^{p+1})\big) = b_n + cp + o(1),$$

so $b_n = cpn + o(n)$ by Cesaro, and $x_n = b_n^{-1/p}$. □

For $\phi = \sin$ one has $c = 1/6$, $p = 2$ and $x_n \sim \sqrt{3/n}$. The general lesson is stated as a search: find the power of x whose increments are asymptotically constant, and the recursion becomes additive. That is the master method for every slowly decaying recursion.

Higher order recursions as maps, periodicity, and the Lyness phenomenon

Proposition 5.6.6. *A recurrence $x_{n+2} = F(x_n, x_{n+1})$ is the iteration of $T(x, y) = (y, F(x, y))$ on the space of consecutive pairs. An order k recurrence is a map on k tuples. Periods, orbits, invariants and fixed points of the sequence are those of T.*

Once you say that, periodicity of a sequence becomes finite order of a map, and a rational invariant of the map foliates the state space into level sets on which the dynamics is simpler. Both halves are used in Putnam 2022 A3.

Theorem 5.6.7 (Lyness map)**.** *The recurrence* $x_{n+2} = (x_{n+1} + a)/x_n$ *preserves*

$$I(x, y) = \frac{(x+1)(y+1)(x+y+a)}{xy}.$$

For $a = 1$ *it is periodic with period* 5 *over any field, wherever defined: with* $x_1 = x$, $x_2 = y$ *one computes* $x_3 = (1+y)/x$, $x_4 = (1+x+y)/(xy)$, $x_5 = (1+x)/y$, $x_6 = x$, $x_7 = y$.

Proof. For the invariant, substitute $(x, y) \mapsto (y, z)$ with $z = (y + a)/x$. Then $z + 1 = (x + y + a)/x$ and $y + z + a = (y + a)(x + 1)/x$ and $yz = y(y + a)/x$, so

$$I(y, z) = \frac{(y+1) \cdot \frac{x+y+a}{x} \cdot \frac{(y+a)(x+1)}{x}}{\frac{y(y+a)}{x}} = \frac{(x+1)(y+1)(x+y+a)}{xy} = I(x, y).$$

The period 5 computation for $a = 1$ is the five displayed substitutions, each one line. □

At $a = 1$ the invariant specialises to $(1 + x)(1 + y)(1 + x + y)/(xy)$, and requiring every iterate to be defined becomes the single condition $I(x, y) \neq 0$, which is why exactly the three lines $x = -1$, $y = -1$, $x + y = -1$ are removed in Putnam 2022 A3. That is the general use of a rational invariant for a birational map: an infinite condition (all iterates defined) becomes membership of a level set, which is a finite condition.

Two pieces of background you should know by name. First, the level sets of the general Lyness invariant are elliptic curves on which the map acts as translation; $a = 1$ is the degenerate case where the translation is by a point of order five, which is the structural reason for period five rather than an accident. Second, recursions of the shape $a_n a_{n+2} = a_{n+1}^b + 1$ are rank two cluster algebra exchange relations, and iterating them always produces Laurent polynomials in the initial data; this is the Laurent phenomenon, and it is why such recursions have surprising integrality. Putnam 2017 A2, with $u_n u_{n-2} = u_{n-1}^2 - 1$, is a small instance. Naming this places the guaranteed integrality inside a studied phenomenon instead of leaving it as a coincidence.

Two more iteration structures worth having

Escape time unfolding. If $F(x) = G(x)F(\sigma(x))$ outside a base region B, with $\sigma(x) < x$ and the iterates of σ reaching B after $k(x)$ steps, then

$$F(x) = \prod_{j=0}^{k-1} G(\sigma^j(x)) \cdot F_{\text{base}}(\sigma^k(x)),$$

and the natural partition of the domain is by escape time. Putnam 2008 A4 is this with $\sigma = \ln$, producing the tower $t_k = e^{t_{k-1}}$ and the closed form $x \ln x \ln^2 x \cdots$. The same unfolding handles $\sigma(x) = x/2$ (dyadic blocks) and $\sigma(x) = \sqrt{x}$ (double exponential blocks).

Self composition at powers of a prime. Since $b_{n+1} = f(b_n)$ means $b_n = f^{\circ n}(b_0)$, and $f^{\circ(m+n)} = f^{\circ m} \circ f^{\circ n}$, the iterates at indices 2^k satisfy $g_{k+1} = g_k \circ g_k$ for $g_k = f^{\circ 2^k}$. Working with the polynomial g_k, specifically its constant term and its linear coefficient, replaces an unmanageable orbit by two integers whose p-adic valuations you can track by induction. That is the first move and the bifurcation point of Putnam 2025 A6, where writing $g_k = A_k + B_k x + x^2 Q_k$ gives, to leading order, $A_{k+1} = A_k(1 + B_k + \cdots + B_k^{p-1})$ and $B_{k+1} = B_k^p$, so the valuations of A_k and of $B_k - 1$ each grow by exactly one per step.

- **WHEN** a sequence is defined by $a_{n+1} = \phi(a_n)$ and you are asked for a limit **DO** solve $\phi(L) = L$ first, then classify each fixed point by $|\phi'|$ **BECAUSE** the limit of a continuous recursion must be a fixed point, and the multiplier decides which fixed point an orbit can reach

- **WHEN** a constant in the problem statement equals $1/\phi'(p)$ or $\phi'(p)$ for some fixed point **DO** treat that constant as the multiplier and build the one step lemma around it **BECAUSE** the examiners derived the constant from the local rate at the fixed point, so reading it that way recovers their derivation
- **WHEN** a first order recursion converges but not geometrically **DO** compute $\phi(x) = x - cx^{p+1} + \cdots$ near the neutral fixed point and substitute $b_n = x_n^{-p}$ **BECAUSE** the substitution makes the increments asymptotically constant, so the orbit becomes an arithmetic progression, Theorem 5.6.5
- **WHEN** a rational recursion appears to be periodic **DO** write it as a map on pairs, compute the first few iterates symbolically, and look for a rational invariant **BECAUSE** periodicity of the sequence is finite order of the map, and an invariant converts the condition that all iterates exist into one algebraic condition, Theorem 5.6.7
- **WHEN** a functional or recursive relation compares $F(x)$ with $F(\sigma(x))$ for a contracting σ **DO** unfold along the orbit of σ and partition the domain by escape time **BECAUSE** each escape time class is a separate block on which the product is finite and explicit
- **WHEN** a recursion is asked about only at indices that are powers of a fixed number **DO** compose the map with itself instead of iterating it, and track a small number of coefficients of the composite **BECAUSE** $f^{\circ 2^{k+1}} = f^{\circ 2^k} \circ f^{\circ 2^k}$, so the data doubles in accuracy each step while staying two or three numbers wide

5.7 Nonlinear recursions and the substitutions that linearise them

A nonlinear recursion is solvable exactly when it is a linear one in disguise. There are five disguises worth knowing, and each has a cue you can read off the algebra.

The squaring family: $y + 1/y$ and $2\cos\theta$

Theorem 5.7.1 (Putnam 2014 A3)**.** *If $t_0 = y_0 + y_0^{-1}$ then $t_k = t_{k-1}^2 - 2$ has the closed form $t_k = y_0^{2^k} + y_0^{-2^k}$. Equivalently, writing $t_0 = 2\cosh\theta_0$, or $t_0 = 2\cos\theta_0$ when $|t_0| \le 2$, gives $t_k = 2\cosh(2^k\theta_0)$ or $2\cos(2^k\theta_0)$.*

Proof. $(y + y^{-1})^2 - 2 = y^2 + 2 + y^{-2} - 2 = y^2 + y^{-2}$, so the substitution conjugates $t \mapsto t^2 - 2$ to $y \mapsto y^2$. Induct. □

The cue is the shape $t^2 - 2$, or anything reducible to it by scaling. The reason it is natural is structural, not lucky: $u \mapsto u^2 - 2$ is conjugate to squaring, and squaring is the one nonlinear map whose iterates you can write down. In the trigonometric form, $t_k = 2T_{2^k}(t_0/2)$ where T_m is the Chebyshev polynomial of the first kind, using $T_m \circ T_n = T_{mn}$.

Chebyshev deserves its own statement because it appears on the Putnam constantly, in Putnam 2013 A6, Putnam 2013 B2, Putnam 2018 A3 and Putnam 2018 B4.

Definition 5.7.2. T_n is defined by $T_0 = 1$, $T_1 = z$, $T_{n+1}(z) = 2zT_n(z) - T_{n-1}(z)$, equivalently by $T_n(\cos\theta) = \cos(n\theta)$.

The two definitions agree because the recursion is the product to sum identity $2\cos\theta\cos n\theta = \cos(n+1)\theta + \cos(n-1)\theta$. Note that the Chebyshev recursion is itself a constant coefficient linear recursion in n with characteristic equation $\lambda^2 - 2z\lambda + 1 = 0$ and roots $e^{\pm i\theta}$ when $z = \cos\theta$, so Theorem 5.2.3 gives the closed form $T_n(z) = \cos(n \arccos z)$ directly. The governing product identity in polynomial form is $2T_mT_n = T_{m+n} + T_{|m-n|}$.

That identity is what solves a whole family of nonlinear three term recursions. For $x_{n+1} = 2x_n x_{n-1} - x_{n-2}$, substitute $x_n = \cos\theta_n$; then

$$2\cos\theta_n\cos\theta_{n-1} - \cos\theta_{n-2} = \cos(\theta_n + \theta_{n-1}) + \cos(\theta_n - \theta_{n-1}) - \cos\theta_{n-2},$$

so if the angles satisfy $\theta_n = \theta_{n-1} + \theta_{n-2}$, the last two terms cancel and $x_{n+1} = \cos\theta_{n+1}$ with the same additive rule. The nonlinear recursion on the values has become a linear recursion on the angles, solved by its characteristic equation: with $\theta_0 = 0$ and $\theta_1 = b$ you get $\theta_n = F_n b$. That is Putnam 2018 B4. The correct name for what makes this legal is d'Alembert's cosine functional equation $f(A+B) + f(A-B) = 2f(A)f(B)$, whose continuous solutions are $\cos(\lambda x)$, $\cosh(\lambda x)$, the constant 1 and the constant 0. Compact parameter gives bounded orbits, noncompact gives orbits escaping to infinity; the same recursion transported through cosh instead of cos is what proves escape.

Riccati: fractional linear maps become matrices

Proposition 5.7.3. *The recursion* $x_{n+1} = \dfrac{\alpha x_n + \beta}{\gamma x_n + \delta}$ *linearises: writing* $x_n = p_n/q_n$ *and* $\begin{pmatrix} p_{n+1} \\ q_{n+1} \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} p_n \\ q_n \end{pmatrix}$ *reproduces it exactly, so* x_n *is a ratio of two solutions of one constant coefficient linear recursion, and the dynamics is governed by the ratio of the eigenvalues.*

The special case you will meet most often is the reciprocal substitution. If $c_{k+1} = \dfrac{c_k}{a + bc_k}$, then $w_k = 1/c_k$ satisfies $w_{k+1} = aw_k + b$, which is Shape 2. In Putnam 2018 A5 the recursion is $c_{k+1} = c_k/(1 + c_k)$ with $c_1 = 1$, so $w_{k+1} = w_k + 1$, $w_k = k$, $c_k = 1/k$. The cue is an unknown appearing in both numerator and denominator; the reaction is to take reciprocals.

Products, floors, and invariants

Products become sums under logarithms. Already stated for $x_n = cx_{n-1}^k$; the same applies to random multiplicative recursions and to any local multiplicative constraint. In Putnam 2022 B4 a cyclic condition forces $d_{m+1} = 2^{\epsilon_m} d_m$ with $\epsilon_m \in \{-1, 0, 1\}$, so $d_m = 2^{\epsilon_1 + \cdots + \epsilon_{m-1}} d_1$ and going once round the cycle forces the exponents to sum to zero. An n dimensional search has become one scale plus a bounded walk.

Floors are removed by algebraic conjugates. For $x_{n+1} = ax_n + \lfloor x_n\sqrt{d} \rfloor$ with a, d positive integers, bound the floor strictly between $x_n\sqrt{d} - 1$ and $x_n\sqrt{d}$; if the conjugate $a - \sqrt{d}$ has absolute value less than one, multiplying by it produces a window of width less than one around $x_{n+1}\sqrt{d}$, which pins the next floor exactly and yields the floor free linear recurrence

$$x_{n+2} = 2a\, x_{n+1} - (a^2 - d)\, x_n,$$

with characteristic roots $a \pm \sqrt{d}$. This is Putnam 2007 B3 in full, with $a = 3$, $d = 5$, giving $x_{n+2} = 6x_{n+1} - 4x_n$ and the check $0 < 3 - \sqrt{5} < 1$. The general phenomenon has a name: $a + \sqrt{d}$ is a Pisot number, an algebraic integer greater than one all of whose other conjugates have absolute value less than one, and the fractional parts of powers of a Pisot number are controlled by the geometrically shrinking contribution of the small conjugates. Checking that the conjugate lies in $(0, 1)$ is what guarantees in advance that the elimination will work.

Invariant hunting for three term recursions. Substitute the recursion into a symmetric candidate $P(x_{n-1}, x_n, x_{n+1})$ and demand that P at index $n+1$ equals P at index n. For $x_{n+1} = 2x_n x_{n-1} - x_{n-2}$ this produces

$$I_n = x_{n-1}^2 + x_n^2 + x_{n+1}^2 - 2x_{n-1}x_n x_{n+1},$$

constant along orbits; the surface $x^2 + y^2 + z^2 - 2xyz = 1$ is the Cayley cubic, parametrised by $(\cos A, \cos B, \cos(A+B))$, which confirms that the cosine substitution above is structural. The

relatives are the Markov equation $x^2 + y^2 + z^2 = 3xyz$ and Vieta jumping. In Putnam 2018 B4 the invariant is not needed for the proof, but it is the algebraic fingerprint that tells you the parametrisation was not a guess.

Casoratian, the discrete Wronskian. If f_k and g_k both satisfy $u_{k+1} = \alpha_k u_k + \beta_k u_{k-1}$ then $C_k = f_k g_{k+1} - f_{k+1} g_k$ satisfies $C_{k+1} = -\beta_{k+1} C_k$, a first order recursion, so C_k is an explicit product. Constant Casoratian is the case $\beta \equiv -1$; the Fibonacci case $\beta \equiv 1$ gives Cassini's identity with the alternating sign. This is the discrete Abel identity, it is what makes a pinned Casoratian a natural object at all, and it generalises Putnam 2010 B4 to every second order linear recursion.

- **WHEN** a recursion is $t \mapsto t^2 - 2$ or reduces to it **DO** substitute $t = y + 1/y$, or $t = 2\cos\theta$ when $|t| \leq 2$ **BECAUSE** that substitution conjugates the map to squaring or to angle doubling, whose iterates are explicit, Theorem 5.7.1
- **WHEN** a recursion has the shape $x_{n+1} = 2x_n x_{n-k} - x_{n-2k}$ **DO** substitute $x_n = \cos\theta_n$ and read off the linear recursion the index arithmetic forces on θ_n **BECAUSE** the product to sum identity converts a product of cosines into a sum, so a multiplicative rule on values is an additive rule on angles
- **WHEN** the unknown appears in both numerator and denominator **DO** take reciprocals, or write the map as a 2×2 matrix acting on a ratio **BECAUSE** fractional linear maps are linear maps on homogeneous coordinates, so the orbit is a ratio of two solutions of one linear recursion
- **WHEN** a recursion contains a floor of an irrational multiple **DO** bound the floor, multiply by the algebraic conjugate, and check the conjugate lies strictly inside the unit interval **BECAUSE** a conjugate of absolute value less than one shrinks the error window below width one, which pins an integer exactly and eliminates the floor
- **WHEN** a three term nonlinear recursion has no obvious closed form **DO** look for a symmetric invariant by substituting the recursion into a candidate and matching **BECAUSE** an invariant confines the orbit to a surface, which converts an infinite dynamical question into a finite algebraic one

5.8 Recursions modulo m

Theorem 5.8.1 (Eventual periodicity, Putnam 2006 A3)**.** *Let $x_k = f(x_{k-1}, \ldots, x_{k-n})$ for $k > n$, with n fixed and f a fixed integer coefficient polynomial. Then for every modulus N the residue sequence $x_k \bmod N$ is eventually periodic.*

Proof. The state is the window $(x_{k-1}, \ldots, x_{k-n})$ modulo N, of which there are N^n. Among the first $N^n + 1$ windows two coincide, and the recursion is deterministic, so everything after the second repeats what followed the first. □

Theorem 5.8.2 (Pure periodicity)**.** *If in addition the step map on windows is invertible, that is if there is a fixed integer coefficient polynomial g with $x_{k-n} = g(x_{k-n+1}, \ldots, x_k)$, then the sequence extends uniquely to a doubly infinite one and the residues modulo every N are purely periodic, with no preperiod.*

Proof. Iterating a map on a finite set gives an orbit shaped like the letter ρ: a tail followed by a cycle. If the map is a bijection the tail is empty, since a repeat $v_{n_1} = v_{n_2}$ can be pulled back n_1 steps by the inverse to give $v_0 = v_{n_2 - n_1}$. □

For a constant coefficient linear recursion the state map is the companion matrix C over $\mathbb{Z}/N\mathbb{Z}$, so the criterion is concrete: the sequence is purely periodic modulo N whenever $\det C$ is a unit modulo N. Fibonacci has $\det\left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right) = -1$, a unit modulo every N, so the Fibonacci sequence is purely periodic modulo every N; that period is the Pisano period $\pi(N)$, with $\pi(2) = 3$, $\pi(4) = 6$,

$\pi(8) = 12$, $\pi(10) = 60$. Putnam 2021 B4 uses $\pi(4) = 6$ and the cycle $0, 1, 1, 2, 3, 1$ to conclude that F_m is 1 or 2 modulo 4 for odd m.

Three uses of this pair of theorems, all live on recent papers.

Putnam 2006 A3 uses Theorem 5.8.1 to convert an order 2006 recursion, whose characteristic polynomial is useless at that degree, into a finite state object; Theorem 5.8.2 then licenses running the recursion backwards from the initial block to find the desired pattern, with periodicity guaranteeing that whatever is found backwards also occurs forwards. Note the decision made there: for a sparse wide gap recursion such as $x_{k+1} = x_k + x_{k-m}$ with large m, a characteristic polynomial closed form is not a productive route, and recognising that is what redirects you to periodicity.

Putnam 2023 B2 and Putnam 2021 A5 use the multiplicative version: $r_{c+1} = gr_c \bmod m$ is an invertible map when $\gcd(g, m) = 1$, hence purely periodic with period $\mathrm{ord}_m(g)$; when g shares a factor with m, strip it, accept a preperiod, then work in the unit group. Putnam 2021 A5 needs the Carmichael function $\lambda(m)$, the exponent of $(\mathbb{Z}/m\mathbb{Z})^\times$, giving $n^{j+\lambda(m)} \equiv n^j$ for every unit n and hence a sequence of period dividing $\lambda(m)$; the general statement is that any sequence $j \mapsto \sum_g \chi(g)^j$ with values in a finite group of exponent e satisfies the exact linear recurrence $u_{j+e} = u_j$, characteristic polynomial $x^e - 1$.

Putnam 2010 A4 is the tower version. For $t_{j+1} = b^{t_j}$, the valuations satisfy $v_p(t_{j+1}) = t_j\, v_p(b)$, which explodes, and the residues satisfy $t_{j+1} \bmod M = b^{\,t_j \bmod \lambda(M)} \bmod M$, so a tower is evaluated by descending it, applying Euler or Carmichael at each level and taking care when $\gcd(b, M) > 1$.

- **WHEN** a recursion is over the integers and the question is about residues **DO** declare the state to be the window of the last n terms modulo N and apply pigeonhole **BECAUSE** there are only N^n windows and the step is deterministic, so a repeat forces periodicity, Theorem 5.8.1
- **WHEN** you want to run a recursion backwards, or want no preperiod **DO** check that the step map on windows is invertible, for a linear recursion that the companion determinant is a unit modulo N **BECAUSE** a bijection of a finite set has purely cyclic orbits, Theorem 5.8.2
- **WHEN** the order of the recursion is huge but the rule is sparse **DO** abandon the characteristic polynomial and argue by periodicity of the state **BECAUSE** a degree 2006 characteristic polynomial gives no usable closed form, whereas the state space argument does not care about the order
- **WHEN** an exponent tower must be evaluated modulo M **DO** reduce the exponent modulo $\lambda(M)$ and recurse down the tower **BECAUSE** the multiplicative order divides the exponent of the unit group, so only the exponent modulo $\lambda(M)$ matters, subject to a coprimality check

5.9 The transfer matrix method

Definition 5.9.1. A *transfer system* for a counting problem is a finite set of states, a vector $v(n)$ of counts refined by state, and a matrix M with $v(n+1) = Mv(n)$, possibly with an added constant vector b.

The reason this is a chapter topic and not a curiosity: whenever it applies, the counts satisfy a constant coefficient linear recurrence whose characteristic polynomial is that of M, and everything in Section 2 becomes available.

Proposition 5.9.2 (Where the states come from, Putnam 2015 B5)**.** *Suppose an element can always be deleted from a valid object of size $n+1$ to leave a valid object of size n. Then the count at level $n+1$ is the sum over level n objects of the number of legal reinsertion slots. The features of an object that this slot count mentions form the state space; grouping objects by those features gives $v(n+1) = Mv(n) + b$, hence a linear recurrence.*

That is a procedure, not an observation. Protocol: (1) find a canonical deletion; (2) ask what you must know about the smaller object to count reinsertions; (3) that list is the state; (4) write the matrix; (5) take the characteristic polynomial.

Instances across the archive. Putnam 2015 B5 is the archetype, and its transfer system is cross checked by the rational generating function with denominator $(1-x)(1-x^2)(1-x-x^3)$. Putnam 2018 B6 has the transfer system $a(k,n)=\sum_{j\in A} a(k-1,n-j)$, one step per term, whose generating function solution is simply $G_k = P^k$ with $P(x)=\sum_{j\in A} x^j$. Putnam 2008 B6 is a transfer system over $\mathbb{F}_2$ coming from a banded Toeplitz matrix, where multiplying the symbol by the polynomial that clears the interior turns the band into the recursion $a_{i+2k+1}=a_i$, whose solution space has dimension $2k+1$; that dimension count is the whole case analysis. Putnam 2011 A6 diagonalises the transfer operator by Fourier analysis on a finite abelian group, so that $v_m=\sum_j c_j\mu_j^m w_j$ and the second largest eigenvalue modulus controls the rate of convergence; that is the same spectral method as Section 2, done in a basis adapted to the group.

Putnam 2025 A5 shows the vector version at its most useful. Let $F(s)$ be the vector of counts of permutations with pattern s refined by last entry. Appending an ascent replaces F by its vector of prefix sums, appending a descent by its vector of suffix sums. Both are linear operators, both preserve componentwise inequalities, and reversal exchanges them, so inequalities of the form $F(x)\le \operatorname{rev} F(y)$ propagate under the paired substitutions. This is the boustrophedon recursion, and for the alternating pattern it is the Seidel and Entringer triangle.

- **WHEN** a count is defined by building objects one element at a time **DO** find a canonical deletion, list what you must know to count reinsertions, and make that list the state **BECAUSE** a bounded state list gives a finite linear transfer system, hence a constant coefficient linear recurrence with a computable characteristic polynomial
- **WHEN** you have a transfer matrix and want asymptotics **DO** diagonalise and keep the dominant eigenvalue, using the second largest modulus for the error term **BECAUSE** $v_m=\sum_j c_j\mu_j^m w_j$, so the largest $|\mu_j|$ dictates the growth and the next one dictates the rate of convergence

5.10 Generating functions

Formal power series are algebra, not analysis

Definition 5.10.1. The ring $K[[x]]$ of formal power series consists of all expressions $\sum_{n\ge 0} a_n x^n$ with $a_n \in K$, added coefficientwise and multiplied by the Cauchy convolution $[x^k](fg)=\sum_{i+j=k} a_i b_j$. A series is invertible in $K[[x]]$ exactly when $a_0\neq 0$; a series with $a_0=0$ and $a_1\neq 0$ is invertible in the field of formal Laurent series, with inverse x^{-1} times a power series.

Every coefficient of a product is a finite sum, so no convergence is needed anywhere. You may divide, substitute a series with zero constant term, take formal derivatives and solve algebraic equations, and none of it requires an analytic justification; convergence for small x, when it holds, is a consequence and not a hypothesis. Putnam 2024 A6 depends on exactly this licence, which is what permits forming f^{-1} and substituting inside a quadratic. State it once in your write-up and then use it freely.

The dictionary, which you must know in both directions:

1. shifting: $\sum a_{n+1}x^n=(F(x)-a_0)/x$, and $\sum a_{n-1}x^n=xF(x)$;
2. weighting by the index: $\sum n a_n x^n = xF'(x)$, so the Euler operator $D=x\frac{d}{dx}$ acts diagonally with $D(x^k)=kx^k$ and $P(D)F=\sum P(n)a_n x^n$ for any polynomial P;
3. partial sums: $\sum_n\big(\sum_{k\le n} a_k\big)x^n = F(x)/(1-x)$;
4. convolution: the product FG has coefficients $\sum_k a_k b_{n-k}$;

5. exponential version: with $F(x) = \sum c_n x^n/n!$, the product FG has coefficients $\sum_k \binom{n}{k} c_k d_{n-k}$, the binomial convolution;

6. extraction: $\sum_{n \equiv r \ (m)} a_n = \frac{1}{m} \sum_{j=0}^{m-1} \zeta^{-jr} F(\zeta^j)$ for ζ a primitive mth root of unity.

Item 2 is what turns an opaque coefficient sequence into a differentiation problem in Putnam 2019 A5 and evaluates $\sum nx^n = x/(1-x)^2$ in Putnam 2021 B2; the operator identity $x^k \partial_x^k = D(D-1)\cdots(D-k+1)$ of Putnam 2019 B4 belongs to the same calculus, and under $x = e^u$ the operator D becomes ∂_u, which is why equations whose kth derivative carries the coefficient x^k become constant coefficient in logarithmic variables. Item 6 is the roots of unity filter, and Putnam 2017 A5 uses it to compress three coupled win probabilities into one complex number.

Rational generating functions are exactly linear recurrences

Theorem 5.10.2. *A sequence satisfies a constant coefficient linear recurrence of order d from some index onward if and only if its generating function is $P(x)/Q(x)$ with* $\deg P$ *finite and $Q(x) = 1 - a_{d-1}x - \cdots - a_0 x^d$, the reversed characteristic polynomial; finitely many exceptional initial terms are absorbed into P.*

Proof. Multiply the recurrence $c_{i+d} = \sum_{j<d} a_j c_{i+j}$ by x^{i+d}, sum over the valid i, and recognise each sum as a shift of F; solve the resulting linear equation for F. Conversely expand P/Q by long division and read off the recurrence from $QF = P$. □

Periodicity is the special case $c_{i+m} = c_i$, whose characteristic polynomial $x^m - 1$ gives denominator $1 - x^m$. That is exactly the device of Putnam 2017 B3: an eventually periodic coefficient sequence yields $F(x) = P(x) + x^n Q(x)/(1-x^m)$, an explicit rational function whose denominator can then be analysed. Putnam 2015 B5 uses the same theorem as an independent cross check on a transfer system.

Once the generating function is rational, partial fractions recover the closed form and reproduce Theorem 5.2.3: a factor $(1-\lambda x)^m$ in the denominator contributes $p(n)\lambda^n$ with $\deg p < m$, because $(1-\lambda x)^{-m} = \sum_n \binom{n+m-1}{m-1} \lambda^n x^n$ and the binomial coefficient is a polynomial in n of degree $m-1$.

Convolution recursions and quadratic equations

A recursion containing a self convolution becomes a quadratic equation for the generating function. The archetype is Catalan: $c_0 = 1$, $c_k = \sum_{i=0}^{k-1} c_i c_{k-1-i}$ gives $C = 1 + xC^2$, hence

$$C(x) = \frac{1 - \sqrt{1-4x}}{2x}, \qquad \sqrt{1-4x} = 1 - 2\sum_{m \geq 1} C_{m-1} x^m,$$

so every coefficient of $\sqrt{1-4x}$ after the constant term is negative. Putnam 2022 A2 uses precisely this to show that a forbidden sign pattern is realisable by a genuine square of an infinite series, so the obstruction in that problem must come from finiteness rather than from signs. The formal square root itself is a recursion: for $b_0 > 0$ there is a unique a with $a^2 = b$, $a_0 = \sqrt{b_0}$, computed by

$$a_k = \frac{1}{2a_0}\Big(b_k - \sum_{i=1}^{k-1} a_i a_{k-i}\Big),$$

the general method being to isolate the term carrying the highest unknown coefficient and treat the rest as known history. Putnam 2024 A6 is the same mechanism with a parameter: if $f = x + axf + bf^2$ then $c_1 = 1$ and $c_m = ac_{m-1} + b\sum_{i=1}^{m-1} c_i c_{m-i}$; Catalan is $(a,b) = (0,1)$ and large Schroder is $(1,1)$.

Exponential and probability generating functions

Labelled structures use exponential generating functions because splitting a labelled object into two parts requires choosing which labels go where, which is a binomial convolution. The canonical shape is

$$F_n = \sum_{i=1}^{n} \binom{n-1}{i-1} L_{i-1} R_{n-i},$$

which arises whenever you classify an object by the position of its largest element, and which reads at the level of exponential generating functions as $F' = L \cdot R$. Putnam 2025 A5 uses it twice: to split a permutation at its maximum, and to derive $2E'(x) = 1 + E(x)^2$ for the Euler zigzag numbers, whose solution with $E(0) = 1$ is $\sec x + \tan x$. Putnam 2006 A4 uses it for the insertion recursion of the Eulerian numbers, and Putnam 2013 B5 for a binomial convolution $a_p = \sum_j \binom{p}{j} b_j a_{p-j}$, which is where Lagrange inversion enters when the convolution is nonlinear.

Probability generating functions are the same object with a probabilistic reading. For an integer valued $X \geq 0$, $G(z) = \mathrm{E}[z^X] = \sum p_n z^n$ has nonnegative coefficients, radius of convergence at least 1, $G(0) = \mathrm{P}(X = 0)$, $G(1) = 1$, and kth derivative at 1 equal to the kth factorial moment; nonnegativity of the coefficients forces every derivative to be nonnegative on $[0, 1)$, which is what powers the Taylor estimate in Putnam 2014 A4. Independence multiplies: if $S = \sum a_k X_k$ with the X_k independent and integer valued, then $\mathrm{E}[z^S] = \prod_k \mathrm{E}[z^{a_k X_k}]$, and $\mathrm{P}(S = b)$ is a coefficient of that product. That is the first move of Putnam 2022 B5.

One bound is worth isolating because it is an entire problem.

Proposition 5.10.3 (Putnam 2018 B6)**.** *If $F(x) = \sum a_n x^n$ has $a_n \geq 0$ then for every $x > 0$ in the domain of convergence and every n, $a_n x^n \leq F(x)$, so $a_n \leq x^{-n} F(x)$.*

Applied with $F = P^k$, $P(x) = \sum_{j \in A} x^j$, this bounds the number of length k sequences from A with sum n by $x^{-n} P(x)^k$, and with $A = \{1, \ldots, 6, 10\}$, $k = 2018$, $n = 3860$, $x = 1/2$ it gives exactly the required bound, since $P(1/2) = 1009/1024 = 2018/2048$. The underlying recursion is $a(k, n) = \sum_{j \in A} a(k-1, n-j)$, and every "stage k count is a nonnegative shifted combination of stage $k-1$ counts" recursion solves as $G_k = P^k G_0$ and inherits the bound.

Snake oil, and the two reindexing moves

The snake oil method is this: to evaluate a sum $S(n) = \sum_k f(n, k)$ in closed form, form $\sum_n S(n) x^n$, interchange the two summations, do the inner sum first, and recognise the result.

Example 5.10.4. Evaluate $S(n) = \sum_{k \geq 0} \binom{n+k}{2k}$. Then

$$\sum_{n \geq 0} S(n) x^n = \sum_{k \geq 0} \sum_{j} \binom{j}{2k} x^{j-k} = \sum_{k \geq 0} x^{-k} \cdot \frac{x^{2k}}{(1-x)^{2k+1}} = \frac{1}{1-x} \sum_{k \geq 0} \Big(\frac{x}{(1-x)^2} \Big)^k,$$

using $\sum_j \binom{j}{m} x^j = x^m / (1-x)^{m+1}$. The geometric series sums to $(1-x)^2 / \big((1-x)^2 - x\big)$, so the whole thing is $(1-x)/(1-3x+x^2)$. By Theorem 5.10.2 the coefficients satisfy the recurrence with characteristic polynomial $t^2 - 3t + 1$, whose roots are φ^2 and $\bar{\varphi}^2$; checking two terms gives $S(n) = F_{2n+1}$.

Notice how that example ties three sections together: the reindexing is Putnam style bookkeeping, the denominator is the every second term subsequence transform of Section 2, and the answer is a Fibonacci family member. That is what "revolving concepts" means.

Two reindexing moves recur. *Slicing by level sets*: if the summand factors as $F(k) G(\phi(k, n))$ then

$$\sum_{k,n} F(k) G(\phi(k, n)) = \sum_{N \in \operatorname{im} \phi} G(N) \sum_{(k,n) \in \phi^{-1}(N)} F(k),$$

valid under absolute summability. With $\phi(k,n) = kn$ this is Dirichlet convolution; with $\phi(k,n) = k2^n + 1$ it is the move that reduces the double sum of Putnam 2016 B6 to a one dimensional sum. *Self similar splitting*: if the index set is invariant under a substitution, splitting along the orbits reproduces a scaled copy of the sum and gives a linear equation $S = (\text{explicit finite part}) + \lambda S$, solved by $S = (\text{explicit})/(1-\lambda)$ when $|\lambda| < 1$. Putnam 2016 B6 has a four line alternative solution by exactly that, and the same one step move evaluates geometric sums, dyadic sums and self similar integrals.

Digit recursions and automatic sequences

If $c_{bk+d} = w_d c_k$, that is if the weight of an integer is the product of weights of its base b digits, then

$$\sum_{0 \le k < b^n} c_k x^k = \prod_{j<n} \Big(\sum_d w_d x^{d\, b^j} \Big),$$

a telescoping product, because choosing each digit independently is exactly what a product of factors does. Putnam 2023 A5 uses this with $b = 3$ and $w_0 = w_2 = 1$, $w_1 = -2$, so each factor is $(x^{3^j} - 1)^2$, turning a sum of 3^{1010} terms into a product of 2020 elementary factors and exhibiting a zero of order $2n$ at $x = 1$. Putnam 2020 B1 is the Thue and Morse case $t_{2k} = t_k$, $t_{2k+1} = -t_k$, where $\prod_{i<m}(1 - x^{2^i}) = \sum_{k<2^m} (-1)^{d(k)} x^k$ has a zero of order exactly m at $x = 1$, which is precisely the statement that the first m moments of the sign sequence vanish. The identity $\prod_{n \ge 0}(1 + x^{2^n}) = 1/(1-x)$ used in Putnam 2016 B6 is the same principle read as uniqueness of binary expansion.

The general lesson is that root multiplicity at $x = 1$ encodes moment vanishing, since applying $D = x\frac{d}{dx}$ fewer than m times and setting $x = 1$ gives zero. That is a two way circuit: to prove moments vanish, exhibit the factorisation; to compute a degree drop, count the order of the zero.

- **WHEN** a sequence satisfies a linear recurrence and you want a closed form or an identity **DO** write the generating function as a rational function with denominator the reversed characteristic polynomial and use partial fractions **BECAUSE** multiplying the recurrence by x^i and summing gives one linear equation for F, Theorem 5.10.2
- **WHEN** a recursion contains a convolution of the sequence with itself **DO** form the generating function; the convolution becomes F^2 and the recursion becomes a quadratic equation **BECAUSE** Cauchy multiplication is exactly convolution, so a convolution recursion is a polynomial identity in F
- **WHEN** the objects being counted carry labels and are split into two independent parts **DO** use exponential generating functions, since the binomial convolution becomes an ordinary product **BECAUSE** choosing which labels go to each part is the binomial coefficient, which is precisely the exponential convolution
- **WHEN** you need an upper bound on one coefficient of a product of nonnegative series **DO** evaluate the whole series at a well chosen positive point and divide by that power **BECAUSE** every term of a nonnegative series is at most the sum, so $a_n \le x^{-n} F(x)$, and optimising x is a one variable calculus problem
- **WHEN** a sum over one index resists evaluation **DO** multiply by x^n, sum over the outer index, exchange the order, and do the inner sum first **BECAUSE** the outer generating function is frequently a geometric or binomial series even when the individual sums are not, which is the snake oil method
- **WHEN** a weight is multiplicative over base b digits **DO** write the generating function as a product over digit positions **BECAUSE** independent digit choices correspond exactly to a product of factors, and the order of the zero at $x = 1$ reads off the vanishing moments

- **WHEN** you need only the terms whose index lies in one residue class **DO** average the generating function against roots of unity **BECAUSE** the orthogonality relation for characters of $\mathbb{Z}/m\mathbb{Z}$ kills every other class

5.11 Asymptotics of recursively defined sequences

When there is no closed form you still want the growth rate, and there are five reliable methods.

Dominant root. For a constant coefficient linear recursion the growth rate is the root of largest absolute value, with the other roots contributing lower order terms. Putnam 2012 A1 uses φ; Putnam 2015 B5 uses the supergolden root of $t^3 = t^2 + 1$, whose conjugate pair has modulus below one, so $u_m = C\psi^m + O(1)$. Putnam 2023 A4 uses the same fact in the form $x_{k+1}/x_k \to$ dominant root.

Integrating the increments. For $a_{n+1} = a_n + f(a_n)$ with $f > 0$ decreasing and $a_n \to \infty$, put $F(a) = \int^a dt/f(t)$. Monotonicity squeezes $F(a_{n+1}) - F(a_n)$ between 1 and $f(a_n)/f(a_n + f(a_n))$, which tends to 1, so by Cesaro $F(a_n) \sim n$ and $a_n \sim F^{-1}(n)$. The instances are worth banking: $f(a) = a^{-p}$ gives $a_n \sim ((p+1)n)^{1/(p+1)}$; $f(a) = e^{-a}$ gives $a_n = \log n + o(1)$; the decreasing sibling $a_{n+1} = a_n - ca_n^k$ gives $a_n \sim (c(k-1)n)^{-1/(k-1)}$. Putnam 2006 B6 and Putnam 2012 B4 are the two live instances, and the method also tells you in advance that the error is of size $1/n$, which calibrates every later estimate. This is the discrete twin of Theorem 5.6.5: the substitution $b_n = x_n^{-p}$ there is the special case $F(a) = a^p/p$ here.

Doubly exponential. $m_{j+1} = \lambda m_j$ with $\lambda > 1$ capped by M runs for at most $\log_\lambda(M/m_0)$ steps, which is where logarithms in answers come from; the quadratic sibling $s_{j+1} \le s_j^2$ has solution $s_j \le s_0^{2^j}$, which is the convergence rate of Newton's method and of repeated squaring. Putnam 2008 A6 uses both, the second giving the sharper length $\log_2 n + \log_2 \log n$.

Sum to integral comparison. If f is increasing then $\int_0^M f \le \sum_{n \le M} f(n) \le \int_1^{M+1} f$, so a running total $T(M) = T(M-1) + c(M)$ with $c(n)$ of the size of a power of n has $T(M)$ of the size of the next power divided by the new exponent. Putnam 2016 B2 uses that, and the harmonic case $\ln(n+1) \le H_n \le 1 + \ln n$, which powers Putnam 2008 B2, is the same comparison at exponent -1.

Power law ansatz for self referential inequalities. If $g(n)$ behaves like n^s then $g(g(n))$ behaves like n^{s^2} and $g(n+1) - g(n)$ like sn^{s-1}: composition squares the exponent and differencing subtracts one. So an inequality $g(n+1) - g(n) \ge g(g(n))^r$ balances exactly when $s - 1 \ge rs^2$, and the maximum of $(s-1)/s^2$ over s is $1/4$ at $s = 2$. That locates the answer to Putnam 2025 B6, the extremal function n^2, and the mechanism of the impossibility proof, all within the first minute and before any rigorous argument is written. The exponent map $\varphi(s) = rs^2 + 1$ then does the rigorous work, as described in the bootstrap discussion above.

For implicitly defined sequences, such as the roots of $\tan x = x$ in Putnam 2024 B3, the method is: write $x_n =$ (branch centre) + (correction), expand the defining equation in the correction, and read off the leading asymptotic from the first order term; iterating gives further terms. Alternatively, use an exact addition formula to relate consecutive terms, as the tangent subtraction formula does there, and bound the gap from that exact relation rather than from a series.

- **WHEN** a positive increasing recursion has increments that are a decreasing function of the current value **DO** integrate the reciprocal of that function and invert **BECAUSE** the increments of $F(a_n)$ are asymptotically 1, so $F(a_n) \sim n$ and the asymptotic follows by inversion

- **WHEN** a self referential inequality involves a composition of the unknown function with itself **DO** substitute a power law, turn composition into squaring of the exponent, and solve the resulting exponent inequality **BECAUSE** composition squares exponents and differencing lowers them by one, so the extremal exponent is the solution of a quadratic and the answer is its discriminant threshold

5.12 The worked families

These five families supply the initial data for most recursion problems on the paper. Learn the recursion, the closed form and the two or three identities that follow from the matrix or the generating function; everything else is derivable.

Fibonacci and Lucas. $F_{n+2} = F_{n+1} + F_n$, $F_1 = F_2 = 1$; Binet $F_m = (\alpha^m - \beta^m)/\sqrt{5}$ with $\alpha, \beta = (1 \pm \sqrt{5})/2$ the roots of $z^2 = z + 1$. With $Q = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ one has $Q^n = \left(\begin{smallmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{smallmatrix}\right)$, so multiplicativity $Q^{a+b} = Q^a Q^b$ gives the addition formula $F_{a+b} = F_a F_{b+1} + F_{a-1} F_b$ and hence $F_{2n} = F_n(F_{n+1} + F_{n-1})$, while multiplicativity of the determinant gives Cassini. Putnam 2021 B4 uses both: the doubling identity forces m odd when F_m is prime, and Cassini supplies a square root of -1 modulo F_m. Telescoping identities: $\sum_{j \le m} F_j = F_{m+2} - 1$, the odd indexed sum is F_{2m}, the even indexed sum is $F_{2m+1} - 1$, and $\sum F_j^2 = F_m F_{m+1}$; Putnam 2020 A5 and Putnam 2019 B5 both use these. Also $\varphi^k = F_k \varphi + F_{k-1}$, which gives an explicit density proof for $\mathbb{Z} + \varphi\mathbb{Z}$ in Putnam 2023 A4. The Lucas sequences generalise: with α, β the roots of $x^2 - Px + Q$, $U_n = (\alpha^n - \beta^n)/(\alpha - \beta)$ and $V_n = \alpha^n + \beta^n$; then $V_n \mid V_{kn}$ for odd k, and, when $\gcd(P, Q) = 1$, $\gcd(U_m, U_n) = U_{\gcd(m,n)}$. Putnam 2015 A2 is the V statement.

Catalan. $C_0 = 1$, $C_k = \sum_{i<k} C_i C_{k-1-i}$, $C_n = \binom{2n}{n}/(n+1)$, generating function $(1 - \sqrt{1-4x})/(2x)$. They count ballot words, monotone lattice paths below the diagonal, and standard Young tableaux of two row shape; Putnam 2024 A3 and Putnam 2024 A6 both use them.

Stirling numbers of the first kind. $\prod_{j<k}(x+j) = \sum_i \left[\begin{smallmatrix} k \\ i \end{smallmatrix}\right] x^i$ with $\left[\begin{smallmatrix} k+1 \\ i \end{smallmatrix}\right] = k \left[\begin{smallmatrix} k \\ i \end{smallmatrix}\right] + \left[\begin{smallmatrix} k \\ i-1 \end{smallmatrix}\right]$. The recursion says exactly that multiplying the rising factorial by $(x+k)$ maps the coefficient vector to k times itself plus its own shift. Putnam 2007 B5 expands an annihilating product in this basis; Putnam 2017 A5 uses the probabilistic reading, in which $X(X+1)\cdots(X+n-1)/n!$ is the probability generating function of the number of cycles of a uniform random permutation, with mean H_n.

Bell and involution numbers. Both come from $y' = Qy$ as in Section 3: $Q = e^x$ gives $y = \exp(e^x - 1)$ and the Bell numbers, $Q = 1 + x$ gives $y = \exp(x + x^2/2)$ and the involution numbers. Putnam 2022 B1 is the general statement of that circuit.

Derangements. $D_m = mD_{m-1} + (-1)^m$, equivalently $D_m = (m-1)(D_{m-1} + D_{m-2})$, with $D_m = m! \sum_{i \le m} (-1)^i / i!$, so $D_m / m! \to 1/e$. Putnam 2020 A4 produces a derangement count as the solution of its normalised recursion, and Putnam 2023 A6 uses the fixed point distribution.

Two further families appear once each and are worth recognising rather than memorising: the Eulerian numbers $A(n,k) = (k+1)A(n-1,k) + (n-k)A(n-1,k-1)$ from the insertion recursion of Putnam 2006 A4, and the Euler zigzag numbers with $\sum E_n x^n / n! = \sec x + \tan x$ from Putnam 2025 A5.

- **WHEN** a Fibonacci identity involving products or indices $a + b$ is required **DO** write it as a statement about Q^n and use multiplicativity of the matrix or of its determinant **BECAUSE** matrix multiplicativity is the addition law, so index arithmetic on Fibonacci numbers is matrix arithmetic and needs no induction
- **WHEN** a count satisfies a self convolution recursion with a single removed element **DO** recognise Catalan and use $C = 1 + xC^2$ **BECAUSE** that recursion is the decomposition of a path or a tableau at its first return, which is exactly self convolution

5.13 Recursion as a proof technique

The last use of recursion is not to compute anything. It is to prove that something exists, or that a process ends, or that a position is losing.

Invariants. A quantity preserved by every legal step separates reachable from unreachable states. Putnam 2020 B2 uses the inversion count of a peg vacancy word, which decreases by exactly the

move distance, so the total distance over a whole game is forced. Putnam 2022 A3 uses the rational invariant of the Lyness map. Putnam 2018 B4 uses the Cayley cubic. Putnam 2023 B1 states the general protocol for counting reachable states: find a property preserved by every move for the upper bound, then show every value of it is attained for the lower bound, usually by working backwards from the target along a strictly decreasing potential.

Monovariants. A quantity that is bounded and strictly monotonic at every legal move forces termination, whatever order the moves are made in. Putnam 2008 A3 is the model: a weighted product serves as the monovariant, and integer values plus strict monotonicity plus boundedness gives termination in finitely many steps. When the natural improvement is layered, meaning that a later coordinate improves while earlier ones may worsen, compress the layers into a single scalar by weighting the coordinates with a strictly increasing weight sequence, chosen so that the gain in a later coordinate dominates any simultaneous loss earlier. That converts "prove termination by layers" into one displayed inequality. Putnam 2017 A4 uses the same descent with an integer potential and reads the parity of the potential to decide whether the terminal value is 0 or 1; Putnam 2023 B1 uses the inversion number of a two letter word, which decreases by exactly one per adjacent sorting swap, so the number of steps is exactly the inversion number.

Backward recursion on a game graph. In an impartial game under normal play, n is a $\mathcal{P}$ position (the player to move loses) exactly when every move leads to an $\mathcal{N}$ position, and an $\mathcal{N}$ position exactly when some move leads to a $\mathcal{P}$ position. That is a recursion computed upward from the terminal positions, and it decides every subtraction game once the move set is known; Putnam 2006 A2 is one. The efficient write-up is not to run the recursion but to guess the set $\mathcal{S}$ and verify three fixed point conditions: it contains every terminal position, every move out of it leaves it, and from every position outside it some move enters it. Those three conditions are precisely the fixed point equations of the backward recursion, so verifying them solves all levels at once instead of descending. Putnam 2013 B6 and Putnam 2020 B2 both take that route. The underlying general theory is the Sprague and Grundy function $G(P) = \operatorname{mex}\{G(P') : P \to P'\}$ with $G(P \oplus Q) = G(P) \oplus G(Q)$ for disjunctive sums. For scoring games Grundy theory does not apply, and you instead record the mutual recursion $A(P) = \max_\mu B(P^\mu)$, $B(P) = \min_\mu A(P^\mu)$, which is well founded because every move strictly decreases a nonnegative integer statistic; Putnam 2022 A5 is that case.

Closure as a least fixed point. The set generated by a seed A under partial maps $f_1, \dots, f_k$ is the least fixed point of $X \mapsto A \cup f_1(X) \cup \cdots \cup f_k(X)$, that is the union of the iterates $S_0 = A$, $S_{j+1} = S_j \cup \bigcup_i f_i(S_j)$. There are exactly two proof methods, and knowing that there are two is the point. To prove the closure is everything, find a complexity measure strictly decreased by some composite of the maps and induct downward. To prove it is not everything, exhibit one closed proper superset. Putnam 2009 A4, Putnam 2017 A1, Putnam 2016 A5 and Putnam 2025 B3 are four instances across four different fields. Two refinements worth banking. If a rule set contains an operation together with something close to its inverse, compose them first, because the composite is frequently far simpler than either rule; in Putnam 2017 A1 composing "square after adding c" with "take the integer square root" gives the plain additive step $n \mapsto n + c$, after which the chase is mechanical, and a closure containing an additive step of size c is an up set within each residue class modulo c that it meets, which is a miniature Frobenius computation done one class at a time. And iterating a monotone map F on subsets of an N element set, starting from a set S with $S \subseteq F(S)$, produces an increasing chain that stabilises within N steps, so reachability questions on finite sets are settled in at most N rounds; that is Putnam 2016 A5.

Dynamic programming inside an existence proof. A recursion on counts can prove that some configuration exists without exhibiting it. Putnam 2020 A5 is the model: with an increasing weight sequence and partial sums U_m, the count of index sets summing to n satisfies $r(n) = r(n - u_m) + r(U_{m-1} - n)$, splitting on whether the top index occurs and complementing inside the rest. Base b digits are the special case, and the complementation is the nines complement. Putnam 2014 B1 is the digit peeling version: reduce N to $\lfloor (N - d_0)/b \rfloor$ where d_0 is forced by the residue, turning a k digit claim into a $(k-1)$ digit claim about a smaller integer. Putnam 2025 B4 is the array

version: the peeling recursion $(S, N, n) \mapsto (S - N,\ N - c_1,\ n - 1)$ subtracts one from every nonzero entry and deletes the first row and column.

- **WHEN** a process may be run in any legal order and you must prove it stops **DO** exhibit one integer valued quantity that is bounded and strictly monotonic at every legal move **BECAUSE** boundedness plus strict monotonicity of an integer forces finiteness regardless of the order of moves
- **WHEN** the improvement at each step is layered rather than scalar **DO** weight the coordinates with a strictly increasing sequence and combine into one weighted sum or product **BECAUSE** a sufficiently steep weighting makes a gain in a later coordinate dominate any loss in earlier ones, converting a lexicographic monovariant into a scalar one
- **WHEN** you must decide who wins a combinatorial game **DO** guess the set of losing positions and verify the three closure conditions rather than running the recursion **BECAUSE** those three conditions are the fixed point equations of the backward recursion, so verification settles every level in one step
- **WHEN** a set is defined as the smallest one closed under given rules **DO** prove an explicit candidate is closed for the upper bound, and reach every element of the candidate from the seed for the lower bound **BECAUSE** the generated set is the least fixed point, hence the intersection of all closed sets, so those two proofs pin it exactly
- **WHEN** a rule set contains an operation and something close to its inverse **DO** compose them and work with the composite **BECAUSE** the composite is frequently a plain translation, and a closure under translation by c is an up set in each residue class modulo c
- **WHEN** an existence claim concerns representations or configurations of a given total **DO** write the count as a recursion splitting on the largest available piece and complementing the rest **BECAUSE** a dynamic programme proves existence by proving a count is nonzero, which needs no construction

5.14 The generalisation bank

This is the standing order made explicit. Whenever you meet a recursion, name its shape here and take the solution method off the shelf. Nothing in this list is optional.

1. $c_{m+1} = rc_m$. Solution $r^m c_0$. Sums to $1/(1-r)$.
2. $u_{k+1} = au_k + b$, $a \neq 1$. Subtract the fixed point $b/(1-a)$; $u_k = a^k(u_0 - c) + c$. If $a = 1$, arithmetic.
3. $f(k) = \lambda_k f(k-1) + R(k)$. Divide by $P_k = \prod \lambda_j$ and telescope.
4. $d_N = r(N)d_{N-1}$. Product $d_1 \prod_{i=2}^{N} r(i)$; for $r(N) = -1/(N+c)$ this is a reciprocal factorial and truncated exponential series appear.
5. $u_n = u_{n-1} + q(n)$, q polynomial of degree d. Polynomial of degree $d+1$; Faulhaber or undetermined coefficients.
6. $u_n = \alpha u_{n-1} + \beta^n$. Homogeneous plus $c\beta^n$, or $cn\beta^n$ in the resonant case $\alpha = \beta$.
7. $x_{n+k} = \sum c_i x_{n+k-i}$. Characteristic polynomial; $\sum_i p_i(n)\lambda_i^n$ with $\deg p_i$ below multiplicity.
8. $q(E)x = c\mathbf{1}$. Particular solution $c/q(1)$; if $q(1) = 0$, a polynomial of degree the multiplicity of the root 1.
9. $u_m = \sum a_i u_{m-i} + c$. Homogenise by $u_m = v_m + c/(1 - \sum a_i)$ when $\sum a_i \neq 1$.

10. $x_{k+1} = Mx_k + v$ on a vector space. $x_k = M^k x_0 + \sum_{j<k} M^j v$; telescopes to $(I - M)^{-1}(I - M^k)v$ when $I - M$ is invertible, otherwise expand binomially.
11. $x_n = cx_{n-1}^k$. Logarithms; $\log|x_n| = \log|c|\frac{k^{n-1}-1}{k-1} + k^{n-1}\log|x_1|$, sign absorbed for even k.
12. $c_{k+1} = c_k/(a + bc_k)$. Reciprocals; $w_{k+1} = aw_k + b$.
13. $t_k = t_{k-1}^2 - 2$. Substitute $y + 1/y$ or $2\cos\theta$.
14. $x_{n+1} = 2x_n x_{n-1} - x_{n-2}$. Substitute $\cos\theta_n$; the angles obey $\theta_n = \theta_{n-1} + \theta_{n-2}$.
15. $x_{n+2} = (x_{n+1} + a)/x_n$. Lyness; invariant $(x+1)(y+1)(x+y+a)/(xy)$; period five at $a = 1$.
16. $a_{n+1} = \phi(a_n)$. Fixed points, multipliers, monotone convergence; neutral case via $b_n = x_n^{-p}$.
17. $a_{n+1} = a_n + f(a_n)$, $f > 0$ decreasing. $a_n \sim F^{-1}(n)$ with $F = \int dt/f$.
18. $x_k = f(x_{k-1}, \ldots, x_{k-n}) \bmod N$. Eventually periodic by pigeonhole on N^n windows; purely periodic when the step is invertible.
19. $t_{j+1} = b^{t_j}$. Descend the tower, reducing exponents modulo $\lambda(M)$.
20. $c_N a_N = b_N + \lambda \sum_{m<N} a_m$. Difference against the same identity at $N - 1$.
21. $F(x) = G(x)F(\sigma(x))$, σ contracting. Unfold by escape time.
22. $c_k = \sum_i c_i c_{k-1-i}$ and relatives. Generating function; quadratic equation; Catalan.
23. $F_n = \sum_i \binom{n-1}{i-1} L_{i-1} R_{n-i}$. Exponential generating functions; $F' = LR$.
24. $c_{bk+d} = w_d c_k$. Product over digit positions.
25. $v(n+1) = Mv(n) + b$ from a deletion map. Transfer matrix; characteristic polynomial of M.
26. $G(k) = G(k-1) + t_k$. Telescoping; identify G and evaluate at the endpoints.
27. $S = (\text{explicit}) + \lambda S$ from self similarity. Solve the linear equation.
28. $s_{j+1} = \varphi(s_j)$ on an exponent or a quality parameter. Bootstrap; the answer is the fixed point or the threshold where fixed points appear.

5.15 The rare ones, learn these last

Everything below appeared once in twenty years. Recognise the name, know what it would buy, and do not spend preparation time on the machinery until the rest of this chapter is automatic.

Beatty sequences and Sturmian words, generated by the continued fraction of the parameter through a substitution recursion, together with the three distance theorem: relevant to Putnam 2015 A4, but not needed for the ideal solution there, which uses a gap bound instead. Gaussian binomial coefficients and partial theta truncations, Putnam 2014 B4. J-fractions and the Hankel determinant theorem, with the Motzkin path and Lindstrom, Gessel and Viennot proofs, Putnam 2024 A6; the practical content there is the reduction lemma, that inverting the series rather than expanding it turns a Hankel determinant question into a one step recursion $h_n(\alpha) = \lambda_1^{n-1} h_{n-1}(\alpha_1)$. Pisot numbers, Putnam 2007 B3, where all you need is the check that the conjugate lies in the unit interval. Lagrange inversion, needed only when a nonlinear binomial convolution must be solved in closed form, as in the harder route to Putnam 2013 B5; the statement to remember is that if $w = x\phi(w)$ with $\phi(0) \neq 0$ then $[x^n]H(w) = \frac{1}{n}[w^{n-1}]\big(H'(w)\phi(w)^n\big)$. Gosper's and Zeilberger's algorithms, named above; their value on the paper is the decision rule they justify, not the computation. The Laurent phenomenon for cluster algebra style exchange relations, Putnam 2017 A2, which explains integrality but is not needed to prove it.

One closing instruction. When a recursion appears on the paper and you cannot see which shelf it comes off, do not start computing terms in the hope of spotting a pattern; that is consumption, not analysis. Ask the three questions in order. What is the state, and is the step a linear map of it? If yes, characteristic polynomial. If no, is there a substitution making it linear, and what is the cue? If still no, is there an invariant or a monovariant, and what does it force? Those three questions cover every recursion in twenty years of papers.

Chapter 6

Real analysis

6.1 What this chapter is for

At A-level you were given the word *limit* and never a definition of it. You were told a sequence "tends to" a value, that $\sin x/x \to 1$, that you may differentiate a series term by term, that a continuous function on a closed interval has a maximum. Every one of those is a theorem, and not one of them was proved to you. The Edexcel syllabus is a catalogue of computations performed on objects it never defines. That is the single largest gap between you and the Putnam, and it is why Real analysis is the second most frequent field in the whole archive, behind Number theory and ahead of everything else.

Understand what the gap costs. It is not that you cannot compute; it is that you cannot certify. A Putnam analysis problem rarely asks you to evaluate something beyond A-level reach, it asks you to prove that the evaluation is legitimate. Putnam 2015 B6 evaluates to $\pi^2/16$, and the whole difficulty is that the series is only conditionally convergent, so the rearrangement producing that value is a claim needing an explicit estimate. Putnam 2020 A6 is the same shape reversed: the pointwise limit of f_N is a square wave of supremum $\pi/4$, yet the Fourier partial sums of that square wave overshoot to 0.926, so concluding $M = \pi/4$ from pointwise convergence is worth nothing.

So: every object gets a definition, every interchange of two limiting operations gets a named theorem authorising it or does not happen, and every hypothesis is read as a purchase. The setter paid for "continuously differentiable" rather than "differentiable" because some step needs that continuity, and finding which step is often the whole solution (Putnam 2008 B5).

By the end you will be able to: state and use the completeness axiom in all four of its working forms; write an epsilon and delta argument without hesitation; deploy Rolle and the mean value theorem as zero-counting machinery rather than as facts; produce the right Taylor remainder for the job; test any series and know exactly which manipulations its mode of convergence permits; recognise on sight when a limit may be taken inside a sum or an integral; extract asymptotics with Stolz, Euler and Maclaurin, and Laplace; and read a regularity hypothesis as a specification of which arguments are legal. Every topic ends with reaction circuits, and every circuit is anchored to Putnam problems from 2006 to 2025 that fired it.

6.2 Completeness, and everything you buy with it

The axiom, stated once

Definition 6.2.1. Let $S \subseteq \mathbb{R}$ be nonempty. A real M is an *upper bound* for S if $s \leq M$ for every $s \in S$. The *supremum* $\sup S$ is the least upper bound: an upper bound M such that no $M' < M$ is an upper bound. Infimum is the mirror image.

Theorem 6.2.2 (Completeness axiom)**.** *Every nonempty subset of* $\mathbb{R}$ *that is bounded above has a supremum*

in $\mathbb{R}$.

This is not a theorem you prove; it is the property that distinguishes $\mathbb{R}$ from $\mathbb{Q}$, and it is the only thing $\mathbb{R}$ has that $\mathbb{Q}$ does not. The set $\{q \in \mathbb{Q} : q^2 < 2\}$ is bounded above and has no rational least upper bound. Everything in this section is a consequence of that one axiom, and it is worth knowing that they are all equivalent to it, because then you know that any of them may be used as the entry point.

The working characterisation, the one you actually use in the arena, is the epsilon form: $M = \sup S$ if and only if $s \le M$ for all $s \in S$ and, for every $\varepsilon > 0$, there exists $s \in S$ with $s > M - \varepsilon$. In other words the supremum is approached from inside. That second clause is what you reach for whenever a problem says "the largest constant such that".

Remark 6.2.3. A supremum need not be attained. The wording "find the largest L such that $f(x) \ge L$ for all x" is by definition asking for $\inf f$, not for a minimum value of f. Putnam 2015 A4 punishes the confusion directly: the infimum is $4/7$, achieved only in the limit along the set $S_x = \{1, 4, 7, \dots\}$, and a candidate who evaluates f at a convenient point and answers $5/7$ has answered the wrong question. Infima are attained when the domain is compact and the function is lower semicontinuous, and not otherwise guaranteed.

Theorem 6.2.4 (Archimedean property)**.** *For every real x there is a positive integer n with $n > x$.*

Proof. If not, $\mathbb{N}$ is bounded above, so $M = \sup \mathbb{N}$ exists. Then $M - 1$ is not an upper bound, so some $n \in \mathbb{N}$ has $n > M - 1$, whence $n + 1 > M$ and $n + 1 \in \mathbb{N}$, contradicting that M bounds $\mathbb{N}$. □

Two consequences. First, $\mathbb{Q}$ is dense in $\mathbb{R}$: for $x < y$ pick $n > 1/(y - x)$ and an integer m with $m/n \in (x, y)$, which is how Putnam 2011 B1 chooses integers a, b with b/a in a prescribed interval. Second, the Archimedean contradiction: if $A \le C/k$ for every positive integer k with C fixed, then $A \le 0$. That closes Putnam 2018 A5 in one line, since the argument gives $f(1) \le f'(1)/k$ for every k and $f(1) = 1$, with no estimate of $f'(1)$ needed.

Monotone convergence, and why it is the workhorse

Theorem 6.2.5 (Monotone convergence)**.** *A nondecreasing sequence of reals that is bounded above converges to its supremum. A nonincreasing sequence bounded below converges to its infimum.*

Proof. Let (a_n) be nondecreasing and bounded above, $L = \sup_n a_n$. Given $\varepsilon > 0$ there is N with $a_N > L - \varepsilon$, and monotonicity gives $L - \varepsilon < a_N \le a_n \le L$ for all $n \ge N$. □

This is the workhorse because it produces a limit *without naming it*. You establish monotone and bounded, you own a limit L, and only then do you hunt for its value, usually by passing the recursion to the limit. Putnam 2016 B1 runs on it (strictly decreasing and positive, so convergent, and only afterwards is the limit identified); so does Putnam 2012 B4 (once $a_n - \log n$ is decreasing and positive, convergence is free). Putnam 2011 A2 is the same move dressed as telescoping: the partial sums come out as $\frac{3}{2}(1 - T_m)$ with $0 < T_m < 1$, so increasing and bounded by $\frac{3}{2}$ is visible at a glance and the value $3/2$ follows once $T_m \to 0$.

- **WHEN** a sequence is defined by $a_{n+1} = \Phi(a_n)$ and you are asked whether it converges **DO** prove monotone and bounded before you try to find the limit, then pass the recursion to the limit using continuity of Φ **BECAUSE** monotone convergence hands you the existence of L for free, and $L = \Phi(L)$ then pins it; hunting for the value first is the zombie route and it fails whenever the limit has no closed form
- **WHEN** partial sums come out in closed form $c(1 - T_m)$ with $0 < T_m < 1$ **DO** declare convergence immediately and spend all remaining effort on $\lim T_m$ **BECAUSE** the closed form

already certifies increasing and bounded above by c, so the convergence half of the question is answered with no estimate at all (Putnam 2011 A2)

- **WHEN** $0 < \Phi(x) < x$ on an interval and the orbit starts inside it **DO** conclude the orbit decreases to a fixed point of Φ, and if Φ has no fixed point in the interval, conclude the orbit must leave it **BECAUSE** a decreasing sequence bounded below converges, and continuity forces the limit to satisfy $L = \Phi(L)$; this is how Putnam 2020 A3 gets $a_n \downarrow 0$ and how Putnam 2008 A4 proves the recursive definition of f terminates

Bolzano and Weierstrass, Cauchy sequences, and compactness on the line

Theorem 6.2.6 (Bolzano and Weierstrass)**.** *Every bounded sequence of reals has a convergent subsequence. In $\mathbb{R}^n$, every sequence in a closed bounded box has a subsequence converging to a point of the box.*

Proof. Bisect repeatedly: one half of the containing interval contains infinitely many terms, choose it, choose a term from it with index larger than the last, iterate. The nested intervals have lengths tending to 0 and their common point is the limit by completeness. For $\mathbb{R}^n$ apply the one-dimensional statement to each coordinate in turn, passing to a further subsequence each time. □

The compactness pattern: take a maximising sequence, extract a convergent subsequence, use continuity to show the limit is a maximiser. Putnam 2012 B5 does this on the parameter box $[0, \log M]^2$, converting a supremum into a maximum; the hypothesis that buys it is g_1, g_2 bounded with values in $[1, \infty)$, which is exactly what puts the parameters in a compact box.

Definition 6.2.7. (a_n) is *Cauchy* if for every $\varepsilon > 0$ there is N with $|a_m - a_n| < \varepsilon$ for all $m, n \geq N$.

Theorem 6.2.8. *A sequence of reals converges if and only if it is Cauchy.*

Proof. Convergent implies Cauchy by the triangle inequality. Conversely a Cauchy sequence is bounded, so has a subsequence tending to some L by Bolzano and Weierstrass, and $|a_n - L| \leq |a_n - a_{n_k}| + |a_{n_k} - L|$ forces the whole sequence to L. □

Cauchy is the criterion for convergence *when you have no candidate limit and no monotonicity*; that is its niche. Its usual Putnam appearance is the Cauchy criterion for series: $\sum a_n$ converges if and only if for every $\varepsilon > 0$ there is N with $|a_{m+1} + \cdots + a_n| < \varepsilon$ for all $n > m \geq N$.

Theorem 6.2.9 (Heine and Borel, and its consequences)**.** *A subset of $\mathbb{R}^n$ is compact (every open cover has a finite subcover) if and only if it is closed and bounded. On a compact set, a continuous real function is bounded, attains its supremum and infimum, and is uniformly continuous.*

Treat "closed and bounded" as an announcement that three facilities are now open: a maximum exists, a positive continuous function has a positive minimum, and continuity upgrades to uniform continuity. Putnam 2022 B6 uses all three; Putnam 2013 B4 uses the extreme value theorem solely to write $M(f) = \max_{[0,1]} |f|$ rather than a supremum; Putnam 2018 A3 uses it to speak of an actual maximiser on a box intersected with a hyperplane; Putnam 2011 A5 uses it to manufacture the Lipschitz constant $M = \max |h'|$ that the whole estimate rests on.

Limit superior, limit inferior, and how to negate them

Definition 6.2.10. For a bounded sequence, $\limsup_n a_n = \lim_N \sup_{n \geq N} a_n$, which exists by monotone convergence since the inner suprema are nonincreasing. For a function near a point, $\limsup_{x \to 0^+} h(x) = \lim_{\delta \to 0^+} \sup_{(0,\delta)} h$.

Install it properly because it always exists in $[-\infty, +\infty]$ whereas the limit may not, so it lets you start an argument before you know anything, and because its negation is clean:

$$\limsup_{x\to 0^+} h(x) < \infty \iff \text{there exist } M, \delta_0 > 0 \text{ with } h(x) \le M \text{ on } (0, \delta_0),$$

$$\limsup_{x\to 0^+} h(x) = \infty \iff \text{there is } x_n \to 0^+ \text{ with } h(x_n) \to \infty.$$

Putnam 2019 A6 is built on the first equivalence: it asks you to prove that either $g'(x) \to 0$ or $\limsup_{x\to 0^+} x^r |g''(x)| = \infty$, and the first move is to assume the second fails, whose negation is the usable pointwise inequality $|g''(x)| \le Mx^{-r}$ on a punctured neighbourhood. Install also the arithmetic $\limsup(1+u) = 1 + \limsup u$, and that the limit exists exactly when limsup and liminf agree: Putnam 2022 B6 reduces the whole proof that $f(0^+) = 1$ to $2L_+ = 1 + L_+$ and $2L_- = 1 + L_-$.

- **WHEN** a problem offers you a disjunction, "either P or Q", and Q is a limsup statement **DO** assume Q fails and convert that failure into an explicit pointwise bound on a punctured neighbourhood, then prove P **BECAUSE** the negation of an infinite limsup is a uniform bound with constants, which is the only form of the hypothesis you can compute with (Putnam 2019 A6)
- **WHEN** you need a limit to exist but cannot see monotonicity or a candidate value **DO** work with lim sup and lim inf, which always exist, and prove they are equal **BECAUSE** this converts an existence question into an equality of two well defined numbers, and the arithmetic of limsup under affine maps is enough to close many such equalities (Putnam 2022 B6)
- **WHEN** a sequence lies in a set S with no accumulation point at its limit **DO** conclude the sequence is eventually constant **BECAUSE** no interval around L short enough contains two points of S; for S the powers of 10, two elements differ by a factor of at least 10, which is how Putnam 2007 A4 upgrades an asymptotic statement into an exact one

6.3 Limits and continuity, done with epsilon and delta

The definitions, and the two characterisations of continuity

Definition 6.3.1. $\lim_{x\to c} h(x) = L$ means: for every $\varepsilon > 0$ there exists $\delta > 0$ such that $0 < |x - c| < \delta$ implies $|h(x) - L| < \varepsilon$. The function h is *continuous at* c if $h(c)$ is defined and $\lim_{x\to c} h(x) = h(c)$, that is, for every $\varepsilon > 0$ there is $\delta > 0$ with $|x - c| < \delta \Rightarrow |h(x) - h(c)| < \varepsilon$.

Read the quantifiers as a game and the definition stops being a formula. The adversary names the tolerance ε, you must produce a window δ. Continuity says you can always win with a δ depending on ε and on the point c; uniform continuity says you can win with a δ depending on ε alone. That difference in quantifier order is exactly what makes uniform continuity strong enough to prove continuous functions on $[a, b]$ integrable, and ordinary continuity not.

Theorem 6.3.2 (Sequential characterisation)**.** *h is continuous at c if and only if $h(x_n) \to h(c)$ for every sequence $x_n \to c$.*

This is the form you use nine times out of ten, because it converts a statement about neighbourhoods into one about sequences, and sequences you can construct. The corollary that does the work: if $h(x_n) = v$ for every n and $x_n \to c$, then $h(c) = v$. Putnam 2012 A3 is exactly this and nothing more.

The algebra of limits and the squeeze theorem you have from A-level as facts; what you did not have is that they are theorems you may chain without anxiety. Putnam 2011 B3 is built from nothing else: differentiability at 0 of fg and of f/g gives continuity of both, hence of f up to sign, and the product rule gives $f^2 = (fg)(f/g)$ differentiable at 0. The answer is yes.

Uniform continuity

Definition 6.3.3. h is uniformly continuous on D if for every $\varepsilon > 0$ there is $\delta > 0$ such that $|x - y| < \delta$ with $x, y \in D$ implies $|h(x) - h(y)| < \varepsilon$.

Theorem 6.3.4. *A continuous function on a compact set is uniformly continuous.*

Proof. Suppose not: there is $\varepsilon > 0$ and points x_n, y_n with $|x_n - y_n| < 1/n$ but $|h(x_n) - h(y_n)| \geq \varepsilon$. Pass to a subsequence with $x_{n_k} \to c$; then $y_{n_k} \to c$ too, and continuity at c forces $|h(x_{n_k}) - h(y_{n_k})| \to 0$, a contradiction. □

That proof is a template: negate, extract a convergent subsequence, contradict continuity at the limit point. Uniform continuity is also what makes the double antiderivative of Putnam 2012 A6 differentiable, with $F(x, y) = \iint_{[0,x]\times[0,y]} f$ of class C^1 and $F_{xy} = F_{yx} = f$, the ingredients being uniform continuity on compact sets, the fundamental theorem of calculus and Fubini. A Lipschitz function is uniformly continuous with $\delta = \varepsilon/K$, which is the hypothesis of Putnam 2014 B6.

The intermediate value theorem, and connectedness

Theorem 6.3.5 (Intermediate value theorem)**.** *If h is continuous on $[p, q]$ and v lies between $h(p)$ and $h(q)$, there is $a \in [p, q]$ with $h(a) = v$.*

Proof. Say $h(p) < v < h(q)$. Let $A = \{x \in [p, q] : h(x) \leq v\}$, nonempty and bounded, and let $a = \sup A$. If $h(a) < v$ then continuity keeps $h < v$ slightly to the right of a, contradicting the supremum; if $h(a) > v$ then $h > v$ slightly to the left, again contradicting it. So $h(a) = v$. □

The structural statement behind it is that a continuous image of a connected set is connected, and the connected subsets of $\mathbb{R}$ are exactly the intervals. That is worth carrying because it generates all the variants at once, and the variants are what the Putnam actually uses:

1. *Sign constancy.* A continuous zero free function on an interval has one constant sign. Putnam 2013 A3 converts "P has no root in $(0, 1)$" into the usable "P has one constant sign on $(0, 1)$".
2. *Omitted values.* If g is continuous on an interval, never takes the value v, and takes some value below v, then $g < v$ throughout. Putnam 2022 B6 uses this three times, once in surjectivity form.
3. *Continuous maps into discrete sets.* A continuous $\mathbb{Z}$ valued function on an interval is constant. That plus the next theorem is the mechanism of Putnam 2008 B5: the mean value theorem makes f' integer valued on a dense set, density spreads it to $\mathbb{R}$, connectedness collapses it to one integer, and the answer is $f(x) = \pm x + n$ with $n \in \mathbb{Z}$. That collapse is where the word "continuously" is spent.
4. *Averaging existence.* If J is continuous on a connected parameter space with average v, then J attains v: otherwise $J - v$ is continuous and nowhere zero, hence of one sign, and a function of one sign has nonzero mean. Putnam 2021 B3 turns "the mean of the disc integrals is zero" into "some disc has integral zero" by this alone.
5. *Construction tool.* Build a continuous family straddling the target and the theorem hands you a member hitting it exactly (Putnam 2009 B2).

The counting version of all this is worth its own statement, because it turns "how many solutions" into arithmetic. If g is continuous on $\mathbb{R}$ with finitely many stationary points, split $\mathbb{R}$ into the maximal intervals on which g is strictly monotone. On each, g is injective with image an interval determined by its endpoint limits, so the number of solutions of $g(x) = b$ is the number of those image intervals containing b, counting a shared boundary level once per piece. That is the engine of Putnam 2022 A1, which converts a two parameter intersection count into a comparison of b with the two turning

point values, and of Putnam 2024 B3, which pins each root of $\tan x = x$ to one branch. Beware the neighbouring false move: for a general smooth curve, meeting at exactly one point and meeting tangentially are independent conditions, so a tangency condition describes a curve in the parameter plane whereas the correct answer set in 2022 A1 is a region. The discrete cousin is root counting by sign changes: a degree k polynomial taking nonzero values of alternating sign at $k+1$ consecutive integers has, by the intermediate value theorem, a root in each unit interval and hence exactly k real roots, all located; the nonvanishing must be checked separately, and in Putnam 2024 B5 a two term relation between consecutive values does it.

Theorem 6.3.6 (Agreement on a dense set)**.** *Two continuous functions on an interval that agree on a dense subset agree everywhere.*

Proof. The difference is continuous and vanishes on a dense set; for any c take $x_n \to c$ from the dense set and use sequential continuity. □

Trivial to prove and constantly load bearing, because Putnam problems routinely hypothesise something about the rationals and ask for a conclusion about all reals. Putnam 2014 B6 has every hypothesis about rational r and a conclusion about $[0,1]$; this theorem is the only step that crosses the gap, and omitting it is a genuine hole in the proof, not a formality.

- **WHEN** the hypotheses concern rationals, or a dense set, and the conclusion concerns an interval **DO** say explicitly that both sides are continuous and invoke agreement on a dense set **BECAUSE** a graded proof that never crosses this gap has proved a strictly weaker statement (Putnam 2014 B6, Putnam 2008 B5)
- **WHEN** a continuous function is known never to take a particular value **DO** immediately upgrade to a one sided inequality valid on the whole interval **BECAUSE** connectedness forbids a continuous function from being on both sides of an omitted value; this converts a negative hypothesis into a usable positive one (Putnam 2022 B6, Putnam 2013 A3)
- **WHEN** a continuous function takes only values in $\mathbb{Z}$, or in any set with no accumulation points **DO** conclude it is constant on each interval of its domain **BECAUSE** the continuous image of a connected set is connected, and a connected subset of a discrete set is a point (Putnam 2008 B5)
- **WHEN** you must show some object in a family has a given property, and the family is a continuous connected family **DO** compute the average over the family and show it equals the target value **BECAUSE** a continuous function attains its own mean on a connected domain, which turns an existence problem into a computation (Putnam 2021 B3)
- **WHEN** you are asked how many solutions an equation has, or for the parameters giving exactly one **DO** cut the domain at the stationary points, read off the image interval of each monotone piece, and count how many contain the target level **BECAUSE** a strictly monotone continuous function is injective with interval image, so counting preimages becomes comparing the target with a finite list of turning values (Putnam 2022 A1, Putnam 2024 B3)
- **WHEN** a polynomial of degree k takes values of alternating sign at $k+1$ consecutive points **DO** claim exactly k real roots, one per interval, after separately checking no value is zero **BECAUSE** the intermediate value theorem places at least one root in each of the k intervals and the degree forbids more (Putnam 2024 B5)
- **WHEN** a continuous function is positive at one point **DO** assert it exceeds half that value on a whole subinterval and use the subinterval **BECAUSE** continuity at a point is exactly the statement that nearby values cannot jump; a single point has no measure and contributes nothing to an integral, a subinterval contributes a definite amount (Putnam 2017 A3, Putnam 2025 B2, Putnam 2006 B5)

6.4 Differentiation: Rolle, the mean value theorem, and the family they generate

Fine print you were never given

$h'(c) = \lim_{x\to c}(h(x)-h(c))/(x-c)$, and differentiability at c implies continuity at c. Three pieces of fine print matter on the Putnam.

First, the chain rule needs differentiability of the outer function *at the actual image point*. Putnam 2011 B3 turns on this: $\sqrt{u}$ is differentiable at u_0 only when $u_0 > 0$, which is why the argument runs when $f(0) \neq 0$ and needs separate treatment when $f(0) = 0$.

Second, if f is differentiable everywhere then $g(x) = f(x+1) - f(x)$ is differentiable with $g'(x) = f'(x+1) - f'(x)$, even though f' need not be differentiable. That remark closes the only real gap in Putnam 2010 A2, where the temptation is to differentiate f' directly, which is illegal.

Third, in several variables existence of the partials does not give differentiability: $F(x,y) = xy/(x^2+y^2)$ with $F(0,0) = 0$ has both partials at the origin and is not continuous there. Continuity of the partials is what does give it.

Theorem 6.4.1. *If the partial derivatives of F exist in a neighbourhood of p and are continuous at p, then F is differentiable at p, and the chain rule applies along any differentiable path through p.*

Putnam 2010 A3 and Putnam 2018 B5 both hand you continuous partials in the problem statement, and in both cases that hypothesis exists for exactly this reason: to legalise the chain rule step. Read it that way when you see it.

Rolle and the mean value theorem

Theorem 6.4.2 (Rolle)**.** *If h is continuous on $[a,b]$, differentiable on (a,b) and $h(a) = h(b)$, then $h'(c) = 0$ for some $c \in (a,b)$.*

Proof. By the extreme value theorem h attains a maximum and a minimum on $[a,b]$. If both are attained at endpoints then h is constant and any interior c works. Otherwise an extremum is attained at an interior c, where the one sided difference quotients have opposite signs, forcing $h'(c) = 0$. □

Theorem 6.4.3 (Mean value theorem)**.** *If h is continuous on $[a,b]$ and differentiable on (a,b), there is $\xi \in (a,b)$ with $h'(\xi) = (h(b)-h(a))/(b-a)$.*

Proof. Apply Rolle to $h(x) - h(a) - \frac{h(b)-h(a)}{b-a}(x-a)$. □

Theorem 6.4.4 (Cauchy mean value theorem)**.** *If h, k are continuous on $[a,b]$ and differentiable on (a,b), there is ξ with $(h(b)-h(a))k'(\xi) = (k(b)-k(a))h'(\xi)$.*

Proof. Apply Rolle to $(h(b)-h(a))k(x) - (k(b)-k(a))h(x)$, which takes the same value at a and at b. □

Notice what the hypotheses do and do not require: continuity on the closed interval, differentiability on the open one, and *no continuity of the derivative*. Putnam 2015 B1 gives you "three times differentiable" rather than C^3 precisely because Rolle does not need more, and saying so in one sentence is what makes the write up correct rather than lucky.

The mean value theorem is a secant to derivative bridge: every secant slope is realised as a derivative value strictly inside. Three standard deployments.

Monotonicity from a sign. $F' \geq 0$ on an interval gives F nondecreasing, $F' > 0$ gives strictly increasing, and no continuity of F' is needed. Every bound in Putnam 2009 B5 comes this way, by choosing an auxiliary combination whose derivative has a determinable sign. Putnam 2021 B6 is the same move made explicit: write $g(x) = \frac{3}{2}(\frac{1}{2} - x) - (\frac{1}{2} - \phi(x))$, check $g(\frac{1}{2}) = 0$ and $g'(x) = -\frac{3}{2}(1-2x)^2 \leq 0$, which also displays where the inequality is tight.

Strict monotonicity from a nonstrict bound. If $g' \leq 0$ with equality only on a finite set then g is strictly decreasing, since $g(v) - g(u) = \int_u^v g'$ and a nonpositive integrand not identically zero has strictly negative integral. Putnam 2022 A1 needs this for the boundary case $a = 1$.

Zero derivative gives a constant on each interval, not on each domain. If $g' \equiv 0$ on an interval then g is constant there, which is how Putnam 2010 A2 turns the periodicity $f'(t+1) = f'(t)$ into an outright constant. The caveat is that the conclusion is per connected component: $\arctan t + \arctan(1/t)$ has zero derivative wherever it is defined, but its domain is $(0,\infty) \cup (-\infty, 0)$, so it is $\pi/2$ on one piece and $-\pi/2$ on the other, evaluated at $t = 1$ and $t = -1$. That identity collapses six arctangent terms to the constant $3\pi/2$ in Putnam 2016 A3, and a solution that reports one constant instead of two is wrong.

Growth from a derivative limit. If $F'(x) \to L$ on (a,∞) then $F(x)/x \to L$, and $L > 0$ forces $F \to +\infty$, since beyond some X you have $F' \geq L/2$ and the mean value theorem gives $F(x) \geq F(X) + \frac{L}{2}(x - X)$. This is Step 4 of Putnam 2009 B5.

Rolle as a zero counting machine

This is the highest yield idea in the section, so install it as a machine rather than a theorem.

Lemma 6.4.5. *If h has at least k distinct real zeros then h' has at least $k-1$ distinct real zeros. The zeros of h' lie in the $k-1$ disjoint open intervals between consecutive zeros of h, so they are automatically distinct with no further argument.*

The disjointness remark is not pedantry: it licenses "at least four *distinct* zeros" at each stage rather than "four zeros, possibly repeated". Iterating, n differentiations cost n zeros. In reverse: if $g^{(n)}$ has no zero on an interval then g has at most n zeros there, which is the engine of Putnam 2014 A5, with the memorable corollary that $e^x = P(x)$ with $\deg P = d$ has at most $d+1$ real solutions.

The generalisation is a standing-order recursion: you can arrange the differentiations to compute any real-rooted linear differential operator.

Theorem 6.4.6. *Let $L(t) = \prod_{i=1}^{n}(t - r_i)$ have all roots real, and let $D = d/dx$. Then $L(D)f$ can be produced, up to a nonvanishing exponential factor, by n successive operations each consisting of one multiplication by an exponential followed by one differentiation. Hence if f has $m > n$ distinct real zeros, $L(D)f$ has at least $m - n$ distinct real zeros.*

Proof. For a single root, $\frac{d}{dx}\left(e^{-rx}u\right) = e^{-rx}(u' - ru) = e^{-rx}(D - r)u$. So multiplying by e^{-rx}, differentiating, and discarding the nonvanishing factor realises $D - r$ and costs exactly one zero by Rolle. Compose the n steps. □

Example 6.4.7. Putnam 2015 B1 asks you to show that $f + 6f' + 12f'' + 8f'''$ has at least two distinct real zeros when f is three times differentiable with at least five. Factor the operator: $f + 6f' + 12f'' + 8f''' = 8\left(D + \frac{1}{2}\right)^3 f$. Repeated root, so a single substitution does all three steps: put $g = e^{x/2}f$ and induct on $\frac{d}{dx}(e^{x/2}u) = e^{x/2}(D + \frac{1}{2})u$ to get $g''' = e^{x/2}(D + \frac{1}{2})^3 f$. Now g has five distinct zeros (the exponential never vanishes), so g' has four, g'' has three, g''' has two, and dividing by $e^{x/2}$ returns two distinct zeros of the target. The choice $c = 1/2$ was not inspired: it is the root of the characteristic polynomial, and the integrating factor e^{cx} is the unique solution of $u' = cu$, which is what turns $D + c$ into a bare derivative.

Rolle applied twice in sequence also produces exact identities rather than inequalities, which is a different use of the same machine: three zeros of h give two zeros of h' and then one zero of h'' strictly between them. That is the mechanism behind the interpolation remainder in the next section, and it is why Putnam 2007 B2 becomes an equation from the start.

Darboux, and where continuity of the derivative is really needed

Theorem 6.4.8 (Darboux)**.** *A derivative has the intermediate value property on an interval, even if it is not continuous.*

Proof. Suppose $f'(a) < \lambda < f'(b)$. Set $g(x) = f(x) - \lambda x$, so $g'(a) < 0 < g'(b)$. Then g decreases initially from a and decreases into b from the left, so the minimum of g on $[a, b]$ is attained at an interior point c, where $g'(c) = 0$, that is $f'(c) = \lambda$. □

Darboux is diagnostic: it tells you which uses of "f' is continuous" are removable. In Putnam 2008 B5 it alone collapses a derivative taking only the values 1 and -1 to a constant, so that use is removable, whereas spreading integrality of f' from a dense set to everywhere does need continuity. Naming which is which is the honest reading of the hypothesis.

L'Hopital, stated so that you may actually use it

Theorem 6.4.9 (L'Hopital)**.** *Let $-\infty \le a < b \le \infty$, let h, k be differentiable on (a, b) with $k'(x) \neq 0$ throughout, and suppose $h'(x)/k'(x) \to L \in [-\infty, \infty]$ as $x \to a^+$. If either*

1. *$h(x) \to 0$ and $k(x) \to 0$, or*
2. *$|k(x)| \to \infty$,*

then $h(x)/k(x) \to L$.

Three pieces are routinely got wrong. The nonvanishing of k' throughout is a hypothesis, not a formality (the proof runs through the Cauchy mean value theorem). In case (2) only the denominator need blow up, which is more than the "∞/∞" slogan suggests. And the implication is one way: if h'/k' has no limit, nothing follows.

In practice prefer Taylor, since L'Hopital applied twice is a second order expansion written inefficiently and Taylor also gives the size of the error. Putnam 2021 A2 shows the alternative: $\lim_{t\to 0} f(t)^{1/t} = \exp(f'(0))$ for $f(0) = 1$ holds because $\log f(t)/t$ is literally the difference quotient of $\log f$ at 0.

Vector valued failure, and the substitutes

Theorem 6.4.10. *For a differentiable $\gamma : [p, q] \to \mathbb{R}^n$ with $n \ge 2$ there need be no t with $\gamma'(t) = (\gamma(q) - \gamma(p))/(q - p)$.*

The witness is $\gamma(t) = (\cos t, \sin t)$ on $[0, 2\pi]$: average velocity zero, $|\gamma'| \equiv 1$. The structural reason is the one that generates counterexamples: a continuous real function on a connected set attains its average because its image is an interval, whereas a map into $\mathbb{R}^n$ has connected image whose closed convex hull may be strictly larger, so the average lies only in the hull. Design the image to surround the origin without containing it. Putnam 2009 A6 asks whether the two variable version holds; the answer is disprove, and knowing this failure in advance saves a session spent trying to prove a false statement.

The surviving substitutes are the integral form $\gamma(q) - \gamma(p) = \int_p^q \gamma'$, the mean value inequality $|\gamma(q) - \gamma(p)| \le (q - p) \sup |\gamma'|$, and pairing with a fixed vector v to reduce to the scalar $t \mapsto \langle v, \gamma(t)\rangle$. Putnam 2018 B5 uses the third for exactly this reason.

Injectivity: monotone operators, and local versus global

Theorem 6.4.11 (Inverse function theorem in one variable)**.** *If G is C^k on an interval with $G' \neq 0$ throughout, then G is strictly monotone, maps the interval onto an interval, and G^{-1} is C^k with $(G^{-1})'(x) = 1/G'(G^{-1}(x))$.*

In several variables the theorem is only local, and that is a trap. The map $F(x,y) = (e^x \cos y, e^x \sin y)$ is smooth with $\det DF = e^{2x} > 0$ everywhere yet satisfies $F(x, y+2\pi) = F(x,y)$, so a nonvanishing Jacobian does not give global injectivity; Hadamard's global inverse theorem repairs it only by adding properness. Recognising that route as wrong is what forces the correct one in Putnam 2018 B5, which is monotonicity of the operator.

Theorem 6.4.12. *If $f : \mathbb{R}^n \to \mathbb{R}^n$ satisfies $\langle f(q) - f(p), q - p\rangle > 0$ for all $p \neq q$, then f is injective. For C^1 maps this follows from positive definiteness of $Df(x) + Df(x)^{\mathsf{T}}$ everywhere, via*

$$\langle f(q) - f(p), q - p\rangle = \int_0^1 v^{\mathsf{T}} Df(p + tv)\, v \, dt, \qquad v = q - p.$$

That integral identity is the multivariable substitute for the mean value theorem, and the symmetric part appears because $v^{\mathsf{T}} A v = v^{\mathsf{T}} \frac{1}{2}(A + A^{\mathsf{T}})v$. Putnam 2018 B5 is the two dimensional instance: its determinant condition together with positive diagonal entries is exactly positive definiteness of the symmetrised Jacobian, and naming the theorem tells you the shape of the proof before any computation.

- **WHEN** a function has k known zeros and you must find zeros of a derivative or of a differential expression in it **DO** count with Rolle, one zero lost per differentiation, and factor any constant coefficient operator into real linear factors realised by exponential substitutions **BECAUSE** $L(D) = \prod(D - r_i)$ and each $D - r$ is a bare derivative after multiplying by e^{-rx}, so n Rolle steps cost exactly n zeros (Putnam 2015 B1, Putnam 2014 A5)
- **WHEN** you need a function to be increasing, or need a strict inequality between two expressions **DO** form the difference, differentiate, determine the sign of the derivative, and evaluate at one endpoint **BECAUSE** the mean value theorem turns a sign condition on the derivative into a global order relation, and this reduces an inequality in two variables to a one variable calculus exercise (Putnam 2009 B5, Putnam 2021 B6, Putnam 2025 A2)
- **WHEN** the derivative bound is nonstrict but vanishes only at isolated points **DO** claim strict monotonicity anyway and justify it with $g(v) - g(u) = \int_u^v g'$ **BECAUSE** a nonpositive integrand that is not identically zero on the interval has strictly negative integral, so isolated equality points do not obstruct strictness (Putnam 2022 A1)
- **WHEN** the problem involves a curve in the plane or a map $\mathbb{R}^n \to \mathbb{R}^n$ and you want to apply the mean value theorem **DO** stop, and instead dot with a fixed vector, or use the integral form of the increment **BECAUSE** the mean value equality is false for vector valued functions, with $(\cos t, \sin t)$ the witness; the scalarised version is the only legal one (Putnam 2009 A6, Putnam 2018 B5)
- **WHEN** you must prove a smooth map is one to one and the Jacobian is nonsingular **DO** do not invoke the inverse function theorem; look for a monotone operator estimate $\langle f(q) - f(p), q - p\rangle > 0$ **BECAUSE** the inverse function theorem is local and $e^x(\cos y, \sin y)$ is the standing counterexample to the global claim (Putnam 2018 B5)
- **WHEN** the problem hands you continuous partial derivatives rather than merely partial derivatives **DO** use the chain rule freely, and note in writing that continuity of the partials is what makes the function differentiable **BECAUSE** existence of partials alone does not imply differentiability, so this hypothesis was purchased specifically to license the chain rule (Putnam 2010 A3, Putnam 2018 B5)

6.5 Taylor's theorem, in every remainder form, and which to reach for

The Taylor polynomial of order $n-1$ at a is $T_{n-1}(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!}(x-a)^k$, and $R_n = f - T_{n-1}$. There is not one theorem here but four, and choosing among them is the skill.

Theorem 6.5.1 (Lagrange form). *If f is n times differentiable on an open interval containing a and x, there is ξ strictly between them with*

$$R_n(x) = \frac{f^{(n)}(\xi)}{n!}(x-a)^n.$$

The case $n = 1$ is the mean value theorem.

Theorem 6.5.2 (Integral form). *If $f \in C^n$ on an interval containing a and x, then*

$$R_n(x) = \frac{1}{(n-1)!}\int_a^x (x-t)^{n-1} f^{(n)}(t)\,dt.$$

Theorem 6.5.3 (Peano form). *If $f^{(n-1)}$ exists at a, then $R_n(x) = o((x-a)^{n-1})$ as $x \to a$, with no hypothesis on $f^{(n)}$ at all.*

Choose by what you need. Lagrange gives an exact identity with an unknown point, which you use when you want to *solve* for something: Putnam 2019 A6 links g, $g'(x)$ and g'' on a window by the second order Lagrange form with an unspecified step, and the whole proof is solving that identity for $g'(x)$. The integral form gives a kernel you can weigh against another kernel: Putnam 2018 A5 turns both $f(1)$ and $\int_0^1 f$ into remainder integrals with kernels $(1-t)^{n-1}$ and $(1-t)^n$, and comparing them yields a bound of order $1/n$. Peano is for limits only. Putnam 2006 B6 is a case where the official route uses a quantitative bound on $(1+x)^m - 1 - mx$ and the ideal route needs only the qualitative limit, so the quantitative form is more work than the problem pays for.

A fourth form, which is the one people fail to have installed, is the interpolation remainder. It is Taylor with the derivative data replaced by values at several nodes.

Theorem 6.5.4 (Interpolation remainder). *Let $h \in C^{k+1}[a,b]$ and let P be the polynomial of degree at most k agreeing with h at nodes $x_0 < \cdots < x_k$ in $[a,b]$. For each $x \in [a,b]$ there is $\xi \in (a,b)$ with*

$$h(x) - P(x) = \frac{h^{(k+1)}(\xi)}{(k+1)!}\prod_{i=0}^{k}(x - x_i).$$

Proof. Fix x not a node, choose K so that $\Psi(t) = h(t) - P(t) - K\prod_i (t - x_i)$ vanishes at $t = x$. Then Ψ has $k+2$ distinct zeros, so by the Rolle machine $\Psi^{(k+1)}$ has a zero ξ, and $\Psi^{(k+1)} = h^{(k+1)} - (k+1)!K$. □

Example 6.5.5. Putnam 2007 B2 asks for $\left|\int_0^\alpha f\right| \le \frac{1}{8}\max|f'|$ when $f \in C^1[0,1]$ and $\int_0^1 f = 0$. Set $g(x) = \int_0^x f$, so $g(0) = g(1) = 0$ and $g'' = f'$. The $k = 1$ interpolation remainder with nodes 0 and 1 gives, for each α, a ξ with $g(\alpha) = \frac{g''(\xi)}{2}\alpha(\alpha-1)$, hence $|g(\alpha)| \le \frac{B}{2}\alpha(1-\alpha) \le \frac{B}{8}$ with $B = \max|f'|$. No case split on the sign of $\alpha - \frac{1}{2}$, no symmetry argument. Sharpness: every inequality used was "bound the derivative by its maximum modulus", so an affine function with that constant derivative saturates all of them at once, and $f(x) = \frac{1}{2} - x$ does it.

Two facts stop you over-trusting Taylor. The function equal to e^{-1/x^2} for $x > 0$ and 0 otherwise is C^∞ with every derivative zero at the origin, so infinite order flatness is no contradiction and C^∞ is strictly weaker than analytic; Borel's theorem says every real sequence is the derivative sequence at 0 of some C^∞ function. Putnam 2018 A5 hands you a function all of whose derivatives at 0 vanish, and Borel is what tells you no cheap power series finish exists, so the sign hypotheses must do all the work.

- **WHEN** a function is pinned to zero at two or more points and you need a bound on it in between **DO** use the interpolation remainder, not symmetry and not a case split **BECAUSE** it gives an exact identity $h(x) - P(x) = \frac{h^{(k+1)}(\xi)}{(k+1)!}\prod(x - x_i)$ from which the bound falls out uniformly (Putnam 2007 B2)

- **WHEN** you must compare two integrals or two functionals of the same function **DO** write each as a Taylor remainder in integral form and compare the kernels **BECAUSE** the integral remainder exposes a weight you can dominate pointwise, whereas Lagrange hides everything in an unknown ξ (Putnam 2018 A5)
- **WHEN** a limit is required and you have reached for a quantitative Taylor bound **DO** ask whether the bare limit suffices, and if so drop the bound **BECAUSE** a quantitative estimate costs more ink and invites error; the ideal route uses the weakest tool that closes the gap (Putnam 2006 B6)
- **WHEN** all derivatives of a function vanish at a point **DO** do not conclude the function vanishes near that point **BECAUSE** e^{-1/x^2} is the standing counterexample and Borel's theorem says the phenomenon is generic, so the conclusion must come from elsewhere (Putnam 2018 A5)

6.6 Sequences and series of numbers

The tests, and the one thing they are all really doing

$\sum a_n$ converges means the partial sums converge, and for nonnegative terms they are nondecreasing, so convergence is exactly boundedness of the partial sums. Every test for nonnegative series is a device for bounding partial sums against something known. Know the comparators cold: $\sum n^{-q}$ converges exactly when $q > 1$, $\sum r^n$ exactly when $|r| < 1$, and the logarithmic scale

$$\sum_n \frac{1}{n \ln n \ln\ln n \cdots (\ln^k n)^{\alpha}} \quad \text{converges} \iff \alpha > 1,$$

the Bertrand family. Carry it because it tells you there is no slowest divergent series: Abel and Dini says that if $\sum a_n$ diverges with partial sums s_n then so does $\sum a_n / s_n$. Putnam 2008 A4 defines f by a tower recursion and asks about $\sum 1/f(n)$; the family tells you in seconds that the answer is divergence and that comparison with any fixed member is doomed, since f eventually exceeds all of them.

Theorem 6.6.1 (Integral test, both directions)**.** *If g is positive and nonincreasing on $[1, \infty)$ then*

$$\int_1^{N+1} g \;\le\; \sum_{n=1}^{N} g(n) \;\le\; g(1) + \int_1^{N} g.$$

Proof. On $[n, n+1]$, $g(n+1) \le g \le g(n)$; integrate and sum. □

Carry both inequalities separately, because only the left one transfers divergence of the integral to the sum, and that left inequality is the entire analytic content of Putnam 2008 A4. Check monotonicity on the whole ray first, including the junctions of any piecewise definition: increasing formulas on adjacent intervals give an increasing union only if they agree at each junction.

Ratio and root tests you have. The root test is strictly stronger and decides every case the ratio test decides. Both are comparisons with a geometric series, which is why both are silent at the boundary.

Alternating series, absolute versus conditional, and Riemann's theorem

Theorem 6.6.2 (Alternating series test, with tail bound)**.** *If $b_1 \ge b_2 \ge \cdots \to 0$ then $\sum_{m \ge 1} (-1)^{m-1} b_m$ converges, and for every ℓ,*

$$\Big| \sum_{m \ge \ell} (-1)^{m-1} b_m \Big| \le b_\ell.$$

A finite head of terms breaking monotonicity does not affect convergence.

The tail bound is the useful half and it is what people forget. Putnam 2015 B6 uses it three separate times: to bound $|T_\ell| \le 1/\ell$ so that the rearranged outer series is absolutely convergent, to bound a truncation error, and to prove uniform convergence for a sum and integral interchange (where the Weierstrass M test fails, because $\sum 1/(2\ell-1)$ diverges). Putnam 2017 B4 needs only convergence, of $\sum(-1)^{n-1}(\ln n)/n$, and the point of knowing it converges is that the unknown constant then cancels between the shifted indices; the answer is $(\ln 2)^2$.

Definition 6.6.3. $\sum a_n$ converges *absolutely* if $\sum |a_n| < \infty$, and *conditionally* if it converges but not absolutely.

Theorem 6.6.4 (Riemann rearrangement)**.** *A conditionally convergent real series can be rearranged to converge to any prescribed real value, or to diverge to* $\pm\infty$*.*

This is the most important licensing fact in the chapter, because it says exactly when a manipulation is free and when it is a claim. Absolute convergence buys everything: the unordered sum exists, equals every iterated sum, and equals the sum over any partition into blocks. Conditional convergence buys nothing beyond the given ordering, except that grouping consecutive terms into fixed finite blocks is always safe. In Putnam 2015 B6 the double family is not absolutely summable ($A(k) \ge 1$, so $\sum A(k)/k$ diverges), so the passage from summation ordered by k to summation ordered by ℓ is the load bearing step and needs an explicit estimate on the difference of the two partial sum sequences; the answer $\pi^2/16$ is worth nothing without it. Putnam 2017 B4 organises everything around finite partial sums with one limit at the end for the same reason. Putnam 2016 B6 regroups by the level sets of $k2^n+1$, a partition into blocks rather than an iterated order, and absolute summability is the one line that licenses it; the answer is 1.

Theorem 6.6.5 (Tonelli for counting measure)**.** *For a family* $(a_p)_{p\in P}$ *of* nonnegative *terms indexed by a countable set, the unordered sum* $\sup\{\sum_{p\in F} a_p : F \subseteq P \text{ finite}\}$ *is independent of ordering and grouping, and every iterated sum over a product index set agrees with it, as elements of* $[0,\infty]$*. No convergence hypothesis is needed.*

Nonnegativity is free permission; reach for it before any estimate. Putnam 2015 B4 splits a triple sum by parity class, reparametrises within each class and factors the result into a product with no epsilon anywhere. Putnam 2021 B2 exchanges a double sum before knowing it is finite, and Putnam 2022 A4 exchanges an infinite sum with an expectation, the same theorem for a probability measure.

Summation by parts, Dirichlet and Abel

Theorem 6.6.6 (Summation by parts)**.** *With* $A_n = \sum_{k\le n} a_k$ *and* $A_0 = 0$*,*

$$\sum_{n=m}^{N} a_n b_n = A_N b_N - A_{m-1} b_m - \sum_{n=m}^{N-1} A_n (b_{n+1} - b_n).$$

This is discrete integration by parts and it is the source of both remaining tests. *Dirichlet*: if the A_n are bounded and b_n decreases to 0, then $\sum a_n b_n$ converges. *Abel*: if $\sum a_n$ converges and b_n is monotone and bounded, then $\sum a_n b_n$ converges. Dirichlet is the general reason a periodic mean zero weight against $(\ln n)/n$ converges at all, which underlies Putnam 2017 B4, the bounded partial sums coming from the mean zero periodicity. Memorise also the tail identity $T_\ell - T_{\ell+1} = a_\ell$ for $T_\ell = \sum_{m\ge\ell} a_m$, which telescopes, and its Abel consequence $\sum_\ell c_\ell T_\ell = \sum_m a_m C_m$ with $C_m = \sum_{\ell\le m} c_\ell$.

Two more facts about sums of positive terms, each priority four in the archive. First, geometric self-domination: for $0 < r \le \frac{1}{2}$ and any n, $\sum_{m>n} r^m \le r^n$, with equality exactly at $r = \frac{1}{2}$. Hence subset sums of $\{2^{-n}\}$ are ordered lexicographically by their index sets, and a real in $[0,1]$ has a unique representing set unless it is a dyadic rational; that is what turns Putnam 2015 A4 into a scheduling problem and pins the extremal set $\{1,4,7,\dots\}$ and the value $4/7$. Second, recurrence

from divergence with vanishing steps: an increasing real sequence tending to infinity with increments tending to zero has reductions modulo 2π meeting every arc of positive length infinitely often. This is strictly weaker than equidistribution and needs only a minimality argument, and it is the closing step of Putnam 2017 A5, where the argument of $P_n(\omega)$ is such a sequence and each of three sectors is entered infinitely often.

Telescoping earns its own line as the cheapest identity in analysis: $\sum_{n=1}^{N}(g(n)-g(n+1)) = g(1)-g(N+1)$, so the infinite sum is $g(1)$ whenever $g \to 0$. Generalise the trigger: any term of the form $g(n)-g(n+m)$ sums to a fixed head of m values minus a vanishing tail, and every mean zero weight of period q is an integer combination of such differences.

Infinite products

Theorem 6.6.7. *For $t_n \in [0,1)$, $\prod_n(1-t_n)$ converges to a strictly positive limit if and only if $\sum_n t_n$ converges; otherwise the product tends to 0.*

Proof. Take logarithms and use $t \le -\log(1-t) \le 2t$ on $[0, \frac{1}{2}]$, handling separately the case where t_n does not tend to 0. □

This is the theorem Putnam 2010 A6 disguises: sampling $\int_0^\infty (f(x)-f(x+1))/f(x)\,dx$ at the integers turns it into a product of one step ratios equal to $f(N)/f(0)$, which tends to 0, so the complementary sum must diverge. Putnam 2014 A3 uses the same criterion in the positive direction, to certify that a product converges to a genuine nonzero limit before computing what it is.

- **WHEN** you want to exchange the order of a double sum, or regroup a sum into blocks **DO** first ask whether the terms are nonnegative or the family is absolutely summable; if yes, exchange freely and say why in one line; if no, prove the exchange with an explicit estimate **BECAUSE** Tonelli gives free permission for nonnegative families and Riemann's rearrangement theorem shows there is no permission at all otherwise (Putnam 2015 B4, Putnam 2021 B2, Putnam 2016 B6 versus Putnam 2015 B6, Putnam 2017 B4)

- **WHEN** a series is alternating with terms decreasing to zero **DO** quote convergence and, more importantly, the tail bound $|\sum_{m\ge\ell}| \le b_\ell$ **BECAUSE** the tail bound is a uniform, computable error estimate and often supplies uniformity where the Weierstrass M test fails (Putnam 2015 B6)

- **WHEN** a series has a factor with bounded partial sums multiplied by a monotone null factor **DO** apply Dirichlet's test via summation by parts **BECAUSE** summation by parts moves the difficulty onto $b_{n+1}-b_n$, which is summable by monotonicity, and the bounded partial sums absorb the rest (Putnam 2017 B4)

- **WHEN** a product of factors $1-t_n$ appears, or a ratio u_{n+1}/u_n telescopes into one **DO** convert to $\sum t_n$ by logarithms and decide convergence there **BECAUSE** $t \le -\log(1-t) \le 2t$ on $[0, \frac{1}{2}]$ makes the product and the sum equiconvergent, and products are otherwise hard to bound (Putnam 2010 A6, Putnam 2014 A3)

- **WHEN** a sum has to be compared with an integral, or an integral with a sum **DO** use the two sided integral test and keep both inequalities **BECAUSE** only one of the two directions transfers divergence, and which one depends on whether the summand is increasing or decreasing (Putnam 2008 A4, Putnam 2007 B6, Putnam 2016 B2, Putnam 2024 B6)

6.7 Sequences and series of functions

Pointwise versus uniform, and what each permits

Definition 6.7.1. $f_n \to f$ *pointwise* on E if $f_n(x) \to f(x)$ for each x. $f_n \to f$ *uniformly* if $\sup_{x\in E} |f_n(x) - f(x)| \to 0$, that is, one N works for all x at once.

Pointwise convergence permits essentially nothing: it does not preserve continuity, does not permit termwise integration, does not preserve boundedness of the family, and says nothing about $\sup_n \sup_x f_n$. Putnam 2020 A6 is built on that last failure. Its f_N tends pointwise to a square wave of supremum $\pi/4$, yet the Fourier partial sums of that square wave overshoot to $\frac{1}{2}\mathrm{Si}(\pi) \approx 0.926$ near the jump and keep overshooting as N grows: the Gibbs phenomenon. The correct structural reading is that f_N is the Cesaro mean σ_{2N+1} of the partial sums, that is, convolution with the Fejer kernel

$$K_m(t) = \frac{1}{m+1}\left(\frac{\sin((m+1)t/2)}{\sin(t/2)}\right)^2 \geq 0, \qquad \int K_m = 1,$$

and a convolution with a nonnegative kernel of unit mass is an average of the values of the function, so it cannot exceed $\sup g = \pi/4$, and it approaches $\pi/4$ at points of continuity. The answer $M = \pi/4$ is then earned rather than asserted.

Uniform convergence, by contrast, buys you exactly three things.

Theorem 6.7.2. *Let $f_n \to f$ uniformly on E.*

1. *If each f_n is continuous then f is continuous.*
2. *If $E = [a,b]$ and each f_n is Riemann integrable then so is f, and $\int_a^b f_n \to \int_a^b f$.*
3. *Differentiation is* not *on this list. What is true: if f_n are differentiable on $[a,b]$, $f_n(x_0)$ converges for some x_0, and f_n' converges uniformly, then f_n converges uniformly to some f with $f' = \lim f_n'$.*

Point (3) is the one to get right: uniform convergence of the functions gives nothing about derivatives, you need uniform convergence of the *derivatives*, and $f_n(x) = \sin(nx)/\sqrt{n} \to 0$ uniformly with $f_n'(x) = \sqrt{n}\cos(nx)$ unbounded is the counterexample.

Theorem 6.7.3 (Weierstrass M test)**.** *If $|f_\ell(x)| \leq M_\ell$ for all $x \in E$ and $\sum_\ell M_\ell < \infty$, then $\sum_\ell f_\ell$ converges uniformly and absolutely on E.*

When the M test fails you are not finished: an alternating or Dirichlet type uniform tail bound can still supply uniformity. That is the situation in Putnam 2015 B6, where $\sum 1/(2\ell-1)$ diverges so the M test is unavailable and the alternating tail bound $1/(2L+1)$ licenses the interchange of the outer sum with the integral. Carry also Dini's theorem: continuous f_n on a compact set, monotone in n, with continuous pointwise limit, converge uniformly automatically.

Power series

Theorem 6.7.4 (Cauchy and Hadamard)**.** *For $\sum a_n(x-c)^n$ the radius of convergence is R with $1/R = \limsup_n |a_n|^{1/n}$. The series converges absolutely for $|x-c| < R$, uniformly on every closed subinterval of $(c-R, c+R)$, and diverges for $|x-c| > R$. On the open interval the sum is C^∞ and may be differentiated and integrated term by term, with the same radius.*

The endpoints are separate business, and that is where Putnam problems live. Abel's theorem repairs one direction: if $\sum a_n$ converges then $\sum a_n x^n \to \sum a_n$ as $x \to 1^-$, which is what legitimises evaluating the Maclaurin series of $\log(1+x)$ at $x = 1$ to get $\ln 2$, a value Putnam 2017 B4 needs. One sided derivatives at an endpoint need an Abel type argument before use, and Putnam 2014 A4 is where a generating function route can lose marks by skipping it.

Theorem 6.7.5 (Identity theorem)**.** *If two power series about c with positive radius agree on a set of points having c as an accumulation point, their coefficients are equal. Consequently two real analytic functions on an interval that agree on a set with an accumulation point inside the interval agree throughout it.*

The working consequence is the one to install: an identity between analytic expressions verified on any interval, however short, or at any sequence of points with an accumulation point, holds wherever both sides are defined and analytic. That is how you legitimately extend an identity proved for $|x|$ small to all x, or proved for positive integers to a polynomial identity. Putnam 2017 B3 sits next to this: a power series with coefficients in $\{0,1\}$ has radius at least 1, $f(2/3) = 3/2$ pins the coefficient sequence, and the load bearing lemma is that a real in $[0,1)$ has an eventually periodic base b expansion if and only if it is rational, proved by pigeonhole on the finitely many remainders in long division.

- **WHEN** you want to pass a limit inside an integral or claim the limit of continuous functions is continuous **DO** prove uniform convergence first, by the M test if the terms are dominated by a summable constant sequence, and by an alternating or Dirichlet tail bound if they are not **BECAUSE** pointwise convergence licenses none of this, and the M test is only one sufficient condition, not the definition (Putnam 2015 B6)
- **WHEN** a problem asks for $\sup_N \sup_x f_N(x)$ for a family converging pointwise **DO** refuse to read the answer off the limit function; either locate the maximum of each f_N by differentiating, or bound the family by a nonnegative kernel argument **BECAUSE** Gibbs shows the supremum of the family can strictly exceed the supremum of the limit (Putnam 2020 A6)
- **WHEN** a family is a convolution with a nonnegative kernel of unit mass **DO** conclude at once that its values lie between the infimum and supremum of the function being averaged **BECAUSE** an average cannot leave the range of what it averages, which is why Cesaro means kill the Gibbs overshoot (Putnam 2020 A6)
- **WHEN** an identity has been proved only on a subinterval, or only for small values, and you need it everywhere **DO** check both sides are analytic on a connected domain and invoke the identity theorem **BECAUSE** analytic functions agreeing on a set with an accumulation point agree on the whole connected domain, so no separate argument is needed for the rest of the range
- **WHEN** you want to differentiate a series term by term **DO** verify that the differentiated series converges uniformly, not the original one **BECAUSE** uniform convergence of f_n says nothing about f_n'; $\sin(nx)/\sqrt{n}$ is the standing counterexample

6.8 The Riemann integral, and the interchange of limit and integral

The definition you were never given

Definition 6.8.1. Let f be bounded on $[a,b]$ and let $P : a = t_0 < \cdots < t_N = b$ be a partition. With $m_i = \inf_{[t_{i-1},t_i]} f$ and $M_i = \sup_{[t_{i-1},t_i]} f$, set $L(f,P) = \sum m_i \Delta t_i$ and $U(f,P) = \sum M_i \Delta t_i$. Then f is *Riemann integrable* if $\sup_P L(f,P) = \inf_P U(f,P)$, and the common value is $\int_a^b f$.

Theorem 6.8.2 (Riemann's criterion)**.** *f is integrable if and only if for every $\varepsilon > 0$ there is a partition P with $U(f,P) - L(f,P) < \varepsilon$. Every continuous function on $[a,b]$ is integrable, because uniform continuity makes every $M_i - m_i$ smaller than $\varepsilon/(b-a)$ once the mesh is small enough. More generally (Lebesgue's criterion) a bounded function is Riemann integrable exactly when its set of discontinuities has measure zero.*

Theorem 6.8.3 (Riemann sums as limits)**.** *If f is integrable on $[a,b]$, then for any sequence of tagged partitions with mesh tending to 0 the Riemann sums converge to $\int_a^b f$. In particular $\frac{1}{n}\sum_{k=1}^{n} f(k/n) \to \int_0^1 f$ for continuous f.*

The Putnam punishes careless use of that last statement. Putnam 2019 B2 asks for $\lim a_n/n^3$ where $a_n = \sum_{k=1}^{n-1} h(k/n)$ and h has a pole of order 3 at $t = 1$. The Riemann sum route is not inefficient here, it is invalid, because $\int_0^1 h$ diverges. What is true is that such a sum grows like n^p times a convergent series in the reversed index $j = n - k$, with the exponent read off the order p of the pole. Hence the normalisation n^3 and the answer $8/\pi^3$.

The fundamental theorem has two halves: $F(x) = \int_a^x f$ is differentiable with $F' = f$ for continuous f, and $\int_a^b F' = F(b) - F(a)$ for integrable F'. From the first comes the averaging principle $\frac{1}{h}\int_a^{a+h} f \to f(a)$, whose almost everywhere version for locally integrable f is the Lebesgue differentiation theorem. That is the standard route from averaged data to pointwise data, and recognising when it is *blocked* is strategic: Putnam 2012 A6 fixes the rectangle area at 1, so the region cannot shrink to a point, and noticing that is the first decision of the problem.

Positivity, the workhorse of integral problems

Theorem 6.8.4. *If h is continuous and nonnegative on $[a, b]$ and $\int_a^b h = 0$, then $h \equiv 0$. If $h \geq k$ pointwise with both continuous and $h(t_0) > k(t_0)$ somewhere, then $\int h > \int k$ strictly.*

Proof. If $h(t_0) > 0$ then by continuity $h > \frac{1}{2}h(t_0)$ on a subinterval of positive length λ, so $\int h \geq \frac{\lambda}{2}h(t_0) > 0$. □

Two lines, and on at least four problems in the archive it is the difference between most of the marks and all of them. Putnam 2006 B5 uses it to force $f(x) = x/2$ on the interior, giving uniqueness of the maximiser and the value $1/16$; Putnam 2013 A6 for the upgrade from $A(S) \geq 0$ to $A(S) > 0$; Putnam 2017 A3 twice; Putnam 2024 A5 and Putnam 2025 B2 in the strict version, where the point is explicit that "equality only on the diagonal" is not itself a proof of strictness and you must exhibit an interior point with strictly positive integrand.

The complementary move is lower bounding by domain restriction: for a nonnegative integrand, restricting the domain to any subset gives a valid lower bound, and a subinterval of fixed length on which the integrand is comparable to its natural scale gives an order of magnitude bound with no computation. Putnam 2011 B5 gets $f(y) \geq 1/(6(1 + y^2))$ in three lines from $[-1, 0]$.

Improper integrals and the interchange theorems

For $g \geq 0$ on $[0, \infty)$ the partial integrals increase, so convergence is exactly boundedness of them, and equivalently the tail $\int_a^\infty g \to 0$; a function bounded below by a positive constant on a set of infinite measure has divergent integral. Putnam 2010 A6 runs entirely on the tail: convergence would supply a tail tending to 0, used as an upper bound for $f(a + 1)/f(a)$, and the contradiction is that an integrand eventually at least $\frac{1}{2}$ cannot have convergent integral.

Now the three interchange theorems, at the level the Putnam uses.

Theorem 6.8.5. *Let f_n be measurable on a measure space.*

1. (Uniform convergence.) *If $f_n \to f$ uniformly on a set E of finite measure then $\int_E f_n \to \int_E f$, with the explicit bound $|\int_E f_n - \int_E f| \leq |E| \sup_E |f_n - f|$.*
2. (Monotone convergence.) *If $0 \leq f_1 \leq f_2 \leq \cdots$ pointwise with $f_n \to f$, then $\int f_n \to \int f$ in $[0, \infty]$.*
3. (Dominated convergence.) *If $f_n \to f$ pointwise and $|f_n| \leq G$ for a fixed integrable G, then $\int f_n \to \int f$.*

Prefer (1) whenever you can: it needs no measure theory and produces an explicit error. Putnam 2021 A4 does exactly that, computing the radial integral exactly and showing it differs from its limit by $O(R^{-2})$ uniformly in θ; the limit is $\frac{\sqrt{2}}{2}\pi \log 2$. The diagnostic that saves you there is the borderline tail: if $F(r, \theta) = h(\theta)r^{-2} + O(r^{-3})$ in the plane then $\iint_{\rho \leq |p| \leq \lambda\rho} F \to \log \lambda \int_0^{2\pi} h$, so each half of the

integrand diverges logarithmically and the difference must never be split. In n dimensions the borderline exponent is $-n$.

Theorem 6.8.6 (Tonelli and Fubini)**.** *For a nonnegative measurable integrand the order of integration (or of a sum and an integral, or of two sums) may always be exchanged, with no convergence hypothesis, and the common value lies in $[0,\infty]$ (Tonelli). For a general integrand the exchange is licensed once absolute integrability of the double integral is known (Fubini). For continuous integrands on a compact rectangle both apply.*

Putnam 2020 B3 uses Tonelli to swap $\sum_n \int_0^L$ with $\int_0^L \sum_n$, the step that collapses the exponential series against e^{-t}. Putnam 2021 B3 combines Fubini with invariance: for regions forming a single orbit $K_g = gK$ of a compact group, the occupation weight $M(P) = \int_G \mathbf{1}[P \in gK]\, dg$ is G invariant, so for $G = SO(2)$ it depends only on $|P|$, converting information about circles centred at the origin into information about off centre discs. Putnam 2019 B4 uses Clairaut ($f_{xy} = f_{yx}$ for C^2 functions, the same interchange one level down) to commute the Euler operators.

Theorem 6.8.7 (Differentiation under the integral sign)**.** *If f and $\partial f/\partial x$ are continuous on $[a,b]\times[c,d]$, then $\frac{d}{dx}\int_c^d f(x,t)\,dt = \int_c^d \partial_x f(x,t)\,dt$. With variable limits,*

$$\frac{d}{dx}\int_{u(x)}^{v(x)} f(x,t)\,dt = f(x,v(x))v'(x) - f(x,u(x))u'(x) + \int_{u(x)}^{v(x)} \partial_x f(x,t)\,dt.$$

Over an infinite interval, replace continuity by a fixed integrable dominating function for $\partial_x f$.

- **WHEN** a nonnegative continuous integrand has integral zero, or two integrals of ordered continuous functions are equal **DO** conclude the integrand is identically zero, or that the two functions coincide **BECAUSE** a continuous function positive at one point exceeds half that value on a subinterval of positive length, which contributes a definite positive amount (Putnam 2006 B5, Putnam 2013 A6, Putnam 2017 A3)
- **WHEN** you must prove a strict inequality between integrals **DO** find one interior point where the integrands differ strictly and say the words "by continuity, on a subinterval" **BECAUSE** pointwise nonstrict domination integrates to nonstrict domination only; strictness must be manufactured on a set of positive measure (Putnam 2024 A5, Putnam 2025 B2)
- **WHEN** a limit must pass inside an integral **DO** try uniform convergence with an explicit error first, then monotone convergence if the terms are nonnegative and increasing, and dominated convergence last **BECAUSE** uniform convergence needs no measure theory and gives a quantitative bound, so it is the cheapest legitimate route (Putnam 2021 A4)
- **WHEN** a double sum or a sum against an integral must be swapped and every term is nonnegative **DO** swap immediately and cite Tonelli in one clause **BECAUSE** nonnegativity removes every convergence hypothesis, so this is free permission and spending an epsilon argument on it is wasted ink (Putnam 2020 B3, Putnam 2021 B2, Putnam 2022 A4)
- **WHEN** a sum $\frac{1}{n}\sum f(k/n)$ appears and f blows up at an endpoint **DO** do not write a Riemann sum; reverse the index at the singular end and read the growth exponent off the order of the pole **BECAUSE** the Riemann sum converges to a divergent integral, so the identification is false; the correct scale is n^p for a pole of order p (Putnam 2019 B2)
- **WHEN** a parameter dependent integral must be differentiated in the parameter **DO** check continuity of $\partial_x f$ on a compact rectangle, or produce an integrable dominating function, then differentiate under the sign **BECAUSE** without the domination the interchange is a claim, and it is exactly the step a grader checks

6.9 Asymptotics

Landau notation as a discipline

$f = O(g)$ means $|f| \le C|g|$ on the stated range; $f = o(g)$ means $f/g \to 0$; $f \sim g$ means $f/g \to 1$. State the range, and say what the implied constant depends on. The habit that pays is to measure the tolerance before choosing a method. Putnam 2007 B6 asks for $n^{n^2/2-Cn}e^{-n^2/4} \le f(n) \le n^{n^2/2+Cn}e^{-n^2/4}$ with C free; since $n^{Cn} = e^{Cn\log n}$ that is exactly $|\log f - \log M| \le Cn\log n$, so a quantity built from n coordinates may lose a factor $n^{O(1)}$ per coordinate. Measuring that first licenses crude box bounds and rules out saddle point analysis as disproportionate. The target shape is a hyperfactorial $H(n) = \prod_{j\le n} j^j$ up to $n^{O(n)}$, which tells you the count must come out as a product of factorial ratios.

Growth hierarchy: for all $\alpha, \beta > 0$, $(\ln x)^\alpha = o(x^\beta)$ and $x^\beta = o(e^{\varepsilon x})$. Putnam 2022 A1 needs only the first, to get $\ln(1+x^2) - ax \to -\infty$ and hence the limits at both ends that guarantee at least one intersection. Install also $x^k \ln x \to 0$ as $x \to 0^+$ for $k > 0$, seen by $x = e^{-u}$; Putnam 2008 B2 needs it twice to pin a constant of integration at a singular endpoint.

Sums against integrals, Euler and Maclaurin, Stirling

Theorem 6.9.1. *If g is nonnegative and nondecreasing on $[1, n]$ then $\int_1^n g \le \sum_{i=1}^n g(i) \le \int_1^n g + g(n)$, and the mirror statement holds for nonincreasing g. A unimodal function is handled by splitting at the mode and adding the at most two terms within distance 1 of it.*

With $g = \log x$ this gives $\log n! = n\log n - n + O(\log n)$ and with $g = x\log x$ it gives $\sum_{i\le n} i\log i = \frac{1}{2}n^2\log n - \frac{1}{4}n^2 + O(n\log n)$, which together are the entire analytic content of Putnam 2007 B6. With $g(x) = x^p$, $p > -1$, it gives $\sum_{n\le M} n^p = M^{p+1}/(p+1) + O(M^p)$, which is how Putnam 2016 B2 finds the growth of a counting function and produces $\alpha = 3/4$, $\beta = 4/3$. Putnam 2024 B6 uses the unimodal version to bound a discrete Gaussian sum by $2 + \sqrt{\pi t} \le 4\sqrt{t}$.

The sharpened version is Euler and Maclaurin, first order:

$$\sum_{k=a}^{b} f(k) = \int_a^b f(t)\,dt + \frac{f(a)+f(b)}{2} + \int_a^b \left(\{t\} - \tfrac{1}{2}\right) f'(t)\,dt,$$

and continuing the expansion produces the Bernoulli number terms $\sum_k \frac{B_{2k}}{(2k)!}(f^{(2k-1)}(b) - f^{(2k-1)}(a))$. Two consequences you should simply know: $H_m = \log m + \gamma + \frac{1}{2m} + O(m^{-2})$, with the crude bracket $\log m < H_m \le 1 + \log m$ good enough for most purposes (Putnam 2012 B4, Putnam 2008 B2); and Stirling, $n! \sim \sqrt{2\pi n}\,(n/e)^n$, equivalently $\log n! = n\log n - n + \frac{1}{2}\log n + \frac{1}{2}\log(2\pi) + O(1/n)$.

Cesaro, Stolz, and Abel summation

Theorem 6.9.2 (Stolz and Cesaro)**.** *If (B_n) is positive, strictly increasing and unbounded and $(A_{n+1} - A_n)/(B_{n+1} - B_n) \to L$, then $A_n/B_n \to L$. The case $B_n = n$ is the Cesaro mean lemma: if $x_n \to c$ then the average of $x_1, \dots, x_n$ tends to c, and if $b_{n+1} - b_n \to L$ then $b_n/n \to L$.*

This is the discrete L'Hopital, and the only general bridge from increments to size. Putnam 2006 B6 is nothing else: the increments of $b_n = a_n^{(k+1)/k}$ tend to $(k+1)/k$, so $b_n/n \to (k+1)/k$ and $a_n^{k+1}/n^k \to ((k+1)/k)^k$. Putnam 2020 A3 converts $b_{n+1} - b_n \to 1/3$ for $b_n = a_n^{-2}$ into $a_n^2 \sim 3/n$, whence $\sum a_n^2$ diverges by comparison with the harmonic series.

Abel summation in analytic form is the pairing of a series with its partial sums above; in limit form it is Abel's theorem, $\lim_{x\to 1^-} \sum a_n x^n = \sum a_n$, which is how you evaluate a conditionally convergent series by summing a power series inside the disc first.

Laplace's method

Theorem 6.9.3. *If a sum of positive terms has summand $h(n)e^{\phi(n)}$ with ϕ concave and maximised at n_0, and $\phi''(n_0) \approx -1/\sigma^2$ with $\sigma \gg 1$, then the sum is of order $h(n_0)e^{\phi(n_0)}\sigma\sqrt{2\pi}$. The rigorous version is two sided: lower bound by restricting to a window of width σ about n_0, upper bound by splitting off the far tail (exponentially small, times polynomially many terms) and comparing the bulk with a Gaussian integral. The integral version is $\int h(t)e^{-\lambda\varphi(t)}dt \sim h(t_0)e^{-\lambda\varphi(t_0)}\sqrt{2\pi/(\lambda\varphi''(t_0))}$.*

Putnam 2024 B6 is a pure Laplace problem: for $F_a(x) = \sum n^a e^{2n} x^{n^2}$ the peak is at $n = t = 1/(1-x)$ with height e^t, width $\sqrt{t}$ and weight t^a, so the normalised sum has order $t^{a+1/2}$ and the threshold is $c = -1/2$. Read the statement type: "the limit is 0 for $a < c$ and ∞ for $a > c$" means the quantity is t^{a-c} times something bounded above and below by positive constants, proved by two inequalities each sharp in the exponent and crude in the constant. Do not compute the constant.

Worth a line alongside it: the two scale argument near a singularity. Work on a window of symbolic length L, obtain a bound $A(x)/L^\alpha + B(x)L^\beta$, optimise over L, then verify the optimal L lies in the admissible range. That verification is where the structural hypothesis is consumed: in Putnam 2019 A6 the hypothesis $r > 1$ is used exactly once, to certify that the optimal window of length proportional to x^r fits inside $[x, 2x]$.

- **WHEN** you know how a sequence increments and want to know how large it is **DO** apply Stolz and Cesaro with the right B_n, usually $B_n = n$ **BECAUSE** it is the discrete L'Hopital and the only general bridge from differences to magnitudes (Putnam 2006 B6, Putnam 2020 A3)
- **WHEN** a problem states a two sided bound with a free constant in an exponent **DO** take logarithms, read off exactly how much slack you have per coordinate, and choose the crudest method that fits inside it **BECAUSE** measuring the tolerance before choosing the method is what rules out disproportionate machinery (Putnam 2007 B6)
- **WHEN** a sum of positive terms has a sharp interior peak **DO** apply Laplace's method: locate the peak, compute the width from the second derivative, and prove matching bounds by a window below and a Gaussian comparison above **BECAUSE** the sum is dominated by a window of width σ around the peak, so the answer is height times width up to constants (Putnam 2024 B6)
- **WHEN** a discrete sum must be replaced by an integral **DO** use the two sided monotone comparison, splitting at the mode if the summand is unimodal **BECAUSE** the error is at most one boundary term per monotone piece, which is exactly the accuracy most Putnam problems need (Putnam 2007 B6, Putnam 2016 B2, Putnam 2024 B6)

6.10 Convexity, Jensen, and the standard inequalities in analytic form

Definition 6.10.1. F is *convex* on an interval if $F(\lambda u + (1-\lambda)v) \le \lambda F(u) + (1-\lambda)F(v)$ for all u, v in the interval and $\lambda \in [0,1]$.

The definition is the chord lying above the graph, and every property falls out of it directly.

Theorem 6.10.2 (Chord slope monotonicity)**.** *If F is convex and $u < v < w$ then*

$$\frac{F(v)-F(u)}{v-u} \le \frac{F(w)-F(u)}{w-u} \le \frac{F(w)-F(v)}{w-v}.$$

Consequently one sided derivatives exist everywhere with $F'_-(t) \le F'_+(t)$, and F is continuous on the interior.

Theorem 6.10.3 (Supporting line)**.** *A finite convex $F : \mathbb{R} \to \mathbb{R}$ admits, at every t, a slope m_t with $F(x) \ge F(t) + m_t(x-t)$ for all x. Hence F is the pointwise supremum of its affine minorants, and at each point the supremum is attained by the supporting line there.*

Proof. Chord slope monotonicity makes $m_t = F'_+(t)$ finite and squeezed between the difference quotients on both sides, which is exactly the stated inequality. □

Putnam 2012 B5 is this theorem and nothing else: it asks you to write $\sup_s(g_1(s)^x g_2(s))$ as $\max_t(xh_1(t) + h_2(t))$, and the supporting line lemma supplies $h_1(t) = m_t$ and $h_2(t) = F(t) - tm_t$, with attainment at $t = x$ making the right hand side a maximum rather than a supremum. Finiteness of F comes from the hypotheses "bounded" and "codomain $[1, \infty)$", which put the exponents in a bounded box and give $\sup_s e^{\alpha_s x + \beta_s} \le M^{|x|+1}$; without them the supremum is infinite for one sign of x and no convexity argument starts.

The structural name is the Legendre and Fenchel transform $F^*(y) = \sup_x(xy - F(x))$, with $F^{**} = F$ for closed proper convex F and $xy = F(x) + F^*(y)$ characterising supporting slopes; the pair above is $h_2(t) = -F^*(h_1(t))$, so what is proved is biconjugation for that family. Example: $F(x) = e^x$ gives $F^*(y) = y \log y - y$ on $y > 0$.

Criteria, one line each. F twice differentiable is convex if and only if $F'' \ge 0$; F differentiable is convex if and only if F' is nondecreasing, and then the tangent line bound $F(x) \ge F(x_0) + F'(x_0)(x - x_0)$ holds, which is how Putnam 2010 B5 turns the single value $f''(0)$ into a floor on a whole ray. Midpoint convexity plus continuity, or plus boundedness on an interval, implies convexity; without a regularity hypothesis it does not.

Theorem 6.10.4 (Jensen)**.** *For φ convex and μ a probability measure, $\varphi\big(\int f\,d\mu\big) \le \int \varphi(f)\,d\mu$. The finite form with weights λ_i summing to 1 is the special case of a discrete μ.*

Jensen with $\varphi = -\log$ gives AM-GM in any weighting; with $\varphi(t) = t^2$ it gives $\int (h - c)^2 = \mathrm{Var}(h) + (\mu(h) - c)^2$, so $\mathrm{Var}(h) = \min_c \int (h - c)^2$ with the minimum at $c = \mu(h)$. That identity is the whole of Putnam 2013 B4: instead of computing $\mu(fg)$ you take the free choice $c = \mu(f)\mu(g)$, collapsing an eight term expansion to four lines. Have the rest installed in integral form: Cauchy and Schwarz, Holder $\int |fg| \le \|f\|_p \|g\|_q$ with $1/p + 1/q = 1$, Minkowski, and the power mean inequality. Their finite versions extend to infinite sums by proving them for each partial sum and passing to the limit once both sides are known finite, the step Putnam 2010 B1 asks you to justify rather than assert.

When the class of competitors is closed under scalar multiplication and both sides of the inequality are homogeneous of the same degree, the smallest admissible constant is the supremum of the ratio, so you may normalise either side to 1; and on a finite dimensional space all norms are equivalent, so the smallest constant is finite before any computation. Putnam 2016 A6 uses this twice: normalising $\max_{[0,1]} |P| = 1$ turns the problem into maximising a linear functional over a bounded convex set, and affine invariance of both sides lets you cut $[0, 1]$ at the roots of P and upgrade "a root somewhere" to "a root at an endpoint of each piece", which is where the bound is sharp. The answer is $5/6$.

A weighted inner product often exposes an extremal problem at once. Putnam 2006 B5 maximises $I(f) - J(f) = \langle f, g\rangle - \|f\|^2$ with $\langle u, v\rangle = \int_0^1 x\,uv\,dx$ and $g(x) = x$; completing the square gives $\|g\|^2/4 - \|f - g/2\|^2$, so the maximum is $\|g\|^2/4 = 1/16$, attained only at $f = g/2$. A linear minus a quadratic is a downward parabola in a Hilbert space, and the observation generalises the problem to arbitrary weights for free.

- **WHEN** a supremum of exponentials, or of any family of functions, must be written as a maximum of affine functions **DO** prove the supremum is convex and finite, then use the supporting line theorem **BECAUSE** a finite convex function is the upper envelope of its affine minorants, with attainment at the point of tangency (Putnam 2012 B5)
- **WHEN** an expression is linear minus quadratic in the unknown function **DO** complete the square in the relevant inner product **BECAUSE** the maximum and the unique maximiser both fall out with no calculus of variations, and the answer generalises to every weight at once (Putnam 2006 B5)

- **WHEN** a variance, or a mean square deviation, appears **DO** replace μ by the most convenient constant c and use $\int (h-c)^2 = \mathrm{Var}(h) + (\mu(h)-c)^2 \geq \mathrm{Var}(h)$ **BECAUSE** variance is the squared distance to the constants, so any constant gives an upper bound for it and you may choose the one that factorises (Putnam 2013 B4)
- **WHEN** an inequality has a free best constant and both sides scale the same way under multiplying the competitor by a scalar, or under an affine change of variable **DO** normalise one side to 1, and if the inequality is affine invariant, partition the interval and apply the normalised form on each piece **BECAUSE** homogeneity makes the best constant a supremum of ratios, and affine invariance transports the sharp configuration to every subinterval (Putnam 2016 A6)
- **WHEN** a finite inequality must be applied to an infinite sum or an integral **DO** prove it for every partial sum and pass to the limit, having first established that both sides converge **BECAUSE** the extension is a theorem, not a notation change, and graders mark the omission (Putnam 2010 B1)

6.11 Functional equations of continuous functions

Theorem 6.11.1 (Cauchy's equation). *If $f : \mathbb{R} \to \mathbb{R}$ satisfies $f(x+y) = f(x) + f(y)$ and is additionally continuous at one point, or monotone on some interval, or bounded above on a set of positive measure, or measurable, then $f(x) = cx$ with $c = f(1)$. Without any such hypothesis there exist non linear solutions, constructed from a Hamel basis of $\mathbb{R}$ over $\mathbb{Q}$ using the axiom of choice.*

Algebra alone gives $f(q) = cq$ on $\mathbb{Q}$; the regularity hypothesis is precisely what transports that to $\mathbb{R}$, and the transport mechanism is agreement on a dense set. The relatives are the same equation conjugated by exp or log, and you should recognise them instantly: $f(x+y) = f(x)f(y)$ gives e^{cx} or 0; $f(xy) = f(x) + f(y)$ on $x, y > 0$ gives $c \log x$; $f(xy) = f(x)f(y)$ gives x^c; Jensen's equation gives affine f. Behind all four: a strictly monotone bijection ϕ conjugates addition to $x * y = \phi^{-1}(\phi(x) + \phi(y))$. Putnam 2012 B1 uses that as an engine rather than a classification, since the class S contains $e^x - 1$ and $\ln(x+1)$ and can therefore convert addition into multiplication; the residual correction, that exponentiating a sum of logs gives $fg + f + g$ rather than fg, is exactly what condition (iii) exists to remove.

More important than the classification is knowing when regularity is load bearing. Given no continuity, boundedness or measurability, any route needing a limit, an integral or an average over infinitely many instances is illegitimate, and the intended route is a finite chain of substitutions. Putnam 2009 A1 is that trap (the average over all squares centred at P is illegal) and Putnam 2008 A1 is its twin. When continuity *is* given, spend it deliberately: Putnam 2022 B6 has solutions $f(x) = 1/(1+cx)$ for $c \geq 0$, and its proof runs on limsup arithmetic, omitted values, and the rule that an identity passes to a limit only when every term converges, so each argument must first be multiplied by the factor its asymptotic dictates.

One further piece of vocabulary. The set of periods of a function is an additive subgroup of $\mathbb{R}$, and a subgroup of $\mathbb{R}$ is either cyclic or dense. Hence a continuous function with arbitrarily small periods is constant. That is the closing move of Putnam 2012 A6, where f ends up depending on one variable and having every positive real as a period.

- **WHEN** a functional equation is stated with no regularity hypothesis at all **DO** refuse every limiting argument and hunt for a finite chain of substitutions **BECAUSE** without regularity the solution set contains Hamel basis pathologies, so any proof that uses a limit is proving something false or unprovable (Putnam 2009 A1, Putnam 2008 A1)
- **WHEN** a functional equation mixes addition and multiplication **DO** conjugate by exp or log to make both operations the same, and account for any shift the generators carry **BECAUSE** a

monotone bijection conjugates addition to $\phi^{-1}(\phi(x)+\phi(y))$, and a shifted generator leaves a residual additive term you must cancel (Putnam 2012 B1)

- **WHEN** a continuous function has arbitrarily small periods **DO** declare it constant **BECAUSE** the period set is a subgroup of $\mathbb{R}$, and a subgroup is cyclic or dense; continuity plus density forces constancy (Putnam 2012 A6)

6.12 Differential relations, read as analysis

The archive is full of problems that look like differential equations and are really inequalities. Four tools cover almost all of them.

Integrating factors turn inequalities into monotonicity. If $u' + au \geq 0$ on $[0, X]$ then $(e^{ax}u)' \geq 0$, so $u(x) \geq u(0)e^{-ax}$, with no division by u and no sign assumption on it. Putnam 2023 A3 applies this to $k = f^2 + g^2$ with $a = 2$ to get $k(x) \geq k(0)e^{-2x} > 0$, which is what makes g/f blow up at the first zero of f. The same problem then uses the polar decomposition $|w'|^2 = (\rho')^2 + \rho^2(\theta')^2$ with $\theta' = (fg' - gf')/(f^2+g^2)$: conditions (c) and (d) added in squares give $|w'| \leq |w|$, hence $|\theta'| \leq 1$, so the curve turns at most one radian per unit time while the problem demands a quarter turn, and the answer is $r = \pi/2$. That is Sturm comparison in Prufer angle form, and recognising it makes the angle route obvious rather than inspired.

Superlinearity forces blow up in finite time. If $y > 0$ with $y' \geq \alpha y^2$ then $(-1/y)' \geq \alpha$, so $x - T < 1/(\alpha y(T))$ for every admissible x and no such y exists on an unbounded interval; Osgood's criterion is that $y' \geq g(y)$ with $g > 0$ and $\int^{\infty} dy/g(y) < \infty$ forces blow up. Putnam 2010 B5 asks for a strictly increasing f with $f' = f \circ f$, and this is the contradiction that makes the answer no.

Regularity bootstraps are free. If $f' = \Phi(f)$ with Φ smooth then $f \in C^k$ gives $\Phi \circ f \in C^k$, so $f' \in C^k$ and $f \in C^{k+1}$, hence $f \in C^{\infty}$. Putnam 2010 B5 needs $f''(x) = f'(f(x))f'(x)$ to exist at all and this supplies it; Putnam 2009 B5 uses the same upgrade. Picard and Lindelof (locally Lipschitz right hand side gives local existence and uniqueness) is the theorem behind "guess and check the initial condition"; know it exists, and prefer a direct construction.

Crude bounds bootstrap into sharp ones. Given $y' = F(x, y)$ and a crude $|y| \leq C_0 x^{\alpha_0}$, substituting bounds y' and yields $|y| \leq C_1 x^{\alpha_1}$, and iterating $\alpha_{k+1} = T(\alpha_k)$ converges to the attracting fixed point, the true exponent. In Putnam 2009 B5, $\alpha_0 = 1$, $T(1) = 1/3 = T(1/3)$, so one iteration gives $f = O(x^{1/3})$, which fed back into $g' = 1 - f^2/x^2$ gives $g' \to 1$, hence $g \to \infty$, and strict monotonicity transfers that to f.

- **WHEN** a differential inequality of the form $u' + au \geq 0$ or $u' \geq \alpha u^{1+\varepsilon}$ appears **DO** multiply by the integrating factor, or divide by the power and integrate, and never divide by u itself **BECAUSE** the integrating factor converts the inequality into monotonicity of an explicit function, which needs no sign assumption on u (Putnam 2023 A3, Putnam 2010 B5)

- **WHEN** two coupled first order inequalities bound each function by the other **DO** square and add to get a bound on the speed by the radius, then work with the polar angle **BECAUSE** $|w'|^2 = (\rho')^2 + \rho^2(\theta')^2$ turns a coupled system into a single bound on the turning rate (Putnam 2023 A3)

- **WHEN** you have a crude growth bound and need a sharp one **DO** feed the crude bound back into the equation and iterate the exponent map **BECAUSE** the true exponent is the attracting fixed point of the induced map, and iterating on the constants instead produces the full asymptotic expansion (Putnam 2009 B5)

6.13 Metric and topological vocabulary at Putnam level

You need six words and one theorem. A set is *open* if every point has a ball inside it, *closed* if its complement is open (equivalently, it contains all its limits of sequences), *dense* in X if its closure is X, *compact* if every open cover has a finite subcover (in $\mathbb{R}^n$: closed and bounded), *connected* if it is not the union of two disjoint nonempty relatively open sets (in $\mathbb{R}$: an interval), and *nowhere dense* if its closure has empty interior. The theorem is Baire: a complete metric space is not a countable union of nowhere dense sets.

Collect what each licenses. Closed plus discrete plus continuous domain gives constancy (Putnam 2008 B5). Closed and bounded gives attained extrema, a positive minimum for a positive continuous function, uniform continuity, and Bolzano and Weierstrass (Putnam 2013 B4, Putnam 2018 A3, Putnam 2022 B6, Putnam 2012 B5). Connected gives the intermediate value theorem and its variants, including attainment of the mean (Putnam 2021 B3). Dense gives transport of identities under continuity (Putnam 2014 B6). A closed nonempty subset of a compact interval has a least element, which is how Putnam 2023 A3 passes to the first zero of f so that $\arctan(g/f)$ has one differentiable branch and no lifting theory is needed. Baire appears rarely, and when it does it is in the form "a countable union of closed sets covering an interval has one with nonempty interior".

One caution about measure. Positive Lebesgue measure is a strong hypothesis and its usual payoff is the Steinhaus difference theorem: if $A \subseteq \mathbb{R}$ has positive measure then $A - A$ contains an interval about 0. It does not apply to an arbitrary two colouring of $\mathbb{R}$, whose colour classes need not be measurable, and Putnam 2022 B3 is designed so that this tempting tool is unavailable. Naming a wrong tool and saying exactly why it fails is part of the analysis, not a detour.

6.14 The recursion atlas for analysis

Recursion is a standing order, so here are the analytic recursions from the archive with their general form, general solution and recognition cue. Every one of them appeared with priority 4 or 5.

1. Telescoping, discrete and continuous. $\sum_{n=a}^{b-1}(u_n - u_{n+1}) = u_a - u_b$, and its continuous parent: for $\Delta_h f(x) = f(x) - f(x+h)$ with $f \geq 0$ decreasing to 0, $\int_a^\infty \Delta_h f = \int_a^{a+h} f$, proved by shifting the second copy and noting the boundary term is at most $f(T) \to 0$. That identity is the pivot of Putnam 2010 A6: it evaluates the numerator in closed form, which is what lets the ratio $f(a+1)/f(a)$ be trapped underneath the tail.

2. Running totals. $T(M) = T(M-1) + c(M)$ with $c(n)$ growing like n^p gives $T(M) \sim M^{p+1}/(p+1)$, by sum to integral comparison. Recognise it whenever a counting function is defined by an increment (Putnam 2016 B2).

3. Neutral fixed point iteration. For $x_{n+1} = x_n - cx_n^{p+1} + O(x_n^{p+2})$ with $c > 0$, substitute $b_n = x_n^{-p}$ to get $b_{n+1} = b_n + cp + o(1)$, hence $b_n = cpn + o(n)$ and $x_n \sim (cpn)^{-1/p}$. The method in general: find the power of x whose increments are asymptotically constant. For $f = \sin$, $c = 1/6$, $p = 2$, $x_n \sim \sqrt{3/n}$, and the refinement $b_n = cpn + K\log n + O(1)$ gives $a_n^2 = 3/n - \frac{27}{5}\log n/n^2 + O(n^{-2})$ (Putnam 2020 A3).

4. Additive into multiplicative. $a_{n+1} = a_n + f(a_n)$ becomes $a_{n+1} = a_n(1 + f(a_n)/a_n)$, and only in that form can the recursion be raised to a power termwise. This is what turns $a_{n+1} = a_n + a_n^{-1/k}$ into $b_{n+1} = b_n(1 + 1/b_n)^{(k+1)/k}$ with $b = a^{(k+1)/k}$, a recursion closed in b alone (Putnam 2006 B6).

5. Exponent bootstrap. $\alpha_{k+1} = T(\alpha_k)$ read off the equation, the true exponent being the attracting fixed point (Putnam 2009 B5). Its discrete cousin is the "eventually" bootstrap of Putnam 2025 B6: if g is strictly increasing on $\mathbb{N}$ with $g(n) \geq n^s$ eventually and $g(m+1) - g(m) \geq g(g(m))^r$, then $g(g(m)) \geq m^{s^2}$, the increments are at least m^{rs^2}, and telescoping over $m \in [n/2, n)$ gives $g(n) \geq n^t$ eventually for every $t < rs^2 + 1$. Iterating $\phi(s) = rs^2 + 1$ from 0 escapes to infinity exactly when $\phi(s) - s \geq \delta > 0$ for all s, that is when the discriminant of $rs^2 - s + 1$ is negative, that is when $r > 1/4$; the answer is $r = 1/4$, realised by $g(n) = n^2$. Writing the lemma with "for all sufficiently

large n" on both sides is what makes the iteration writable in the arena, because you never track constants.

6. Regularity bootstrap. $y \in C^k \Rightarrow F(x,y) \in C^k \Rightarrow y' \in C^k \Rightarrow y \in C^{k+1}$, iterated to C^∞; the same induction gives analyticity for analytic F (Putnam 2010 B5, Putnam 2009 B5). The abstract scheme is: closure of a class under an operator T, plus an inequality I_k valid on the class, yields I_{k+1} by applying I_k to Tu and undoing T. Putnam 2018 A5 is exactly that with $T = d/dx$ on the absolutely monotonic functions.

7. Halving time blocks. Given $q(x)$ decaying to 0, set $x_0 = 0$ and $x_{n+1} = \min\{x > x_n : q(x) < q(x_n)/2\}$. Infinitely many such times exist, the blocks are disjoint, and a uniform positive lower bound on each block sums to infinity. This is the safe fallback route for Putnam 2010 A6.

8. Rolle chains and operator factorisation. Covered above: n differentiations cost n zeros, and $L(D) = \prod(D - r_i)$ is realised by n exponential substitutions (Putnam 2015 B1).

9. Minkowski sums of intervals. $I_{a,b} = I_{a-1,b} + I_{1,0} = I_{a,b-1} + I_{0,1}$ with $I_{0,0} = \{0\}$ solves to $I_{a,b} = aI_{1,0} + bI_{0,1}$, and generally reachable displacements with step sets J_i used a_i times form the interval $\sum a_i J_i$, of length $\sum a_i|J_i|$, with endpoints the corresponding integer combinations. The widening tells you in advance that a squeeze must be done one endpoint at a time (Putnam 2016 B5).

10. Dyadic and annular decomposition. For an integrand decaying like r^{-2} in the plane, annuli of fixed ratio λ carry equal limiting mass $m = \log\lambda \int_0^{2\pi} h$, so the integral over the disc of radius $R\lambda^k$ grows like km: logarithmic divergence is the telescoping sum of a constant sequence. For decay $r^{-n-\varepsilon}$ the masses are geometric and the total converges (Putnam 2021 A4).

- **WHEN** a recursion has the form $x_{n+1} = x_n - cx_n^{p+1} + \cdots$ with a neutral fixed point at 0 **DO** substitute $b_n = x_n^{-p}$ and read off $b_n \sim cpn$ **BECAUSE** the substitution is not a guess: it is the unique power whose increments are asymptotically constant, which is what makes the recursion additive and hence solvable (Putnam 2020 A3)
- **WHEN** a recursion adds a function of the current term **DO** factor it as $a_{n+1} = a_n(1 + f(a_n)/a_n)$ before doing anything else **BECAUSE** only in multiplicative form can you raise the recursion to a power and close it in a new variable (Putnam 2006 B6)
- **WHEN** an induction would need constants tracked through infinitely many steps **DO** restate every bound with "for all sufficiently large n" on both sides **BECAUSE** the eventually quantified lemma composes with itself with no bookkeeping, which is the difference between a writable proof and an unwritable one (Putnam 2025 B6)
- **WHEN** a quantity decays to zero and you need infinitely many disjoint regions of definite size **DO** define halving times $x_{n+1} = \min\{x > x_n : q(x) < q(x_n)/2\}$ **BECAUSE** each block then contributes at least a fixed amount, and infinitely many blocks give divergence (Putnam 2010 A6)
- **WHEN** an integral or sum over an unbounded region sits exactly at the borderline decay rate **DO** decompose into annuli or dyadic blocks of fixed ratio and compare the block masses **BECAUSE** equal block masses mean logarithmic divergence and geometric block masses mean convergence, and this diagnosis takes one line (Putnam 2021 A4)

6.15 What to learn last

Everything above earns its place by recurrence. The following appeared once each in twenty years of papers, so postpone them until the rest is automatic.

The Fejer and Riesz theorem (a nonnegative trigonometric polynomial is $|P(e^{i\theta})|^2$) and the Herglotz and Bochner theorem (positive definiteness equals nonnegativity of the Fourier symbol) appeared in Putnam 2013 B2 and Putnam 2013 A6; the second tells you that in two or more variables

the autocorrelation representation is unavailable, so proving the symbol nonnegative is the only route. Bernstein's theorem on completely monotone functions (Putnam 2018 A5) gives a three line solution to a problem you can also do by hand, but needs measure theory and a Helly compactness argument, so it is installed knowledge, not arena reconstruction. The Pompeiu problem and Delsarte's two radius theorem (Putnam 2019 A4) tell you at once that spherical means of a single radius are not injective, so you commit to constructing within seconds; the construction is the Fourier transform of surface measure, $\sin|\xi|/|\xi|$ in $\mathbb{R}^3$, which vanishes at $|\xi| = \pi$, putting $\sin(\pi z)$ in the kernel. Rademacher's theorem (Putnam 2014 B6) is the tempting wrong tool there, since an almost everywhere statement cannot see the rationals. The Glaisher and Kinkelin constant (Putnam 2007 B6) and the Green's functions for $-u''$ that explain the constant $1/8$ and the kernel $\min(x, y)$ (Putnam 2007 B2, Putnam 2014 A2) are Layer 3 context, not solving tools.

Two more items belong here for a different reason: they are named so that you stop looking. Lambert type series $\sum_{n\geq 0} 1/(q^n + c)$ have no elementary closed form and are known to be irrational, so recognising one (Putnam 2016 B6) is what stops the zombie route immediately and forces the structural approach. And the Steinhaus difference theorem (Putnam 2022 B3) is unavailable against an arbitrary colouring. Knowing that a tool does not exist is worth as much as knowing one that does, because it is the difference between an hour spent and an hour saved.

Chapter 7

Polynomials, complex numbers and functional equations

7.1 What this chapter is for

You already own the quadratic discriminant, the factor and remainder theorems for real polynomials, the relations between roots and coefficients up to the quartic, de Moivre, the nth roots of unity and complex loci. That is Edexcel core pure plus FP1 and FP2, and it is not enough. The Putnam does not ask you to solve a cubic. It asks you to prove that a polynomial you never write down has all its roots real, or that two polynomials in an infinite family are coprime, or that a function satisfying a strange identity must be $\sqrt{1-x^2}$.

The gap has four parts.

First, you treat a polynomial as a function. The Putnam treats it as an element of a ring $k[x]$: a degree, a divisibility theory, a formal derivative that exists over $\mathbb{F}_p$ where limits do not, and an identity theorem that upgrades an approximate statement about large integers into an exact statement about coefficients. Putnam 2007 A4 and 2009 B4 are won at that upgrade and nowhere else.

Second, you know Vieta up to degree four. You do not know the elementary symmetric functions as a system, you do not know Newton's identities, and you have never used the fact that every symmetric polynomial in the roots is an integer polynomial in the e_m. Putnam 2022 A6 is unsolvable without Newton.

Third, you have no finite difference calculus. The shift operator E, the difference operator $\Delta = E - I$, the annihilation law $\Delta^m P = 0$ for $\deg P < m$, and Newton's forward formula are the discrete twin of the calculus you already have, and they carry Putnam 2008 A5, 2010 B4, 2019 B5, 2020 B1, 2023 A5 and 2024 B5 outright.

Fourth, you have no theory of where roots are. Cauchy's bound, Enestrom and Kakeya, Gauss and Lucas and the left half plane positivity lemma are what let coefficient information reach the roots and back. Putnam 2014 A5, 2018 B2, 2019 A3 and 2021 A6 are all root location problems in different coats.

To that I add the two fields the Putnam keeps welding onto polynomials: complex numbers used as plane geometry and as a filter, and functional equations, which are not a topic at all but a protocol for interrogating an unknown function with substitutions until it confesses.

At the end of this chapter, when you meet a polynomial you should already be asking: what ring is it in, what is its degree, where are its roots, what is its reversal, what does Δ do to it, and what symmetric function of the roots does the question actually want. When you meet a functional equation you should already be listing the substitutions the equation permits, and asking whether the substitution map has finite order or a fixed point. That reflex is the whole content of what follows.

7.2 Polynomials as formal objects

Definition 7.2.1. Let R be a commutative ring. A polynomial over R is a sequence $(a_0, a_1, a_2, \dots)$ of elements of R with only finitely many nonzero, written $\sum_i a_i x^i$. Addition is termwise; multiplication is the convolution $(ab)_k = \sum_{i+j=k} a_i b_j$. The set of these is the ring $R[x]$. The degree of a nonzero polynomial is the largest i with $a_i \neq 0$; by convention $\deg 0 = -\infty$.

The point of the definition is that a polynomial is its coefficient list, not the function it induces. The two are different objects, and over $\mathbb{F}_p$ they differ in substance: $x^p - x$ is a nonzero polynomial of degree p whose associated function on $\mathbb{F}_p$ is identically zero, by Fermat. Whenever you argue "these two expressions agree, therefore they are the same polynomial", you have used a theorem, and you must know which one.

Lemma 7.2.2 (Degree law)**.** *If R is an integral domain then $\deg(fg) = \deg f + \deg g$ and the leading coefficient of fg is the product of the leading coefficients. Hence $R[x]$ is an integral domain and its units are the units of R.*

Proof. The coefficient of x^{m+n} in fg is $f_m g_n$, a product of nonzero elements of a domain, hence nonzero; higher coefficients vanish because every contributing product has a factor beyond the degree. □

That one line is doing serious work at Putnam 2024 A2, where the entire case analysis is "a divisor of a nonzero constant is a nonzero constant, and if $A \mid B$ with $B \neq 0$ then $\deg A \leq \deg B$".

Theorem 7.2.3 (Division with remainder)**.** *Let R be commutative and let $D \in R[x]$ be monic of degree d. For every $g \in R[x]$ there are unique $Q, R_0 \in R[x]$ with $g = QD + R_0$ and $\deg R_0 < d$.*

Proof. Induct on $\deg g$. If $\deg g < d$ take $Q = 0$. Otherwise subtract $g_n x^{n-d} D$, which cancels the leading term because D is monic, and apply the inductive hypothesis. Uniqueness: a difference of two representations gives $(Q_1 - Q_2)D = R_2 - R_1$ of degree less than d, and the degree law for a monic D forces $Q_1 = Q_2$. □

Monic is the hypothesis that matters. Over $\mathbb{Z}$ you cannot divide by $2x + 1$, but you can always divide by anything monic, and that is why so many Putnam arguments quietly normalise to monic first. Putnam 2010 B6 is precisely this: division by the monic $xP(x)$ shows that $\mathbb{R}[x] = \operatorname{span}\{1, x, \dots, x^{d-1}\} + D\mathbb{R}[x]$, which converts an induction over all exponents k into a single spanning statement.

Theorem 7.2.4 (Factor theorem and the root bound)**.** *Let R be a commutative ring, $f \in R[x]$, $a \in R$. Then $(x - a) \mid f(x) - f(a)$. If R is an integral domain and $f \neq 0$, then f has at most $\deg f$ roots in R.*

Proof. For the first part, $f(x) - f(a) = \sum_i a_i (x^i - a^i)$ and $x^i - a^i = (x - a)(x^{i-1} + x^{i-2}a + \cdots + a^{i-1})$. For the second, induct on the degree: if $f(a) = 0$ write $f = (x - a)g$; any other root b gives $(b - a)g(b) = 0$, and in a domain with $b \neq a$ this forces $g(b) = 0$, so the other roots are roots of g, which has degree $\deg f - 1$. □

Integral domain is not decoration. Over $\mathbb{Z}/8$ the polynomial $x^2 - 1$ has four roots. Every time you say "at most d roots" you are asserting that you are over a domain, usually a field, and on the Putnam that field is often $\mathbb{F}_p$: Putnam 2025 B5 bounds an exceptional set by 6 purely because three quadratics over $\mathbb{F}_p$ have at most two roots each, and never computes the set at all. That is the lazy route in its purest form.

Corollary 7.2.5 (Identity theorem)**.** *Let k be an infinite integral domain. A polynomial in $k[x]$ vanishing at infinitely many points is the zero polynomial. Hence two polynomials of degree at most d agreeing at $d + 1$ distinct points are equal, and the map from polynomials to polynomial functions is injective.*

This is the single most reused fact in the field, and it takes three shapes in the arena.

The first is the upgrade from an asymptotic identity to an exact one. Putnam 2007 A4 asks for all real f sending repunits to repunits. After the affine normalisation $g(x) = 9f((x-1)/9)+1$ (legitimate because $F(x) \mapsto F(\alpha x+\beta)$ with $\alpha \neq 0$ is a degree preserving ring automorphism of $\mathbb{R}[x]$), the hypothesis reads $g(10^m) = 10^c(10^m)^d$ for all large m, the exponents stabilising because g has a leading term. Two polynomials agreeing at infinitely many points 10^m are equal, so $g(x) = 10^c x^d$. What did the work was not the arithmetic of repunits but the fact that infinitely many agreements is one agreement.

The second is the collapse of an uncountable family of conditions to finitely many. Putnam 2009 B4 calls a two variable polynomial balanced when its mean over every circle centred at the origin is zero; that mean is a polynomial in the radius r vanishing for all $r > 0$, so its coefficients vanish and the dimension count follows.

The third is genericity. Putnam 2015 A6 needs $B - tI$ invertible. You may not assume it, but you may prove the identity for all but at most n values of t, note both sides are polynomials in t of degree at most n, and conclude at $t = 0$. That is the standard licence to assume nondegeneracy: prove it generically, then let the identity theorem carry you into the degenerate case with no analysis at all.

- **WHEN** a hypothesis holds for all sufficiently large n, or for all r in an interval, or for all but finitely many values of a parameter, and both sides are polynomial in that parameter **DO** declare the two sides equal as polynomials and compare coefficients **BECAUSE** a nonzero polynomial has finitely many roots, so infinitely many agreements is an identity of coefficients (Putnam 2007 A4, 2009 B4, 2015 A6)
- **WHEN** you want to prove $\deg R \geq m$ **DO** manufacture m distinct roots of R and quote the root bound **BECAUSE** degree is bounded below by the number of distinct roots, which converts a construction into an inequality (Putnam 2008 A5)
- **WHEN** a count or a bound over $\mathbb{F}_p$ involves how often a fixed polynomial condition holds **DO** bound the count by the degree instead of solving the condition **BECAUSE** over a field, degree d means at most d roots, which is usually all the inequality needs (Putnam 2025 B5)
- **WHEN** a divisibility $A \mid B$ is derived in $k[x]$ with $B \neq 0$ **DO** immediately write $\deg A \leq \deg B$ and see which degrees survive **BECAUSE** the degree law makes divisibility a numerical constraint, and most cases die there (Putnam 2024 A2)

7.3 Roots, multiplicity and the formal derivative

Theorem 7.3.1 (Fundamental theorem of algebra)**.** *Every nonconstant $f \in \mathbb{C}[x]$ has a root in $\mathbb{C}$. Hence $f = c\prod_{j=1}^{d}(x-\alpha_j)$ with $d = \deg f$, and $\mathbb{C}[x]$ is a unique factorisation domain whose primes are the monic linear polynomials.*

I do not prove this; it is analysis, not algebra, and no Putnam solution ever reproves it. What you must own is its consequences. Over $\mathbb{R}$, nonreal roots of a real polynomial occur in conjugate pairs, so every real polynomial factors into monic linear factors and monic quadratics $x^2 - 2\operatorname{Re}(\alpha)x + |\alpha|^2$. That factorisation is load bearing at Putnam 2021 A6, where a positivity lemma is proved factor by factor and would be unprovable in one piece.

Unique factorisation in $\mathbb{C}[x]$ turns a polynomial equation into a combinatorial choice. Putnam 2007 B4 wants the number of real pairs with $P^2 + Q^2 = X^{2n} + 1$ and $\deg P > \deg Q$. Write $A = P + iQ$ and σ for coefficientwise conjugation: the equation becomes $A\,\sigma(A) = X^{2n} + 1$. The right side is squarefree with $2n$ roots in n conjugate pairs, so every divisor is a unit times $\prod_{m\in S}(X - \zeta_m)$, and the equation says S and its conjugate are complementary: one root from each

pair, times a constant that the degree and leading coefficient argument pins to ± 1. The count is 2^{n+1}. An additive problem became a subset selection because you passed to a UFD.

Definition 7.3.2. The formal derivative on $R[x]$ is $\left(\sum a_i x^i\right)' = \sum i a_i x^{i-1}$. It is R linear and satisfies the product rule and the chain rule $(g \circ h)' = (g' \circ h)h'$, all provable by checking on monomials.

The word formal matters: this exists over $\mathbb{F}_p$, over $\mathbb{Z}$, over any commutative ring, with no limits anywhere.

Proposition 7.3.3 (Multiple root criterion)**.** *Over a field, z is a root of h of multiplicity at least two if and only if $h(z) = h'(z) = 0$.*

Proof. Write $h = (x - z)q$. Then $h' = q + (x - z)q'$, so $h'(z) = q(z)$, which vanishes exactly when $(x - z)$ divides q. □

In characteristic p that is the whole story only up to multiplicity $p - 1$. The mth derivative carries a factor $m!$, so it cannot detect multiplicity p or beyond: $x^p - 1$ has derivative identically zero. The repair is the Hasse derivative, the coefficient extractor $P^{(i)}/i!$ from the Taylor expansion, which is defined over any ring. Putnam 2019 A5 lives exactly on this boundary, and its induction is valid only because the orders in play stay between 1 and $p - 1$; if you skip the range check the proof is wrong.

Proposition 7.3.4 (Exact Taylor expansion)**.** *For $g \in R[x]$ and indeterminates or elements a, h,*

$$g(a + h) = \sum_{i \geq 0} \frac{g^{(i)}(a)}{i!} h^i,$$

a finite identity in which every Hasse coefficient $g^{(i)}(a)/i!$ lies in R when $R = \mathbb{Z}$. In particular $g(a + h) \equiv g(a) + g'(a)h \pmod{h^2}$.

Proof. Both sides are additive in g, so check $g = x^m$: the binomial theorem gives $(a + h)^m = \sum_i \binom{m}{i} a^{m-i} h^i$, and $\binom{m}{i}$ is exactly the ith Hasse coefficient of x^m at a. Integrality is the integrality of binomial coefficients. □

This is the algebraic core of Hensel's lemma, and it is the single tool of Putnam 2008 B4: taking the increment $t = p$ turns the injectivity hypothesis modulo p^2 into the statement $p \nmid h'$, and taking $t = yp^2$ makes h affine on each fibre of $\mathbb{Z}/p^3 \to \mathbb{Z}/p^2$, which is exactly the lifting statement wanted. It also drives Putnam 2025 A6, where the 2-adic valuation of $b_{2^{k+1}} - 2b_{2^k}$ is computed by expanding an iterate as a value plus a derivative term plus a controlled tail.

Here is the cleanest illustration, and you should be able to reproduce it in three minutes.

Example 7.3.5 (Putnam 2024 A2)**.** For which real polynomials p is there a real polynomial q with $p(p(x)) - x = (p(x) - x)^2 q(x)$?

Set $r = p - x$, the fixed point polynomial. Then $p(p(x)) = p(x + r) = p(x) + r\,p'(x) + r^2 s(x)$ for some polynomial s, by the exact Taylor expansion with increment r. Hence

$$p(p(x)) - x = r + r\,p' + r^2 s = r\,(1 + p') + r^2 s,$$

so the condition is $r^2 \mid r(1 + p')$, that is $r = 0$ or $r \mid 1 + p'$.

If $r = 0$ then $p(x) = x$. If $1 + p' = 0$ then $p = -x + c$ and both sides are zero. Otherwise $1 + p' \neq 0$ has degree at most $\deg p - 1$, so $\deg r \leq \deg p - 1$. But $r = p - x$ has degree $\deg p$ as soon as $\deg p \geq 2$, which kills every degree above one; and $\deg p = 0$ makes $(c - x)q = 1$, impossible. That leaves $p = mx + b$ with $r = (m - 1)x + b$ dividing the nonzero constant $1 + m$, so r is constant, so $m = 1$. The answer is $p(x) = x + c$ or $p(x) = -x + c$: translations and involutions, exactly the maps for which the second iterate is either trivial or the identity.

- **WHEN** a hypothesis involves $g(a+h)$ or an iterate $p \circ p$, and you want divisibility by h^2 or by $(p-x)^2$ **DO** write the exact Taylor expansion $g(a+h) = g(a) + g'(a)h + h^2(\dots)$ and read the condition off the linear term **BECAUSE** the expansion is an identity in the polynomial ring, so it converts a composition into a value plus a derivative plus a controlled tail (Putnam 2024 A2, 2025 A6, 2008 B4)
- **WHEN** you must produce a repeated root, or exclude one **DO** differentiate formally and test $h(z) = h'(z) = 0$, checking the characteristic first **BECAUSE** multiplicity at least two is exactly a common root of h and h', but only up to multiplicity $p-1$ in characteristic p (Putnam 2011 B6, 2019 A5)
- **WHEN** a real polynomial identity must be proved factor by factor **DO** factor over $\mathbb{R}$ into monic linear and monic quadratic pieces first **BECAUSE** the real factorisation exists by the fundamental theorem plus conjugate pairing, and multiplicative properties then only need checking on the pieces (Putnam 2021 A6)

7.4 Symmetric functions: Vieta, Newton and the power sum machine

Definition 7.4.1. For $u_1, \dots, u_n$ the elementary symmetric functions are defined by the generating identity

$$\prod_{j=1}^{n}(1+xu_j) = \sum_{k=0}^{n} e_k(u)\, x^k,$$

so $e_0 = 1$, $e_1 = \sum u_j$, $e_n = \prod u_j$. The power sums are $s_j = \sum_i u_i^j$.

Go back to the generating identity whenever you are unsure of a sign or an index; every Vieta computation falls out of it. Expanding $\prod_j (t-u_j)$ instead gives the coefficient form: if $p(x) = c_n x^n + \cdots + c_0$ has roots $a_1, \dots, a_n$ with multiplicity, then

$$e_m(a_1, \dots, a_n) = (-1)^m \frac{c_{n-m}}{c_n}.$$

You know this for $n \le 4$. Three things about it are new.

First, the two ends are usually all you need. Putnam 2023 A2 asks, for even n and a monic real p of degree $2n$ with $p(1/k) = k^2$ for $1 \le |k| \le n$, for the other real x with $p(1/x) = x^2$. Multiply the condition by x^{2n}: with the reversal $p^*(x) = x^{2n}p(1/x)$, which is a polynomial of degree at most $2n$ with $p^*(0) = 1$ because p is monic, set $q(x) = x^{2n+2} - p^*(x)$. Then q is monic of degree $2n+2$, $q(0) = -1$, and $q(k) = 0$ for the $2n$ values $1 \le |k| \le n$. By the factor theorem $q(x) = \prod_{k=1}^{n}(x^2-k^2)\cdot(x^2+bx+c)$. Now use Vieta at the two ends: q has no x^{2n+1} term because $\deg p^* \le 2n$, and the first factor is even, so $b = 0$; and the constant terms give $(-1)^n(n!)^2 c = -1$, so with n even, $c = -1/(n!)^2$. The remaining roots are $x = \pm 1/n!$. Not one coefficient of p was ever computed. That is the discipline: a determined linear system does not oblige you to solve it, only to identify the symmetric functions the answer depends on.

Second, distinctness and degree are preconditions, not formalities. Knowing that p vanishes at n given values identifies them as all the roots only if $\deg p = n$ and the values are pairwise distinct. In Putnam 2006 A5 the polynomial Q_n built from the n tangents has degree exactly n because its leading coefficient $i^n(1+\omega)$ is nonzero, and the $a_k = \tan(\theta + k\pi/n)$ are distinct because tan is injective modulo π and the angles are pairwise incongruent modulo π. This is the most common place to write the correct final answer and still lose marks.

Third, and this is the real content: every symmetric polynomial in the roots is a polynomial with integer coefficients in the e_m. So any question about the roots that does not distinguish between them is a question about the coefficients, and you never find the roots.

Theorem 7.4.2 (Newton's identities)**.** *With e_k and s_j as above for n quantities,*

$$j\,e_j = \sum_{i=1}^{j} (-1)^{i-1} e_{j-i}\, s_i \qquad (1 \le j \le n),$$

and for $j > n$,

$$s_j = e_1 s_{j-1} - e_2 s_{j-2} + \cdots + (-1)^{n-1} e_n s_{j-n}.$$

Proof. Let $P(t) = \prod_i (1 - u_i t) = \sum_k (-1)^k e_k t^k$. Logarithmic differentiation gives

$$-\frac{P'(t)}{P(t)} = \sum_i \frac{u_i}{1 - u_i t} = \sum_{m \ge 0} s_{m+1} t^m,$$

so $-P'(t) = P(t) \sum_{m \ge 0} s_{m+1} t^m$. The coefficient of t^{j-1} on the left is $(-1)^{j-1} j e_j$ and on the right is $\sum_{k=0}^{j-1} (-1)^k e_k s_{j-k}$; multiplying by $(-1)^{j-1}$ and substituting $i = j - k$ gives the first family. The second family is the statement that $P(t) \sum s_{m+1} t^m$ has no coefficients beyond degree $n - 1$ on the left, since $\deg P' \le n - 1$. □

The second family deserves a name in your head: it says the sequence (s_j) satisfies a linear recurrence of order n whose characteristic polynomial is the monic polynomial itself. That is unsurprising once you see it, since $s_j = \sum_i r_i^j$ and each r_i^j satisfies that recurrence. It is the bridge in both directions between "I know the roots' powers" and "I know the coefficients".

Example 7.4.3 (Putnam 2022 A6)**.** You want the largest m such that there are $-1 < x_1 < \cdots < x_{2n} < 1$ with $\sum_{i=1}^{n} \left(x_{2i}^{2k-1} - x_{2i-1}^{2k-1} \right) = 1$ for $k = 1, \ldots, m$. Because $(-x)^{2k-1} = -x^{2k-1}$, form the multiset $R = \{x_2, x_4, \ldots, x_{2n}\} \cup \{-x_1, -x_3, \ldots, -x_{2n-1}\}$ of size $2n$. The hypothesis is exactly $s_{2k-1}(R) = 1$ for $k \le m$. Newton then converts vanishing or constancy of odd power sums into information about the odd indexed e_j, and the answer is $m = n$. The construction that attains it is the Chebyshev node configuration $x_j = \cos \frac{(2n+1-j)\pi}{2n+1}$, whose power sums evaluate by roots of unity orthogonality; and the reason the construction stops at $k = n + 1$ is that the orthogonality hypothesis $u \not\equiv 0$ fails there.

Two further symmetric facts you must own.

Proposition 7.4.4 (Odd power sums detect negation symmetry)**.** *A finite multiset $R \subset \mathbb{C}$ satisfies $R = -R$ if and only if $s_j(R) = 0$ for every odd j. Equivalently, a monic polynomial of odd degree N with $e_j = 0$ for all odd $j \le N$ is an odd function.*

Proposition 7.4.5 (Equal power sums force equal multisets)**.** *Two multisets of size at most N with equal jth power sums for $j = 1, \ldots, N$ are equal.*

The second follows from Newton (equal power sums give equal e_j, hence the same polynomial) in characteristic zero, or from invertibility of the Vandermonde matrix, whose determinant is $\prod_{i<j} (r_j - r_i)$.

Two named inequalities orbit this material and are worth knowing by name only. Maclaurin: for nonnegative x_i with $S_k = e_k / \binom{n}{k}$, $S_1 \ge S_2^{1/2} \ge \cdots \ge S_n^{1/n}$. Newton: if $\sum a_k x^k$ has all real roots then $p_k^2 \ge p_{k-1} p_{k+1}$ for $p_k = a_k / \binom{n}{k}$. Each appeared once, at Putnam 2019 A3 and 2014 B4, and in both cases as a check, since Newton's inequalities are necessary and not sufficient for real rootedness. Thirty seconds, then move on.

Example 7.4.6 (Putnam 2019 A1)**.** Determine all values of $A^3 + B^3 + C^3 - 3ABC$ for nonnegative integers A, B, C. Newton for three variables gives $s_3 = e_1^3 - 3e_1e_2 + 3e_3$, so the expression is $s_3 - 3e_3 = e_1^3 - 3e_1e_2$, which depends on e_1 far more than on anything else. If $3 \mid e_1$ then 9 divides $e_1^3 - 3e_1e_2$; if $3 \nmid e_1$ then $e_1^3 \equiv \pm 1$ and the whole expression is ± 1 modulo 3. Either way it is never congruent to 3 or 6 modulo 9, and every other nonnegative integer occurs. The answer is $\{m \ge 0 : m \not\equiv 3, 6 \pmod 9\}$.

The same expression factors as $(A+B+C)(A^2+B^2+C^2-AB-BC-CA) = \frac{1}{2}(A+B+C)\left((A-B)^2+(B-C)^2+(C-A)^2\right)$, which gives its sign for free and vanishes only when the three are equal. That factorisation is also the entire content of Putnam 2006 B1: the curve $x^3+3xy+y^3=1$ is $x^3+y^3+(-1)^3-3xy(-1)=0$, hence the union of the line $x+y=1$ and the single point $(-1,-1)$ where the quadratic factor $\frac{1}{2}((x-y)^2+(x+1)^2+(y+1)^2)$ vanishes. An equilateral triangle then has its apex at $(-1,-1)$ and its base on the line at distance $3/\sqrt{2}$, giving side $\sqrt{6}$ and area $3\sqrt{3}/2$. Test this factorisation before you attempt anything else on a two or three variable cubic with cube terms and a coefficient 3 cross product.

- **WHEN** a question asks for a quantity symmetric in the roots **DO** express it in the e_m by the generating identity and read the e_m off the coefficients **BECAUSE** symmetric functions of the roots are integer polynomials in the coefficients, so the roots never need to be found (Putnam 2006 A5, 2019 A3, 2023 A2)
- **WHEN** power sums are given and coefficients are wanted, or the reverse **DO** run Newton's identities in the appropriate direction **BECAUSE** Newton is the only bridge between the two languages, and the high index family is a linear recurrence with the polynomial as characteristic polynomial (Putnam 2022 A6, 2019 A1, 2015 A6)
- **WHEN** odd power sums vanish or are all equal **DO** test for the symmetry $R=-R$ and conclude the polynomial is odd **BECAUSE** odd power sums are exactly the obstruction to negation symmetry (Putnam 2022 A6)
- **WHEN** you are about to apply Vieta to roots you have manufactured **DO** check the degree equals the number of roots and that they are pairwise distinct, in writing **BECAUSE** Vieta identifies these as all the roots only under both conditions, and this is the standard place marks are lost (Putnam 2006 A5)
- **WHEN** a two or three variable cubic has cube terms and a $3xyz$ style cross term **DO** test $x^3+y^3+z^3-3xyz=(x+y+z)(x^2+y^2+z^2-xy-yz-zx)$ before anything else **BECAUSE** that identity collapses a curve into a line and a point, or a congruence into a statement about e_1 alone (Putnam 2006 B1, 2019 A1)

7.5 Finite differences, the shift operator and interpolation

This is the section with the highest ratio of Putnam yield to page count, and A-level gave you nothing of it.

Definition 7.5.1. On functions or sequences define the shift $(EP)(x)=P(x+1)$, the identity I, and the difference operator $\Delta = E-I$, so $(\Delta P)(x)=P(x+1)-P(x)$. More generally $(\Delta_h P)(x)=P(x+h)-P(x)$.

Lemma 7.5.2 (Degree behaviour)**.** *For nonconstant P, $\deg \Delta_h P = \deg P - 1$ for $h\neq 0$, and the leading coefficient is multiplied by $d\,h$ where $d=\deg P$. Hence $\Delta_{h_1}\cdots\Delta_{h_m}P=0$ whenever $\deg P<m$, and $\Delta_{h_1}\cdots\Delta_{h_m}x^m = m!\,h_1\cdots h_m$. More generally $E-tI$ preserves degree and multiplies the leading coefficient by $1-t$ when $t\neq 1$, and lowers the degree by exactly one when $t=1$.*

Proof. $(x+h)^d-x^d = dhx^{d-1}+\cdots$ by the binomial theorem, and the top terms of lower degree monomials cannot interfere. For $E-tI$, the leading coefficient of $P(x+1)-tP(x)$ is $(1-t)$ times that of P. □

The $t\neq 1$ half of that lemma is the entire engine of Putnam 2008 A5. There, f,g real polynomials place $(f(k),g(k))$ for $k=1,\dots,n$ at the vertices of a regular n-gon counterclockwise. Package them as $P=f+ig\in\mathbb{C}[z]$, noting $\deg P=\max(\deg f,\deg g)$ exactly because f_j+ig_j vanishes only when both real coefficients do. A regular n-gon with centre b is $\{b+a\omega^k\}$ with $\omega=e^{2\pi i/n}$,

and the orientation preserving similarities $z \mapsto az + b$ form a group, so after a degree preserving normalisation the hypothesis reads $Q(k) = \omega^k$ for $k = 1, \dots, n$. Now apply $E - \omega I$: it preserves degree, and $(E - \omega I)Q$ vanishes at $k = 1, \dots, n-1$. So $\deg Q = \deg(E - \omega I)Q \geq n - 1$, giving $\max(\deg f, \deg g) \geq n - 1$. There is a one shot variant: $\Delta^{n-1}Q(1) = \omega(\omega - 1)^{n-1} \neq 0$, which also shows the bound is attained.

Theorem 7.5.3 (Newton's forward difference formula)**.** *For any P and any nonnegative integer t,*

$$P(t) = \sum_{k \geq 0} \binom{t}{k} (\Delta^k P)(0),$$

the sum terminating at $k = \deg P$. More generally $P(x) = \sum_{k \geq 0} (\Delta^k P)(x_0) \binom{x - x_0}{k}$.

Proof. $E = I + \Delta$ and the two commute, so $E^t = (I + \Delta)^t = \sum_k \binom{t}{k} \Delta^k$ by the binomial theorem for commuting operators; evaluate at 0. The sum terminates because $\Delta^{d+1} P = 0$ for $\deg P \leq d$. □

This is the discrete Taylor series, and the binomial coefficients $\binom{x}{k}$ are the discrete monomials. Two consequences you should keep permanently: a polynomial is integer valued on the integers if and only if it is an integer combination of the $\binom{x}{k}$, and any "values at consecutive integers" interpolation is solved in one line by writing the difference table.

Example 7.5.4 (Putnam 2019 B5)**.** Let p have degree 1008 with $p(2n + 1) = F_{2n+1}$ for $n = 0, \dots, 1008$. Substitute $P(t) = p(2t + 1)$, so P has degree 1008 with $P(t) = F_{2t+1}$ at $t = 0, \dots, 1008$, unique because a difference of two interpolants would have 1009 roots. You want $P(1009)$, one node beyond the data, and Newton gives $P(1009) = \sum_{k=0}^{1008} \binom{1009}{k} (\Delta^k P)(0)$, truncated at $k = 1008$ precisely because the degree is 1008. The missing top term is the whole answer: running the same computation for small N gives $P(N + 1) = F_{2N+3} - F_{N+2}$, so $p(2019) = F_{2019} - F_{1010}$.

State the general principle on its own, because it recurs: the difference table of a sequence obeying a linear recurrence with characteristic roots λ_i has rows that are again combinations of geometric sequences, because Δ acts on λ_i^n as multiplication by $\lambda_i - 1$. Diagonalise Δ on the geometric components, apply $(I + \Delta)^t$, truncate at the degree.

Example 7.5.5 (Putnam 2020 B1)**.** With $d(n)$ the binary digit sum, evaluate $S = \sum_{k=1}^{2020} (-1)^{d(k)} k^3$ modulo 2020. Over an aligned block $16m, \dots, 16m + 15$ the sign factorises over the four binary digits, so the block sum is $(I - E^1)(I - E^2)(I - E^4)(I - E^8)$ applied to k^3, up to sign: four difference operators annihilate a cubic, so every aligned block of sixteen contributes zero. Since $2016 = 16 \cdot 126$, only $k = 2016, \dots, 2020$ survive, with signs $+, -, -, +, -$ from digit sums $6, 7, 7, 8, 7$ and residues $-4, -3, -2, -1, 0$, so $S \equiv -64 + 27 + 8 - 1 = -30 \equiv 1990$.

Example 7.5.6 (Putnam 2023 A5)**.** With $f(k)$ the number of ones in base three, the operator $\sum_{k=0}^{3^{1010}-1} (-2)^{f(k)} E^k$ factorises over the 1010 ternary digits as $\prod_{j=0}^{1009} \left(E^{2 \cdot 3^j} - 2E^{3^j} + I \right) = \prod_j \left(E^{3^j} - I \right)^2$, which is 2020 difference operators applied to $(z + k)^{2023}$ and drops the degree from 2023 to 3. That is why the problem says "all complex numbers": two of the three roots are not real. Palindromy of the weights, since $f(3^n - 1 - k) = f(k)$ because the base three digits complement to $2 - d_j$, makes the cubic odd about $z = -(3^{1010} - 1)/2$, so that point is a root and the others are the conjugate pair $-(3^{1010} - 1)/2 \pm i\sqrt{9^{1010} - 1}/4$.

Interpolation itself is the same theorem in a different basis.

Theorem 7.5.7 (Lagrange interpolation)**.** *Given distinct nodes $x_0, \dots, x_d$ in a field and prescribed values, there is exactly one polynomial of degree at most d taking those values, namely $\sum_i y_i \ell_i(x)$ with $\ell_i(x) = \prod_{m \neq i} \frac{x - x_m}{x_i - x_m}$.*

Uniqueness is the identity theorem again. The basis polynomial ℓ_i is the unique degree d polynomial equal to 1 at x_i and 0 elsewhere, so it manufactures indicator patterns to order: Putnam 2014 A4 needs a cubic which is 1 at 0 and 0 at $1, 2, 3$, and $\ell_0(x) = -(x-1)(x-2)(x-3)/6$ arrives already factored, which makes the sign check for $n \geq 4$ immediate. A second corollary: a polynomial of degree d taking rational values at $d+1$ distinct rational points has rational coefficients, because Lagrange writes the coefficients as rational combinations of the values.

Finally, the discrete antiderivative.

Proposition 7.5.8. *For a polynomial g of degree d there is a unique polynomial F of degree $d+1$ with $F(x) - F(x-1) = g(x)$ identically and $F(0) = 0$; then $F(n) = \sum_{z=1}^{n} g(z)$ for positive integers n.*

Because the identity holds on all of $\mathbb{R}$, you may evaluate it at negative or non integer points, which is exactly how Putnam 2024 B5 transports the zeros of g into zeros of the counting polynomial F and thereby proves that a combinatorial count has nonnegative coefficients.

- **WHEN** a sum runs over a block whose weights factorise over digits, or a hypothesis involves $P(x+h) - P(x)$ **DO** write it as a product of difference operators and count how many you have against the degree **BECAUSE** m difference operators annihilate every polynomial of degree less than m, so the whole block collapses to zero (Putnam 2020 B1, 2023 A5)
- **WHEN** a polynomial is specified by its values at consecutive integers and you want a value outside the range **DO** build the difference table and apply Newton's forward formula, watching where the truncation bites **BECAUSE** $E^t = (I + \Delta)^t$ terminates at the degree, and the missing top term is usually the entire answer (Putnam 2019 B5, 2008 A5)
- **WHEN** a hypothesis relates $P(x+1)$ and $P(x-1)$ to $P(x)$ **DO** convert it into a statement about $\Delta^k P$ and read off a degree bound **BECAUSE** $\Delta^{k+1} P = 0$ if and only if $\deg P \leq k$, so shift relations are degree statements in disguise (Putnam 2010 B4)
- **WHEN** you need a polynomial with a prescribed 0 and 1 pattern on a finite set **DO** write the Lagrange basis polynomial directly, and keep it factored **BECAUSE** ℓ_i is the unique such polynomial, and its factored form makes sign analysis immediate (Putnam 2014 A4)
- **WHEN** a combinatorial count is claimed to be a polynomial in n with a sign property **DO** find the polynomial F with $F(x) - F(x-1) = g(x)$ and evaluate the identity outside the combinatorial range **BECAUSE** the identity is an identity of polynomials, not of counts, so it may be evaluated at zeros of g (Putnam 2024 B5)

7.6 Reversal, palindromes and the substitution $y = z + 1/z$

Definition 7.6.1. For p of degree d the reversal is $p^*(x) = x^d p(1/x)$, whose coefficient list is that of p read backwards. If $p(0) \neq 0$, the roots of p^* are exactly the reciprocals of the roots of p. A polynomial is palindromic when $p^* = p$.

Reversal is free information; compute it reflexively. It turns an upper bound theorem for root moduli into a lower bound theorem, which is how Putnam 2014 A5 gets its two incompatible facts. It links the polynomial with coefficients $n, n-1, \ldots, 1$ at Putnam 2018 B2 to the one with coefficients $1, 2, \ldots, n$. It supplies the constant term -1 in the Putnam 2023 A2 computation above, precisely because p was monic. And it transfers a root count from a truncated exponential back to a polynomial of factorial coefficients at Putnam 2011 B6, because $x \mapsto \lambda/x$ is a bijection of $\mathbb{F}_p^\times$ and both have constant term 1.

For elementary symmetric functions the reversal law is

$$e_k(1/u_1, \ldots, 1/u_n) = \frac{e_{n-k}(u)}{e_n(u)},$$

which is the same statement in the language of coefficients.

Palindromy pays twice. First, the roots come in pairs $z, 1/z$, so half the analysis is a reflection of the other half; Putnam 2022 A1 reads $r_- r_+ = 1$ off the palindromic quadratic $ax^2 - 2x + a$ and places one root in $(0,1)$ and one in $(1,\infty)$ without solving. Second, palindromy is preserved under multiplication, so if p is palindromic then so is p^2, which halves the verification of the extremal example in Putnam 2022 A2.

Third, and most usefully, for even degree $2m$ the substitution $y = z + 1/z$ halves the degree: divide by z^m and use that $z^j + z^{-j}$ is a polynomial in y. That polynomial is $2T_j(y/2)$, which is why this substitution and Chebyshev are the same fact wearing different clothes. For $y > 0$ real, $y + 1/y = 2\cosh(\ln y)$; for $y = e^{i\theta}$ on the unit circle, $y + 1/y = 2\cos\theta$; and the identity $(y + 1/y)^2 - 2 = y^2 + 1/y^2$ is the double angle formula written multiplicatively. That is exactly why Putnam 2014 A3, with $a_0 = 5/2$ and $a_k = a_{k-1}^2 - 2$, is solved by writing $a_k = 2^{2^k} + 2^{-2^k}$: the recursion is angle doubling, and the infinite product telescopes to $3/7$.

A caveat worth clarifying: at Putnam 2014 B4 the substitution $y = z + 1/z$ is the tempting nominal route and must be rejected, because the polynomial there is palindromic only in the weak sense $T_j = T_{n-j}$ of its evaluation points, and the substitution does not simplify the sign pattern that carries the proof. Recognise the shape, then check that it actually buys something.

- **WHEN** a polynomial has a coefficient list worth reading backwards, or you need a lower bound on root moduli **DO** write down the reversal $p^*(x) = x^d p(1/x)$ and apply your upper bound theorem to it **BECAUSE** reversal inverts the roots, converting every upper bound into a lower bound and back (Putnam 2014 A5, 2018 B2, 2011 B6)
- **WHEN** the coefficients read the same in both directions **DO** use $r \mapsto 1/r$ on the roots at once, and for even degree substitute $y = z + 1/z$ **BECAUSE** palindromy is exactly invariance under root inversion, and the substitution halves the degree through Chebyshev (Putnam 2022 A1, 2022 A2, 2014 A3)
- **WHEN** a recursion has the shape $a_k = a_{k-1}^2 - 2$ or $x_{n+1} = 2x_n x_{n-1} - x_{n-2}$ **DO** substitute $a = y + 1/y$ or $x = \cos\theta$ and read the induced recursion on the exponent or the angle **BECAUSE** these are the multiplicative forms of the double angle and product to sum identities, so the nonlinear recursion becomes linear in the exponent (Putnam 2014 A3, 2018 B4)

7.7 Root location: where the zeros actually live

Coefficient information reaches the roots only through theorems of this section. Learn all four; they are short.

Theorem 7.7.1 (Cauchy bound)**.** *If $P(x) = x^n + \sum_{i<n} a_i x^i$ with $|a_i| \le M$ then every complex root satisfies $|\alpha| \le 1 + M$.*

Proof. If $|\alpha| > 1$, divide $P(\alpha) = 0$ by α^n and put $t = |\alpha|^{-1} \in (0,1)$: then $1 \le \sum_{k\ge 1} Mt^k < Mt/(1-t)$, so $1 - t < Mt$ and $|\alpha| = 1/t < 1 + M$. □

The proof is worth more than the statement, because keeping more terms sharpens it. Retain the first m terms on the exact side, in a sector where they have nonnegative real part: the threshold becomes $t^m + t - 1 > 0$, whose unique positive root τ_m increases to 1, giving $|\alpha| < 1/\tau_m$. For coefficients in $\{0,1\}$ that is $|\alpha| < 2$ in general, $|\alpha| < \frac{1+\sqrt{5}}{2}$ when $\operatorname{Re}\alpha \ge 0$, and $|\alpha| < \frac{3}{2}$ when $|\arg\alpha| \le \frac{\pi}{4}$. One recursion in the number of retained terms generates the whole family, so you can see in advance what each extra term buys.

Theorem 7.7.2 (Enestrom and Kakeya)**.** *If $0 < a_0 \le a_1 \le \cdots \le a_d$ and $d \ge 1$ then every root of $f = \sum a_t x^t$ satisfies $|z| \le 1$. Consequently, for arbitrary positive coefficients every root satisfies*

$$\min_k \frac{a_{k-1}}{a_k} \;\le\; |z| \;\le\; \max_k \frac{a_{k-1}}{a_k}.$$

Proof. Compute $(x-1)f(x) = a_d x^{d+1} - \sum_{t=1}^{d}(a_t - a_{t-1})x^t - a_0$. If $f(z) = 0$ with $|z| > 1$, then $a_d|z|^{d+1} \le \sum_t (a_t - a_{t-1})|z|^t + a_0 \le |z|^d\,(\sum_t (a_t - a_{t-1}) + a_0) = a_d|z|^d$, a contradiction. The annulus form follows by applying this to $f(Rx)$ for the upper bound and to the reversal for the lower. □

That theorem, together with reversal, is the whole of Putnam 2014 A5.

Example 7.7.3 (Putnam 2014 A5)**.** $P_n(x) = 1 + 2x + \cdots + nx^{n-1}$; prove P_j and P_k are coprime for $j \neq k$. Take $j < k$ and write $P_k = P_j + x^j Q$ with $Q = (j+1) + (j+2)x + \cdots + nx^{n-j-1}$. A common root z of P_j and P_k is nonzero (both have constant term nonzero), so $Q(z) = 0$ too. Enestrom and Kakeya on P_j gives $|z| \le \max_t \frac{t}{t+1} = \frac{j-1}{j}$; on Q it gives $|z| \ge \min_t \frac{j+t}{j+t+1} = \frac{j+1}{j+2}$. Since $(j-1)(j+2) < j(j+1)$, the two windows are disjoint. No common complex root, hence coprime in $\mathbb{Q}[x]$, hence coprime in $\mathbb{Z}[x]$ by Gauss's lemma since both have constant term 1.

The general theorem behind that is strict log concavity: if $a_t a_{t+2} < a_{t+1}^2$ for all t, so that the ratios a_t/a_{t+1} are strictly increasing, then all truncations of the series are pairwise coprime, because the head's window ends strictly before the tail's window begins. Putnam 2014 A5 is the instance $a_t = t + 1$. That is how a one off becomes a circuit.

Theorem 7.7.4 (Gauss and Lucas)**.** *Every root of p' lies in the convex hull of the roots of p; if p has only simple roots, no root of p' is a vertex of that hull.*

Proof. If $p'(w) = 0$ and $p(w) \neq 0$ then $0 = p'(w)/p(w) = \sum_k \frac{1}{w - r_k}$. Conjugating, $\sum_k \frac{w - r_k}{|w - r_k|^2} = 0$, so with $\lambda_k = |w - r_k|^{-2} > 0$ we get $w = (\sum_k \lambda_k r_k) \,/\, \sum_k \lambda_k$, a convex combination. □

At Putnam 2018 B2 this powers the alternative route: the roots of $h_n(z) = 1 + z + \cdots + z^n$ are the $(n+1)$th roots of unity other than 1, all on the unit circle, so the roots of $h_n' = g_n$ lie strictly inside the disc, and reversal transfers this to $f_n = n + (n-1)z + \cdots + z^{n-1}$, which therefore has no root in the closed unit disc.

The direct route there is the coefficient telescope, and the mechanism generalises. If the coefficients of $a(z) = \sum c_k z^k$ are in arithmetic progression, $(z-1)a(z)$ collapses to a geometric series plus boundary terms, since shift and subtract cancels the common difference termwise. For f_n that gives $(z-1)f_n(z) = z^n + z^{n-1} + \cdots + z - n$. If $|z| \le 1$ and $f_n(z) = 0$ then $|z^n + \cdots + z| = n$ with n terms of modulus at most 1, so by the equality case of the triangle inequality every z^k is 1, forcing $z = 1$; but $f_n(1) = n(n+1)/2 \neq 0$. In general $(z-1)\sum c_k z^k = \sum (c_{k-1} - c_k) z^k$ plus boundary terms, which is summation by parts for polynomials and works for any simple difference pattern.

The equality case used there is worth stating alone: if $|z_1 + \cdots + z_k| = |z_1| + \cdots + |z_k|$ then all nonzero z_i share an argument, and if additionally all $|z_i| = 1$ then they are all equal. That is the entire content of the character argument at Putnam 2011 A6.

Two sign lemmas complete the toolkit.

Lemma 7.7.5. *If every coefficient of P is nonnegative and the leading coefficient is positive, then $P(x) > 0$ for real $x > 0$, so P has no positive real root. Conversely, if F is monic real of degree at least one with every root in the open left half plane, then every coefficient of F is strictly positive.*

Proof. The first claim is immediate. For the second, factor over $\mathbb{R}$ into $x + c$ with $c > 0$ and $x^2 - 2\,\mathrm{Re}(\alpha)x + |\alpha|^2$ with $-2\,\mathrm{Re}(\alpha) > 0$; a product of polynomials with all coefficients positive has all coefficients positive. □

The converse of the second half is false: $x^4 + x^3 + x^2 + x + 1$ has positive coefficients and roots with positive real part. Know the counterexample, because it is the difference between a lemma and a superstition.

Example 7.7.6 (Putnam 2021 A6)**.** P has coefficients in $\{0, 1\}$ and factors as QR with both nonconstant; must $P(2)$ be composite? Yes. Normalise Q, R monic, their leading coefficients multiplying

to 1, so it suffices that $|Q(2)| \geq 2$ for every nonconstant monic integer factor. The sharpened Cauchy bounds give $\mathrm{Re}(\alpha) < \frac{3}{2}$ for every root of P, hence of Q. Shift by the midpoint of 1 and 2: $\widetilde{Q}(x) = Q(x + \frac{3}{2}) = \sum c_i x^i$ has all roots in the open left half plane, so every $c_i > 0$ and $Q(2) - Q(1) = 2\sum_{i \text{ odd}} c_i 2^{-i} > 0$. Finally $Q(1) \geq 1$, since $Q(1) \leq 0$ would give a real root of Q in $[1, \infty)$, contradicting that P has no positive real root. Hence $Q(2) \geq 2$.

Read the shape. The constant $\frac{3}{2}$ is not chosen, it is the midpoint of the two points compared, and the Taylor shift identity $F(b) - F(a) = 2\sum_{i \text{ odd}} c_i h^i$ with $c = \frac{a+b}{2}$, $h = \frac{b-a}{2}$ turns a comparison of values into a sign statement about recentred coefficients. Nothing was guessed.

Two rare items belong here. Descartes' rule of signs (the number of positive real roots with multiplicity is at most the number of sign changes in the coefficient list, and differs from it by an even number) appeared once, at Putnam 2022 A2, and only to be rejected in thirty seconds, because it constrains roots and not coefficients. Rouche's theorem (if $|g| < |f|$ on a closed contour then f and $f + g$ have the same number of zeros inside) is worth carrying as a citable fact but appears nowhere in the 2006 to 2025 corpus. Nor does Sturm's theorem on counting real roots by sign changes in the Sturm sequence; the only Sturm on the Putnam is the comparison theorem for differential equations at Putnam 2023 A3, a different subject. Learn Sturm sequences last, or not at all.

- **WHEN** coefficients are bounded and you need to keep roots away from a point **DO** apply the Cauchy bound, and if it is not sharp enough, redo its proof retaining more leading terms in a sector **BECAUSE** the bound is a geometric series estimate, so it improves as you retain terms, with threshold $t^m + t - 1 > 0$ (Putnam 2021 A6)
- **WHEN** coefficients are positive and monotone, or you need an annulus containing all roots **DO** quote Enestrom and Kakeya, and apply it to the reversal for the lower bound **BECAUSE** multiplying by $(x - 1)$ telescopes the monotone coefficients into a one line triangle inequality (Putnam 2014 A5, 2014 B4, 2018 B2, 2019 A3)
- **WHEN** two polynomials in a family must be shown coprime **DO** locate the roots of each in disjoint annuli, usually by splitting one as head plus x^j times tail **BECAUSE** coprimality over $\mathbb{Q}$ is exactly having no common complex root, so disjoint root windows finish it (Putnam 2014 A5)
- **WHEN** a statement about the roots of p' is wanted **DO** quote Gauss and Lucas and draw the convex hull of the roots of p **BECAUSE** the logarithmic derivative expresses a root of p' as a positive weighted average of the roots of p (Putnam 2018 B2)
- **WHEN** coefficients are in arithmetic progression, or differ in a simple pattern **DO** multiply by $(z - 1)$ and telescope **BECAUSE** shift and subtract cancels the common difference termwise, leaving a geometric series plus boundary terms (Putnam 2018 B2)
- **WHEN** you must compare $F(a)$ with $F(b)$ for a polynomial with root information **DO** shift to the midpoint $\frac{a+b}{2}$ and read the sign of the odd degree recentred coefficients **BECAUSE** $F(b) - F(a) = 2\sum_{i \text{ odd}} c_i h^i$, so root location in a half plane becomes a comparison of values (Putnam 2021 A6)

7.8 Real rootedness, log concavity and interlacing

To prove that a polynomial has all roots real you need a certificate, not a feeling. There is exactly one elementary certificate and you should reach for it every time.

Proposition 7.8.1 (Sign alternation certificate). *If a real polynomial f of degree n takes values of alternating sign at $n + 1$ points $x_0 > x_1 > \cdots > x_n$, then f has a root in each of the n disjoint intervals (x_j, x_{j-1}) by the intermediate value theorem, and since $\deg f = n$ these are all the roots, each simple and real.*

Everything else in such a problem exists only to make the $n+1$ evaluations computable.

Example 7.8.2 (Putnam 2014 B4)**.** Show all roots of $\sum_{k=0}^{n} 2^{k(n-k)}x^k$ are real. Here $a_k = 2^{k(n-k)}$ and a two line exponent computation gives $a_k^2 = 4a_{k-1}a_{k+1}$ exactly. That is the equality case of Hutchinson's theorem: if $a_k > 0$ and $a_k^2 \geq 4a_{k-1}a_{k+1}$ for all k then every zero is real and negative, and strict inequality (Kurtz) makes them simple. The constant 4 is the discriminant threshold of the local quadratic $a_{k-1} + a_k y + a_{k+1}y^2$, so it is not arbitrary. Knowing you are in the extremal case tells you in advance that the leading estimate will cancel exactly, which is the single hardest moment in the proof; without that prediction you would abandon the route. Enestrom and Kakeya independently confines all roots to the annulus $2^{1-n} \leq |z| \leq 2^{n-1}$ and cross checks the intervals, and the answer is that the roots are real, simple and negative, one in each interval $(-2^{2j-n}, -2^{2j-2-n})$.

Log concavity is the recurring shape here. A positive sequence with $a_k^2 \geq a_{k-1}a_{k+1}$ is log concave; log concavity is necessary for real rootedness (that is Newton's inequalities in disguise) and the factor 4 version is sufficient. Real nonpositive roots give the coefficient consequences directly: if p has positive leading coefficient and all roots in $(-\infty, 0]$ then $p = c\prod(x + s_i)$ with $c > 0$ and $s_i \geq 0$, so all coefficients are nonnegative, log concave and unimodal. That implication is exactly the mechanism of Putnam 2024 B5, where a combinatorial counting polynomial is shown to have nonnegative coefficients by locating its roots on the nonpositive reals. The converse fails, as $x^2 + x + 1$ shows.

Interlacing, the statement that the roots of p' separate those of p for real rooted p, and its generalisation to pairs of polynomials with a common interlacing, does not appear anywhere in the 2006 to 2025 corpus. Know the word, know that it is the main modern tool for real rootedness, and learn it last.

- **WHEN** a polynomial is asserted to have all roots real **DO** find $n+1$ points at which you can determine the sign, and prove the signs alternate **BECAUSE** sign alternation plus the intermediate value theorem plus the degree bound is a complete certificate, and it is the only elementary one (Putnam 2014 B4)
- **WHEN** the coefficients satisfy $a_k^2 \geq 4a_{k-1}a_{k+1}$ **DO** quote Hutchinson: all zeros are real and negative **BECAUSE** the factor 4 is the discriminant threshold of the local quadratic, and the extremal case predicts exact cancellation in your estimates (Putnam 2014 B4)
- **WHEN** a counting function must be shown to have nonnegative coefficients **DO** prove the interpolating polynomial has all roots real and nonpositive **BECAUSE** real nonpositive roots factor as $c\prod(x + s_i)$ with $s_i \geq 0$, forcing nonnegative and in fact log concave coefficients (Putnam 2024 B5)

7.9 Divisibility, coprimality and irreducibility over the integers

Theorem 7.9.1. *For a field k, $k[x]$ is Euclidean with respect to degree, hence a principal ideal domain and a unique factorisation domain. Consequently $\gcd(p, q) = 1$ if and only if $up + vq = 1$ for some $u, v \in k[x]$.*

Bezout is the usable form. Any identity $pU - qV = 1$ proves coprimality outright; Putnam 2010 B4's shift and subtract route needs exactly that to conclude $p \mid p(x+1) + p(x-1)$.

Proposition 7.9.2. *The Euclidean algorithm does not see the field it is run in, so $\gcd(f, g)$ computed in $\mathbb{Q}[x]$ equals the one computed in $\mathbb{C}[x]$. Hence $f, g \in \mathbb{Q}[x]$ are coprime if and only if they have no common root in $\mathbb{C}$.*

That is the bridge from algebra to geometry that Putnam 2014 A5 crosses. To descend from $\mathbb{Q}[x]$ to $\mathbb{Z}[x]$ you need one named theorem.

Theorem 7.9.3 (Gauss's lemma)**.** *The product of primitive polynomials in $\mathbb{Z}[x]$ is primitive (a polynomial is primitive when the gcd of its coefficients is 1). Consequently a primitive polynomial that factors over $\mathbb{Q}$ factors over $\mathbb{Z}$ into factors of the same degrees, and $\mathbb{Z}[x]$ is a unique factorisation domain.*

I do not reprove it; the proof is the observation that reduction modulo a prime p is a ring homomorphism $\mathbb{Z}[x] \to \mathbb{F}_p[x]$ and $\mathbb{F}_p[x]$ is a domain, so if p divided every coefficient of a product it would divide every coefficient of a factor. That single observation, evaluation and reduction commuting, is itself a circuit and appears at Putnam 2007 B1.

The rational root theorem is the degree one case: a rational root a/b in lowest terms of $\sum c_i x^i$ has $b \mid c_n$ and $a \mid c_0$. For a quadratic it reduces to a perfect square condition on the discriminant, which is the entire content of Putnam 2011 B2: the triples (p, q, r) of primes for which $px^2 + qx + r$ has a rational root are exactly those with $q^2 - 4pr$ a perfect square, and classifying those shows that the primes appearing in seven or more triples are 2 and 5.

Now the honest part. Eisenstein's criterion (if a prime π divides $a_0, \ldots, a_{n-1}$, does not divide a_n, and $\pi^2 \nmid a_0$, then the polynomial is irreducible over $\mathbb{Q}$) does not appear once in the 2006 to 2025 corpus. Nor does irreducibility by reduction modulo p (if $\bar{f}$ is irreducible of the same degree over $\mathbb{F}_p$, then f is irreducible over $\mathbb{Q}$). Both are standard and both are cheap to state, so install them by name, but they are not where the marks are. What does appear, at Putnam 2021 A6, is the neighbouring circle of ideas: if the coefficients of P lie in $\{0, \ldots, b-1\}$ then $P(b)$ is the integer whose base b digits are those coefficients, and Cohn's criterion (extended by Brillhart, Filaseta and Odlyzko) says that if $P(b)$ is prime for some base $b \geq 2$ then P is irreducible over $\mathbb{Z}$. Recognising the family tells you immediately that the proof must be root location plus a shift, not a congruence argument. That is what knowing a theorem by name buys you even when you do not use it.

- **WHEN** two polynomials over $\mathbb{Q}$ must be proved coprime **DO** show they have no common complex root **BECAUSE** the Euclidean algorithm is field independent, so coprimality over $\mathbb{Q}$ and over $\mathbb{C}$ are the same statement (Putnam 2014 A5)
- **WHEN** coprimality over $\mathbb{Q}$ is established and you need it over $\mathbb{Z}$ **DO** check primitivity, usually by a constant term of 1, and quote Gauss's lemma **BECAUSE** Gauss's lemma is the only legitimate descent from $\mathbb{Q}[x]$ to $\mathbb{Z}[x]$ (Putnam 2014 A5)
- **WHEN** an identity $pU - qV = 1$ appears, or you can engineer one **DO** declare p and q coprime at once **BECAUSE** Bezout in a Euclidean domain is an if and only if, so the identity is a certificate (Putnam 2010 B4)
- **WHEN** the coefficients of a polynomial are digits in some base **DO** translate values into digit strings and expect a root location proof, not a congruence proof **BECAUSE** Cohn's criterion places the problem in the irreducibility family whose proofs run through root location (Putnam 2021 A6)

7.10 Resultants and discriminants at Putnam depth

Definition 7.10.1. For $f = c\prod_i (x - \alpha_i)$ of degree m and $g = d\prod_j (x - \beta_j)$ of degree n, the resultant is

$$\operatorname{Res}(f, g) = c^n d^m \prod_{i,j} (\alpha_i - \beta_j) = c^n \prod_i g(\alpha_i) = (-1)^{mn} d^m \prod_j f(\beta_j).$$

It vanishes exactly when f and g have a common root. The discriminant of f is $\operatorname{disc}(f) = \frac{(-1)^{m(m-1)/2}}{c} \operatorname{Res}(f, f')$, which vanishes exactly when f has a repeated root.

At Putnam depth you almost never compute a resultant as a determinant. What you use is the middle expression: a product of a fixed polynomial's values over all roots of another polynomial is a single algebraic quantity, and if the second polynomial is one you know, that quantity is one evaluation.

Example 7.10.2 (Putnam 2015 A3)**.** Compute $\log_2 \prod_{a=1}^{2015} \prod_{b=1}^{2015} \left(1 + e^{2\pi i ab/2015}\right)$. Fix a and set $g = \gcd(a, 2015)$, $m = 2015/g$. As b runs over a full residue system, $e^{2\pi i ab/2015}$ runs over the mth roots of unity, each hit g times. Now use the cyclotomic factorisation $\prod_{\zeta^m=1}(x - \zeta) = x^m - 1$ at $x = -1$: $\prod_{\zeta^m=1}(1 + \zeta) = 1 - (-1)^m$, which is 2 for odd m. Since 2015 is odd every m is odd, so the inner product is 2^g and the answer is $\sum_{a=1}^{2015} \gcd(a, 2015)$. That sum is multiplicative with value $2p - 1$ at a prime p, so it equals $9 \cdot 25 \cdot 61 = 13725$.

Naming the general principle is what makes the move instant next time: a product of a fixed polynomial's values over all nth roots of unity is, up to a sign fixed by degree, the resultant with $x^n - 1$, and when the fixed polynomial is linear that resultant is a single evaluation of $x^n - 1$.

The discriminant at Putnam depth is nearly always the quadratic one, and two facts about it are worth having ready. A quadratic $Ax^2 + Bx + C$ has a repeated root exactly when $B^2 - 4AC = 0$, and its value at its own vertex $x = -B/(2A)$ is $-(B^2 - 4AC)/(4A)$. That pair carries both discriminant computations at Putnam 2007 A1, including the shortcut for evaluating one curve's quadratic at the other's vertex. It also identifies the threshold $p = 1/4$ at Putnam 2022 B5 as a discriminant rather than an accident: the palindromic Laurent polynomial $px + p/x + (1 - 2p)$ is nonnegative on the unit circle exactly when $pt^2 + (1 - 2p)t + p$ has real roots, that is when $1 - 4p \geq 0$. And over $\mathbb{F}_p$ for odd p, the number of roots of $ax^2 + bx + c$ is 2, 1 or 0 according as $b^2 - 4ac$ is a nonzero square, zero or a non square; that is exactly where the dichotomy in Putnam 2022 A3 comes from.

One methodological warning, from Putnam 2007 A1. Given $A = 0$ and $B = 0$, the derived equation $A - B = 0$ is implied but not equivalent to the pair. If you subtract two equations you have lost information, and you must keep one of the originals alongside. Write that down; it is a standard and expensive mistake.

- **WHEN** a product runs over all roots of a known polynomial **DO** recognise it as a resultant and evaluate the other polynomial at those roots, or evaluate the known polynomial at the roots of the other **BECAUSE** $\prod_i g(\alpha_i)$ is symmetric in the α_i, hence computable from the coefficients, and for $x^n - 1$ it is a single substitution (Putnam 2015 A3)
- **WHEN** a tangency, a repeated root or a threshold appears **DO** compute the discriminant and, for a quadratic, remember the vertex value is $-\operatorname{disc}/(4A)$ **BECAUSE** tangency is a repeated root, which is exactly the vanishing of the discriminant (Putnam 2007 A1, 2022 B5, 2022 A3)
- **WHEN** you are tempted to subtract two given equations **DO** keep one original alongside the difference **BECAUSE** $A - B = 0$ is implied by but not equivalent to the pair, so subtracting alone weakens the hypothesis (Putnam 2007 A1)

7.11 Polynomials over finite fields

Everything in this section is beyond A-level and all of it recurs.

The root bound over a field, applied to $\mathbb{F}_p$, is Lagrange's theorem on polynomial congruences: a nonzero polynomial of degree d over $\mathbb{F}_p$ has at most d roots. That, and nothing else, bounds the exceptional set at Putnam 2025 B5.

Theorem 7.11.1 (Frobenius)**.** *In a commutative ring of characteristic p, the map $a \mapsto a^p$ is a ring homomorphism; in particular $(a + b)^p = a^p + b^p$ and, iterating, $(a + b)^{p^m} = a^{p^m} + b^{p^m}$.*

Proof. $\binom{p}{i} \equiv 0 \pmod{p}$ for $0 < i < p$, since p divides the numerator $p!$ and not the denominator. □

This is the freshman's dream, and it is true. Three uses appear in the corpus and all three are worth memorising. Over $\mathbb{F}_p[x]$ it gives $(x - c)^p = x^p - c^p$, so $1 + x + \cdots + x^{p-1} = (x^p - 1)/(x - 1) = (x - 1)^{p-1}$: at Putnam 2019 A5 that collapses the problem to a single statement about the order of vanishing at 1, and the answer is $(p - 1)/2$. For a matrix M over $\mathbb{F}_p$ it gives that $M^{p^m} = I$ is equivalent to $(M - I)^{p^m} = 0$, that is M unipotent, which is the whole proof at Putnam 2012 A5.

And on the prime field itself $x^p = x$, so over $\mathbb{F}_2$ every element satisfies $x^2 = x$ and a quadratic form becomes a linear one, which is the second independent consequence needed at Putnam 2011 A4.

Alongside it, keep the order of vanishing as a valuation: $\operatorname{ord}_c(uv) = \operatorname{ord}_c(u) + \operatorname{ord}_c(v)$ and $\operatorname{ord}_c(u+v) \geq \min$, with divisibility by $(x-c)^n$ being exactly $\operatorname{ord}_c \geq n$. Tracking a single integer through a computation instead of manipulating polynomials is what makes the Putnam 2019 A5 solution half a page.

For the structure theory, the facts to cite by name: for every prime p and every $n \geq 1$ there is a field $\mathbb{F}_{p^n}$ with p^n elements, unique up to isomorphism (equivalently, an irreducible polynomial of degree n over $\mathbb{F}_p$), and $\mathbb{F}_{p^n}^{\times}$ is cyclic of order $p^n - 1$. At Putnam 2014 B5 that existence is the one genuine knowledge gap in the problem: without it there is no construction at all.

Putnam 2011 B6 is the showcase for the whole section. You must show $\sum_{k=0}^{p-1} k!\, n^k$ is not divisible by p for at least $(p+1)/2$ values of n. The coefficients $c_k = k!$ satisfy $c_k = k c_{k-1}$, and a first order coefficient recursion always converts into a first order differential relation on the generating polynomial: here $x^2 f'(x) + (x-1)f(x) + 1 = 0$ exactly over $\mathbb{F}_p$, because $p! \equiv 0$. Dually the truncated exponential $e_{p-1}(x) = \sum_{k \leq p-1} x^k/k!$ satisfies $e' = e - x^{p-1}/(p-1)!$, and Wilson's theorem turns that into $g' = g + x^{p-1}$, so at any root of g the derivative equals 1: the same nonzero constant at every root. Reversal transports the root count between f and g, and the multiple root criterion turns the constant derivative into a bound. Every tool in this chapter is present in one problem.

Chevalley and Warning, that a system of polynomials over $\mathbb{F}_p$ in more variables than the sum of the degrees has a number of common zeros divisible by p, is a beautiful theorem and does not appear anywhere in the 2006 to 2025 corpus. Install the statement, do not drill it.

- **WHEN** you are in characteristic p and a pth power or a pth root of unity appears **DO** apply Frobenius and expand $(a+b)^{p^m} = a^{p^m} + b^{p^m}$ **BECAUSE** the binomial coefficients in between are divisible by p, so the map is a ring homomorphism (Putnam 2019 A5, 2012 A5, 2011 A4)
- **WHEN** a divisibility by a power of $(x-c)$ is wanted **DO** track ord_c as a single integer through the whole computation **BECAUSE** order of vanishing is a valuation, additive on products and superadditive on sums, so you never manipulate the polynomials themselves (Putnam 2019 A5)
- **WHEN** coefficients satisfy a first order recursion such as $c_k = \lambda(k) c_{k-1}$ **DO** convert to a first order differential relation on the generating polynomial, with the truncation supplying the error term **BECAUSE** differentiation shifts and weights the coefficients, so a one step recursion becomes a one term differential equation (Putnam 2011 B6)
- **WHEN** multiplicity arguments are needed over $\mathbb{F}_p$ **DO** check that the multiplicities in play are at most $p-1$ before using iterated derivatives **BECAUSE** the mth derivative carries a factor $m!$, so it is blind beyond multiplicity $p-1$; use Hasse derivatives otherwise (Putnam 2019 A5, 2011 B6)

7.12 Chebyshev and the special families

Of all the named polynomial families, the Putnam reuses one constantly and the others essentially never. Learn Chebyshev properly.

Definition 7.12.1. $T_0 = 1$, $T_1 = y$, $T_{n+1}(y) = 2yT_n(y) - T_{n-1}(y)$. The second kind U_n satisfies the same recursion with $U_0 = 1$, $U_1 = 2y$.

Proposition 7.12.2. *$T_n(\cos\theta) = \cos(n\theta)$; $T_m(T_n(y)) = T_{mn}(y)$; and $2T_mT_n = T_{m+n} + T_{|m-n|}$.*

Proof. The first is induction on the product to sum identity $2\cos\theta\cos n\theta = \cos(n+1)\theta + \cos(n-1)\theta$, which is exactly the recursion. The second and third then follow by evaluating at $y = \cos\theta$ and using the identity theorem, since both sides are polynomials agreeing on the infinite set $[-1,1]$. □

Note how the identity theorem does the work in that last step: a trigonometric identity valid for real angles becomes an algebraic identity valid over $\mathbb{C}$. Keep the first few: 1, y, $2y^2-1$, $4y^3-3y$, $8y^4-8y^2+1$, $16y^5-20y^3+5y$. The characteristic equation of the recursion is $\lambda^2-2y\lambda+1=0$ with roots $e^{\pm i\theta}$ when $y=\cos\theta$, which is why the closed form is $\cos(n\arccos y)$ and why the substitution $y=z+1/z$ of the previous section is the same object.

Four Putnam uses, in ascending subtlety.

Multiple angle: Putnam 2018 A3 maximises $\sum_{i=1}^{10}\cos(3x_i)$ subject to $\sum\cos(x_i)=0$, and $\cos 3x = T_3(\cos x) = 4\cos^3 x - 3\cos x$ turns it into an optimisation on the box $[-1,1]^{10}$ with answer $480/49$. Putnam 2013 B2 expands $(2\cos 2\pi x+1)^2$ back into $\{1,\cos 2\pi x,\cos 4\pi x\}$ using T_2, giving maximum $f(0)=3$. Putnam 2013 A6 uses $\cos 2\alpha = 2\cos^2\alpha - 1$ to turn a symbol into a polynomial on the square, where the constraint $z,y\in[-1,1]$ is essential because the polynomial is negative outside it.

Recursion: Putnam 2018 B4 has $x_0=1$, $x_1=x_2=a$, $x_{n+1}=2x_nx_{n-1}-x_{n-2}$. Read $2x_nx_{n-1}$ as product to sum: with $x_n=\cos\theta_n$ the index arithmetic forces $\theta_{n+1}=\theta_n+\theta_{n-1}$, so $\theta_n=F_nb$ and $x_n=T_{F_n}(a)=\cos(F_nb)$. For $|a|>1$ the same computation runs with cosh, giving $|x_n|=\cosh(F_nt)\geq 1$ and no zero, so $x_n=0$ forces the compact branch and hence periodicity.

Nodes: the numbers $\cos\frac{j\pi}{N}$, $j=1,\dots,N-1$, are the interior extremal nodes of T_N, and $\cos\theta=(e^{i\theta}+e^{-i\theta})/2$ turns sums of their powers into sums of roots of unity, which evaluate by orthogonality. That is the source of the optimal configuration at Putnam 2022 A6, automatically strictly increasing and inside $(-1,1)$.

Extremal: $2^{1-n}T_n$ is the monic real polynomial of degree n of smallest supremum norm on $[-1,1]$, namely 2^{1-n}. That is the equioscillation theorem, stated without proof, and it is the standard reason a Chebyshev polynomial answers a minimax question. Do not assume it answers every extremal question: at Putnam 2016 A6 the smallest C with $\int_0^1|P|\leq C\max_{[0,1]}|P|$ over cubics with a root in $[0,1]$ is found by contact conditions instead. If Q has degree at most 3, $Q\leq\mu$ on $[0,1]$ and $Q(p)=\mu$ at an interior point, then $\mu-Q$ has a double root at p; one more root and one value determine Q, and $C=5/6$. Equality conditions in an extremal problem are contact conditions, and contact conditions are double roots.

Legendre polynomials, Hermite polynomials and the general theory of orthogonal polynomials do not appear in the 2006 to 2025 corpus in their own right; the only trace is the Hankel determinants and J-fractions at Putnam 2024 A6. Learn the three term recursion every orthogonal family satisfies and that its roots are real, simple and interlacing, and stop there.

- **WHEN** $\cos(n\theta)$ or a multiple angle appears alongside $\cos\theta$ **DO** substitute $y=\cos\theta$ and write $T_n(y)$, then work on the box $[-1,1]$ **BECAUSE** T_n is defined by exactly this identity, so it converts a trigonometric optimisation into a polynomial one with an explicit constraint set (Putnam 2018 A3, 2013 B2, 2013 A6)

- **WHEN** a nonlinear recursion contains $2x_nx_{n-k}$ **DO** set $x_n=\cos\theta_n$ (or cosh) and read off the induced linear recursion on the index **BECAUSE** the product to sum identity is exactly the recursion for T_n, so the nonlinearity is angle addition (Putnam 2018 B4)

- **WHEN** a configuration of points in $(-1,1)$ must make many power sums simultaneously nice **DO** try the Chebyshev nodes $\cos\frac{j\pi}{N}$ **BECAUSE** their powers become sums of roots of unity, which evaluate uniformly by orthogonality (Putnam 2022 A6)

- **WHEN** an extremal problem has an equality case at an interior point **DO** write down the contact conditions, which are double roots, and solve for the extremal polynomial **BECAUSE** a maximum attained in the interior forces $\mu-Q$ to have a repeated root there, and enough contact conditions determine Q (Putnam 2016 A6)

7.13 Complex numbers as geometry

A-level gave you modulus, argument, de Moivre, roots of unity and loci. What it did not give you is the habit of using $\mathbb{C}$ as the coordinate system for plane geometry, in which similarities are multiplication and the configuration becomes one algebraic equation.

The orientation preserving similarities are exactly $z \mapsto az + b$ with $a \neq 0$, forming a group; the reversing ones are $z \mapsto a\bar{z} + b$; rotation by θ about c is $z \mapsto c + e^{i\theta}(z - c)$. Everything follows without a single angle chase. The vertices of a regular n-gon with centre b, counterclockwise, are $v_k = b + a\omega^k$ with $\omega = e^{2\pi i/n}$, a rotation orbit, and that is what converts the geometric hypothesis of Putnam 2008 A5 into the single equation $Q(k) = \omega^k$. If z, w are opposite vertices of a square, the other two are $\frac{iz+w}{1+i}$ and $\frac{z+iw}{1+i}$, because the centre is $\frac{z+w}{2}$ and the half diagonals differ by multiplication by i; at Putnam 2019 B1 that turns "these four points form a square" into a divisibility test by $1 + i$ in the Gaussian integers, making the count $5n + 1$ immediate.

Three unit circle identities are worth carrying permanently.

Proposition 7.13.1. *If $|z| = 1$ and $z \neq 1$, then $|1 - z|^2 = 2 - (z + \bar{z}) = 2\,\mathrm{Re}(1 - z)$, so $u = 1 - z$ satisfies $\mathrm{Re}(u) = |u|^2/2$ and $\mathrm{Re}(1/u) = \frac{1}{2}$.*

Proposition 7.13.2. *For $\cos\phi \neq 0$, $1 + i\tan\phi = e^{i\phi}/\cos\phi$ and $1 - i\tan\phi = e^{-i\phi}/\cos\phi$, hence $\frac{1+i\tan\phi}{1-i\tan\phi} = e^{2i\phi}$.*

Proposition 7.13.3. *$\prod_{k=0}^{n-1}(X - \zeta^k) = X^n - 1$ with $\zeta = e^{2\pi i/n}$, equivalently $\prod_{k=0}^{n-1}(1 - w\zeta^k) = 1 - w^n$. Consequences: $\prod_{k=0}^{n-1} 2\sin(\phi + k\pi/n) = 2\sin(n\phi)$; for odd n, $\prod_{k=0}^{n-1} 2\cos(\phi + k\pi/n) = (-1)^{(n-1)/2} 2\cos(n\phi)$; hence $\prod_k \tan(\phi + k\pi/n) = (-1)^{(n-1)/2}\tan(n\phi)$, and by logarithmic differentiation $\sum_k \tan(\phi + k\pi/n) = n\tan(n\phi)$.*

The last proposition solves Putnam 2006 A5 outright: for $a_k = \tan(\theta + k\pi/n)$ the requested ratio is $n\tan(n\theta)$ over $(-1)^{(n-1)/2}\tan(n\theta)$, that is $(-1)^{(n-1)/2}n$. The main route uses the second proposition instead, converting the n tangents into n points on the unit circle and hence into the roots of $z^n = \omega$, producing a polynomial whose Vieta data is $e_1/e_n = c_{n-1}/c_0$ for odd n: only two coefficients ever need computing.

Definition 7.13.4. A Mobius transformation is $z \mapsto \frac{az+b}{cz+d}$ with $ad - bc \neq 0$, acting on the Riemann sphere. Composition corresponds to 2×2 matrix multiplication up to a scalar, so the order of such a map is the order of its matrix in the projective general linear group.

Proposition 7.13.5. *Mobius transformations send the family of circles and lines to itself, and send a circle to a line exactly when the circle passes through the pole. A Mobius map with two distinct fixed points is conjugate to $z \mapsto \lambda z$; with one fixed point, to $z \mapsto z + 1$.*

The circles to lines statement is what makes an inversion a forced move rather than a manoeuvre. At Putnam 2020 B5 the four constraints are $|z_j| = 1$, $z_j \neq 1$, and the excluded point 1 lies on the constraint circle; so $w = 1/(1 - z)$ maps $\{|z| = 1\} \setminus \{1\}$ bijectively onto the line $\mathrm{Re}\, w = \frac{1}{2}$ (the inverse $z = (w - 1)/w$ turns $|z| = 1$ into $|w - 1| = |w|$, the perpendicular bisector of 0 and 1) and linearises every hypothesis at once.

Example 7.13.6 (Putnam 2020 B5)**.** Show $3 - z_1 - z_2 - z_3 - z_4 + z_1z_2z_3z_4 \neq 0$ for $|z_j| = 1$, $z_j \neq 1$. Put $u_j = 1 - z_j$. Then $\sum(1 - u_j) = 4 - e_1(u)$ and $\prod(1 - u_j) = \sum_k (-1)^k e_k(u)$, so the expression equals $e_2(u) - e_3(u) + e_4(u)$: the constant and linear parts cancel and only the top three symmetric functions survive. Now pass to $v_j = 1/u_j$, which lie on the line $\mathrm{Re}\, v = \frac{1}{2}$ by the chord identity. The reversal law gives $(e_2 - e_3 + e_4)/e_4 = 1 - e_1(v) + e_2(v)$. Finally, $\mathrm{Re}\, v = \frac{1}{2}$ means $\bar{v} = 1 - v$, so $\prod_j(1 - v_j) = \overline{\prod_j v_j}$, that is $\sum_{m=0}^{4}(-1)^m e_m(v) = \overline{e_4(v)}$. Substituting, the claim becomes $\overline{e_4(v)} + e_3(v) - e_4(v) \neq 0$, four points on a line and a purely real linear condition. Every step was forced by the position of the excluded point.

Two more Mobius facts. The Cayley transform $C(x) = \frac{1+ix}{1-ix}$ maps $\mathbb{R} \cup \{\infty\}$ bijectively to the unit circle with $C(\tan\phi) = e^{2i\phi}$; the tangent addition map $T(x) = \frac{x+t}{1-tx}$ has fixed points $\pm i$ for every t, and C sends $i \mapsto 0$, $-i \mapsto \infty$, so CTC^{-1} is multiplication by a unimodular constant. That is why the substitution at Putnam 2006 A5 is forced: the family is generated by repeated angle addition, and this is the coordinate in which that action is multiplication. Second, the six maps x, $1/x$, $1-x$, $\frac{1}{1-x}$, $\frac{x-1}{x}$, $\frac{x}{x-1}$ form the anharmonic group of order six, isomorphic to S_3, and x, $\frac{x-1}{x}$, $\frac{1}{1-x}$ are its unique subgroup of order three, exactly the cycle at Putnam 2016 A3. The matrix of $\sigma(x) = 1 - 1/x$ has trace 1 and determinant 1, so its characteristic roots are primitive sixth roots of unity and σ has order three. That is why, not merely that.

- **WHEN** a plane geometry configuration involves rotation, similarity, regular polygons or squares **DO** put the points in $\mathbb{C}$ and write the configuration as one algebraic equation using $z \mapsto az + b$ **BECAUSE** orientation preserving similarities are exactly multiplication and translation, so geometry becomes algebra with no angle chasing (Putnam 2008 A5, 2019 B1)
- **WHEN** a constraint set is the unit circle with one point removed, and the removed point is on the circle **DO** invert at that point with a Mobius map **BECAUSE** a Mobius map takes a circle through its pole to a line, so every constraint becomes real linear at once (Putnam 2020 B5)
- **WHEN** a substitution has finite order, or a family is generated by repeated angle addition **DO** find the coordinate in which the action is multiplication, via the associated matrix and its eigenvalues **BECAUSE** Mobius composition is matrix multiplication up to scalars, so order and normal form are eigenvalue statements (Putnam 2016 A3, 2006 A5)
- **WHEN** tangents of angles in arithmetic progression appear **DO** apply $\frac{1+i\tan\phi}{1-i\tan\phi} = e^{2i\phi}$ to move them onto the unit circle, or quote the product and sum formulas **BECAUSE** the Cayley transform is the bijection between $\mathbb{R} \cup \{\infty\}$ and the circle, and it turns equally spaced angles into roots of $z^n = \omega$ (Putnam 2006 A5)
- **WHEN** a sum of unit vectors attains its maximum modulus **DO** invoke the equality case of the triangle inequality and conclude they are all equal **BECAUSE** equality forces a common argument, and unimodularity then forces equality (Putnam 2011 A6, 2018 B2)

7.14 Roots of unity: orthogonality, filters and cyclotomy

Theorem 7.14.1 (Orthogonality)**.** *Let ζ be a primitive Nth root of unity. Then $\sum_{\ell=0}^{N-1} \zeta^{\ell u}$ equals N if $N \mid u$ and 0 otherwise; hence $\sum_{\ell=1}^{N-1} \zeta^{\ell u} = -1$ for $u \not\equiv 0$.*

Proof. Either sum the geometric series, or use the group theoretic proof, which generalises: for a finite abelian group G and a homomorphism $\chi : G \to k^\times$, choose h with $\chi(h) \neq 1$; reindexing $g \mapsto hg$ gives $\Sigma = \chi(h)\Sigma$, so $\Sigma = 0$. When χ is trivial, $\Sigma = |G|$. □

The second proof is the one to keep, because it explains why the reindexing move is forced rather than lucky, and because the same identity underlies character sums and Gauss sums. It is the abstract statement used at Putnam 2021 A5 and, through the character theory of a finite abelian group, at Putnam 2011 A6.

Corollary 7.14.2 (Roots of unity filter)**.** *For $f(X) = \sum_m \alpha_m X^m$ and ζ a primitive kth root of unity,*

$$\sum_{m \equiv r \ (k)} \alpha_m = \frac{1}{k} \sum_{j=0}^{k-1} \zeta^{-jr} f(\zeta^j).$$

That is the standard device for extracting an arithmetic progression of coefficients, and it is what makes Putnam 2017 A5 tractable: the triple of winning probabilities for three players collapses

into the single complex number $P_n(\omega) = c_n + a_n\omega + b_n\omega^2$, with $f(1) = 1$ carrying the only other information. The associated geometric picture is worth carrying too: for nonnegative a, b, c summing to 1, the number $z = c + a\omega + b\omega^2$ has $a - b$, $c - a$, $b - c$ equal to $\frac{2}{\sqrt{3}}$ times the imaginary parts of z, ωz, $\omega^2 z$, so "player A leads" is "the argument of z lies in a fixed 120 degree sector", and cyclic relabelling is multiplication by ω. A statement about which of three numbers is largest became a statement about an argument, and the rotation argument then runs itself.

Cyclotomic factorisation is $x^n - 1 = \prod_{b=0}^{n-1}(x - \zeta^b)$ and, refined, $x^n - 1 = \prod_{d|n} \Phi_d(x)$ where Φ_d is the minimal polynomial of a primitive dth root of unity, of degree $\varphi(d)$ and with integer coefficients. Evaluating the coarse factorisation at a chosen point collapses a product over all nth roots of unity into one algebraic expression, as at Putnam 2015 A3. The polynomial $h_n(z) = 1 + z + \cdots + z^n = \frac{z^{n+1}-1}{z-1}$ has as roots exactly the $(n+1)$th roots of unity other than 1, all simple and unimodular; at Putnam 2018 B2 that identifies the convex hull for Gauss and Lucas. Substituting $w = \rho z$ multiplies the coefficient of z^k by ρ^k and divides every root by ρ, so a polynomial with coefficients in geometric progression has all its roots on one circle; that is exactly the extremal example at Putnam 2019 A3, where the constraints $1 \le b_0 < \cdots < b_{2019} \le 2019$ give $\mu \ge (\prod |z_i|)^{1/2019} = (b_0/b_{2019})^{1/2019} \ge 2019^{-1/2019}$ by AM-GM and Vieta, attained by $b_k = 2019^{k/2019}$.

Two further uses. Trigonometric sums in arithmetic progression telescope when multiplied by $2\sin$ of half the common difference: for $\sin x \neq 0$, $\sum_{n=0}^{N} \sin((2n+1)x) = \frac{\sin^2((N+1)x)}{\sin x}$, $\sum_{n=0}^{N} \cos((2n+1)x) = \frac{\sin((2N+2)x)}{2\sin x}$, and the Dirichlet kernel is $\frac{1}{2} + \sum_{n=1}^{N} \cos(nx) = \frac{\sin((N+\frac{1}{2})x)}{2\sin(x/2)}$. That closes both ingredients at Putnam 2020 A6, where the answer is $\pi/4$. Second, roots of unity as an alphabet: at Putnam 2011 B4 you need unimodular u, v with $\bar{u}v = i$, forcing $v = iu$, so up to a global phase the alphabet is $\{1, i\}$. A real alphabet such as $\{1, -1\}$ cannot do this, because $\bar{u}v$ is then symmetric in the pair; the choice is solved for, not guessed, and it explains why the real correlation matrix is structurally blind to one of the two statistics.

Finally, a finite nonempty subset of a group closed under the operation is a subgroup, so a finite multiplicatively closed subset of $\mathbb{C}$ is $\{0\}$, the group μ_m of mth roots of unity, or $\mu_m \cup \{0\}$. That is the one dimensional model at Putnam 2022 B2.

- **WHEN** you need the sum of coefficients in an arithmetic progression of indices, or three symmetric outcomes must be tracked together **DO** evaluate the generating function at the kth roots of unity and average with the filter **BECAUSE** orthogonality kills every residue class but one, so k evaluations recover k separate sums (Putnam 2017 A5)

- **WHEN** a sum runs over all Nth roots of unity **DO** apply orthogonality directly, and if a product runs over them, use $\prod(x - \zeta^b) = x^n - 1$ at a chosen point **BECAUSE** these two identities evaluate every uniform sum and product over μ_N in one line (Putnam 2022 A6, 2015 A3)

- **WHEN** coefficients are in geometric progression **DO** rescale z to make them constant and read off that all the roots lie on one circle **BECAUSE** substituting $w = \rho z$ scales coefficient k by ρ^k and divides all roots by ρ (Putnam 2019 A3)

- **WHEN** a trigonometric sum has arguments in arithmetic progression **DO** multiply by $2\sin$ of half the common difference and telescope **BECAUSE** the product to sum identity turns each term into a difference $g(n+1) - g(n)$ (Putnam 2020 A6)

- **WHEN** a two valued or three valued outcome must be encoded so that a bilinear form separates two statistics **DO** solve for the alphabet inside the unit circle rather than guessing **BECAUSE** the required relation $\bar{u}v = i$ is antisymmetric, which no real alphabet can produce (Putnam 2011 B4)

7.15 Complex packaging: two real objects as one

A recurring reroute deserves its own name. Whenever a problem carries two real quantities that transform together, package them as one complex quantity, because the packaging is faithful and the algebra collapses.

The degree version: for real f, g, the polynomial $P = f + ig$ has $\deg P = \max(\deg f, \deg g)$ exactly, because $f_j + ig_j = 0$ if and only if both real coefficients vanish. So "at least one of f, g has degree at least $n-1$" is exactly "$\deg P \geq n-1$", with no loss. This is what makes Putnam 2008 A5 a one variable problem.

The norm version: for a, b in any commutative ring containing $\mathbb{R}$, $a^2 + b^2 = (a+ib)(a-ib)$. This is the norm form $N(u) = u\sigma(u)$ of the quadratic extension by i, and it is multiplicative, which in coordinates is the Brahmagupta and Fibonacci identity $(a^2+b^2)(c^2+d^2) = (ac-bd)^2 + (ad+bc)^2 = (ac+bd)^2 + (ad-bc)^2$. It converts an additive equation into a multiplicative one, which is the only step that makes the count at Putnam 2007 B4 finite. Alongside it, know the instant obstruction: if $P(\rho)^2 + Q(\rho)^2 = 0$ for real ρ then both vanish, so $(X-\rho)^2$ divides $P^2 + Q^2$; hence a real polynomial that is a sum of two real squares is nonnegative on $\mathbb{R}$ with all real roots of even multiplicity, which rules out targets such as $X^{2n} - 1$ immediately.

The conjugation version: $\sigma\left(\sum a_k X^k\right) = \sum \overline{a_k} X^k$ is a ring automorphism of $\mathbb{C}[X]$ of order two whose fixed ring is exactly $\mathbb{R}[X]$, with $\sigma(X-\zeta) = X - \bar{\zeta}$. So $A \mapsto \left(\frac{A+\sigma A}{2}, \frac{A-\sigma A}{2i}\right)$ is a bijection onto ordered pairs of real polynomials, and the reality of P and Q becomes a symmetry condition on the root set of A.

The rank version: $e^{i(a+b)} = e^{ia}e^{ib}$, so a matrix whose (j,k) entry is a function of $a_j + b_k$ becomes a rank one complex outer product. At Putnam 2009 A3 that is the master idea: the cosine matrix has real rank at most two, not one, because taking a real part of a rank one complex matrix can double the rank, and hence the determinant vanishes for $n \geq 3$ and the limit is 0.

The Hermitian version: a symmetric statistic and an antisymmetric statistic can be packaged as $T + iW$ with T symmetric and W antisymmetric, making the combination Hermitian; that is the frame of Putnam 2011 B4.

- **WHEN** two real polynomials, two real sequences or two real statistics move together **DO** package them as one complex object $f + ig$ or $T + iW$ **BECAUSE** the packaging is faithful on degrees and coefficients, so nothing is lost and one variable replaces two (Putnam 2008 A5, 2011 B4, 2009 A3)
- **WHEN** an equation involves a sum of two squares **DO** factor it as a norm $(a+ib)(a-ib)$ and use multiplicativity **BECAUSE** a sum of two squares is the norm form of $\mathbb{Z}[i]$ or $\mathbb{R}[i]$, which converts addition into multiplication and makes unique factorisation available (Putnam 2007 B4, 2019 B1)

7.16 Functional equations: the substitution repertoire

A functional equation is not a topic, it is a protocol. The unknown is a function; the equation is a family of constraints indexed by the substitutions you are permitted to make. Your first act is never to guess the answer, it is to list the substitutions the domain allows and notice which collapse the equation. Guess second, and write the guess down: in almost every Putnam functional equation the solution family is recognisable in ten seconds, and knowing it tells you which branches of the proof are real.

Here is the repertoire, in the order you should try it.

Diagonal substitutions. If the equation is symmetric in x and y, set $y = x$; if a cyclic identity holds in three variables, set two equal first. At Putnam 2008 A1, from $f(x,y) + f(y,z) + f(z,x) = 0$, setting all three equal gives $3f(x,x) = 0$, then $z = x$ gives antisymmetry $f(x,y) + f(y,x) = 0$. Only then take a basepoint: $z = b$ gives $f(x,y) = f(x,b) - f(y,b)$, so $g(x) = f(x,b)$ works. Order matters,

and doing the basepoint first leaves an unresolved term. In general a function on pairs satisfying the cyclic zero identity is a coboundary, with potential unique up to an additive constant, exactly as antiderivatives are. At Putnam 2022 B6 the diagonal turns $f(xf(y)) + f(yf(x)) = 1 + f(x+y)$ into $2f(xf(x)) = 1 + f(2x)$, revealing that the true unknown is $h(x) = xf(x)$; once you have an auxiliary function, ask its limits at the ends of the domain, its image, and the values it omits.

Basepoint substitutions. Fix one variable at a distinguished value (0, 1, a fixed point) and see what one variable function drops out.

Cyclic substitutions of finite order. If $\sigma^k = \mathrm{id}$, writing the equation at $x, \sigma x, \ldots, \sigma^{k-1}x$ gives k linear equations in the k unknowns $f(\sigma^j x)$. For k odd the alternating sum around the cycle cancels everything except $2f(x)$; for k even it telescopes to zero and gives only a consistency condition, which is why odd order substitutions are the solvable ones. At Putnam 2016 A3, $\sigma(x) = 1 - 1/x$ has order three, so

$$2f(x) = \arctan x - \arctan \sigma(x) + \arctan \sigma^2(x),$$

and integrating gives $\int_0^1 f = 3\pi/8$.

Normalising by an explicit factor. For $f(x) = c(x)f(\varphi(x))$ with c, φ explicit, find an explicit h with $c(x)h(\varphi(x)) = h(x)$; then $g = f/h$ satisfies the bare invariance $g = g \circ \varphi$. This is the integrating factor idea transplanted. At Putnam 2012 A3, with $c(x) = \frac{2-x^2}{2}$ and $\varphi(x) = \frac{x^2}{2-x^2}$, the identity $1 - \varphi(x)^2 = \frac{4(1-x^2)}{(2-x^2)^2}$ shows $h(x) = \sqrt{1-x^2}$ is the correct normaliser, and the answer is $f(x) = \sqrt{1-x^2}$.

Fixed points of the iteration map. Check every fixed point of φ, not only the attracting one. At a fixed point t_0 the equation reads $f(t_0) = c(t_0)f(t_0)$, pinning $f(t_0) = 0$ whenever $c(t_0) \neq 1$, with no limiting argument. At Putnam 2012 A3, $\varphi(1) = 1$, so $x = \pm 1$ give $f(\pm 1) = 0$ algebraically and save a continuity argument at the boundary; the attracting fixed point 0 is the only place the iteration argument is actually needed.

Conjugation to a normal form. If the equation speaks in powers, take logarithms until it speaks in displacements. Setting $g(u) = \log f(e^u)$ turns the power map x^c into cu; setting $h(s) = \log g(e^s)$ turns that into the translation $s + \log c$. At Putnam 2016 B5 the hypothesis becomes chainable only after exactly two logarithms, when it reads "a displacement in $[\log 2, \log 3]$ maps to a displacement in $[\log 2, \log 3]$", and the answer is $f(x) = x^c$ for $c > 0$. The exponents 2 and 3 rather than 2 and 4 are load bearing: $\log 3/\log 2$ is irrational, whereas for 2 and 4 the function $h(s) = s + \delta \sin(2\pi s/\log 2)$ with $0 < \delta < \log 2/(2\pi)$ gives a non power solution. Always ask which hypothesis your proof spent.

Scaling symmetries. If f solving the equation implies $x \mapsto f(cx)$ does too, solutions come in one parameter families and no substitution can pin a value until you spend the symmetry by a normalisation. That is Putnam 2022 B6, where normalising the least solution of $f = \frac{1}{2}$ to lie at $x = 1$ is what makes the limits computable; the answer is $f(x) = \frac{1}{1+cx}$ for $c \geq 0$.

Geometric indexing. When a function on a space is constrained by a linear relation over every instance of a geometric figure, isolate one point value by finding a finite sub-family of instances whose relations, summed with integer coefficients, cancel every unknown but one. That is Putnam 2009 A1, where $f(A) + f(B) + f(C) + f(D) = 0$ over every square forces $f \equiv 0$; the bifurcation point is the decision to introduce midpoints and overlap the squares rather than grind a system with unfixed parameters.

Shared substitutions in coupled systems. When every equation of a coupled system has its nonlinear terms as multiples of the same full product of the unknowns, introduce that product as a new variable. Putnam 2009 A2 is the model: fgh satisfies its own equation and f falls out afterwards.

Injectivity, surjectivity and monotonicity as levers. Injectivity lets you cancel the unknown; surjectivity lets you choose arguments freely. For integer valued functions monotonicity transports bounds: $g(n+1) - g(n) \geq 1$ forces g strictly increasing with $g(n) \geq n$, so $g(u) \geq v$ gives $g(g(u)) \geq g(v)$. At Putnam 2025 B6 that transport applies eventual bounds at the argument $g(m)$, and integer exponents such as n^5 and n^4 avoid floors entirely; the answer is $r = \frac{1}{4}$.

Existence and uniqueness are two proofs. "Prove f is unique and find it" needs a derivation that at most one function works and a direct verification that the derived formula does. Without the second, the first proves nothing, because a vacuous hypothesis derives anything.

- **WHEN** a functional equation is symmetric in two variables, or cyclic in three **DO** substitute the diagonal first, before any basepoint **BECAUSE** collapsing variables extracts antisymmetry or vanishing on the diagonal, which later substitutions need (Putnam 2008 A1, 2022 B6)
- **WHEN** the equation relates $f(x)$ to $f(\sigma x)$ for an explicit substitution σ **DO** compute the order of σ; if it is odd, write the equation around the whole cycle and take the alternating sum **BECAUSE** for odd order the alternating sum telescopes to $2f(x)$; for even order it telescopes to zero and gives nothing (Putnam 2016 A3)
- **WHEN** the equation has the shape $f(x) = c(x)f(\varphi(x))$ with explicit c and φ **DO** hunt for an explicit h with $c(x)h(\varphi(x)) = h(x)$ and divide by it **BECAUSE** that reduces the equation to the bare invariance $g = g \circ \varphi$, which the iteration then solves (Putnam 2012 A3)
- **WHEN** an iterated map appears in a functional equation **DO** substitute every fixed point of the map, not only the attracting one **BECAUSE** at a fixed point the equation becomes algebraic and pins the value with no limit (Putnam 2012 A3)
- **WHEN** the hypothesis is stated in powers or in products **DO** take logarithms until it is stated in displacements **BECAUSE** two logarithms turn a power map into a translation, which is the only form in which such a condition chains (Putnam 2016 B5)
- **WHEN** the equation is invariant under $f(x) \mapsto f(cx)$ or a similar rescaling **DO** stop hunting for substitutions and normalise the symmetry away first **BECAUSE** a one parameter family of solutions means no substitution can determine a specific value (Putnam 2022 B6)
- **WHEN** a function on a space is constrained over every copy of a geometric figure **DO** find a finite integer combination of instances cancelling all unknowns but one **BECAUSE** the constraints are linear, so isolating one value is a linear algebra problem over the configurations (Putnam 2009 A1)
- **WHEN** a problem says prove the function is unique and find it **DO** write two separate proofs, at most one and at least one **BECAUSE** the derivation alone leaves open that no function exists, making the formula meaningless (Putnam 2012 A3)

7.17 Functional equations: regularity, and the classified families

Some functional equations are determined by algebra alone; some are not, and knowing which is which is the difference between a valid proof and a fantasy.

Theorem 7.17.1 (Cauchy)**.** *If $g : \mathbb{R} \to \mathbb{R}$ satisfies $g(x+y) = g(x) + g(y)$ and is additionally continuous at one point, or monotone on an interval, or bounded on a set of positive measure, then $g(x) = cx$. Without such a hypothesis there are pathological solutions, constructed from a Hamel basis of $\mathbb{R}$ over $\mathbb{Q}$ using the axiom of choice, whose graphs are dense in the plane.*

Proof sketch. Additivity gives $g(qx) = qg(x)$ for rational q, so g is $\mathbb{Q}$ linear and determined on $\mathbb{Q}$ by $c = g(1)$. Any of the regularity hypotheses forces the $\mathbb{Q}$ linear map to be $\mathbb{R}$ linear: continuity extends the identity from $\mathbb{Q}$ to $\mathbb{R}$ by density; monotonicity or local boundedness rules out the dense graph that a non linear $\mathbb{Q}$ linear map necessarily has. □

Know this precisely, because it is a trap in both directions. Sometimes regularity is not given and you must not assume it: Putnam 2008 A1 has enough substitution freedom to pin f by pure algebra, and reaching for continuity is both unnecessary and invalid. Sometimes regularity is given and is

exactly what makes the answer rigid: Putnam 2022 B6 says continuous, and the answer $\frac{1}{f} - 1 = cx$ is a Mobius twisted Cauchy equation. Equations mixing $f(xf(y))$ with $f(x+y)$ have Mobius or exponential answers, so you know the shape before any analysis.

Theorem 7.17.2 (Jensen). *A continuous Θ on $\mathbb{R}$ with $\Theta(a) + \Theta(b) = 2\Theta\left(\frac{a+b}{2}\right)$ for all a, b is affine.*

The proof is the dyadic subdivision recursion: an equation determining the midpoint value from the endpoints determines the function at every dyadic subdivision point of an interval, and continuity extends it to the interval. The same recursion handles Cauchy's equation, midpoint convexity implying convexity, and every midpoint type equation. At Putnam 2012 A6 the combined hypothesis says the antiderivative of f along a horizontal line satisfies Jensen's equation, so f is constant along that line; the answer is yes, $f \equiv 0$.

Theorem 7.17.3 (d'Alembert). *The continuous solutions of $f(A+B) + f(A-B) = 2f(A)f(B)$ for all real A, B are $f(x) = \cos(\lambda x)$, $f(x) = \cosh(\lambda x)$, $f \equiv 1$ and $f \equiv 0$.*

I cite this rather than prove it. What it buys is that the cosine substitution at Putnam 2018 B4 is forced, not lucky: any recursion encoding an addition law $\theta_{n+1} = \theta_n + \theta_{n-1}$ can be transported through any solution of the d'Alembert equation, the compact parameter (cosine) giving bounded orbits and the noncompact one (hyperbolic cosine) giving orbits escaping to infinity. The dichotomy in the problem is the dichotomy in the classification.

Three more classified situations, each of which appeared once and each of which you should recognise on sight.

Shift equations from sliding windows. For continuous g with antiderivative G, "the integral of g over every window of width T equals c" is exactly $G(z+T) = G(z) + c$, whose general solution is cz/T plus an arbitrary T periodic function; so g is T periodic with mean c/T. A single fixed width is therefore never a rigidity hypothesis: it has a large kernel. At Putnam 2019 A4, integrating over every unit sphere in $\mathbb{R}^3$ slices, by the Archimedes projection property, to exactly this with $T = 2$ and $c = 0$, so $f(x,y,z) = \sin(\pi z)$ is a counterexample and the answer is no. Contrast Putnam 2012 A6, where the unit area rectangles have varying widths and do force $f \equiv 0$. Count the degrees of freedom in the family of constraints before guessing which way the answer goes.

Rigidity from lattice preservation. A C^1 map with $f\left(\frac{1}{m}\mathbb{Z}\right) \subseteq \frac{1}{m}\mathbb{Z}$ for every $m \geq 1$ is affine with integer slope, and in $\mathbb{R}^n$ affine with integer matrix. The proof splits: an analytic half, that mesh secant slopes are integers at every level of the refinement $\frac{1}{m}\mathbb{Z} \subseteq \frac{1}{jm}\mathbb{Z}$, so the derivative is an integer constant; and an arithmetic half, that multiplication by k preserves exact denominators only when $\gcd(k,b) = 1$ for every b, that is $k = \pm 1$. At Putnam 2008 B5 the answer is $f(x) = \pm x + n$ for integers n.

Delay and iterative differential equations. Putnam 2010 A2 wants all differentiable f with $f'(x) = \frac{f(x+n)-f(x)}{n}$ for every positive integer n. Testing $f = e^{\lambda x}$ gives $\lambda = \frac{e^{\lambda n}-1}{n}$; $n = 1$ alone has infinitely many complex roots besides 0, and only $n = 1$ together with $n = 2$ forces $\lambda^2 = 0$. Exactly two witnesses are load bearing, and the answer is $f(x) = cx + d$. Putnam 2010 B5 asks whether a strictly increasing f can satisfy $f'(x) = f(f(x))$: recognise the type, since the unknown appears at an argument depending on the unknown. That is an iterative differential equation, the standard existence and uniqueness theorems do not apply, there is no separation of variables, and the handling is to derive monotonicity, regularity and growth and argue globally. The answer is no, and recognising the type is what stops you wasting the page.

Separation from a vanishing mixed partial. If g is C^2 on a rectangle with $g_{xy} \equiv 0$ then $g(x,y) = u(x) + v(y)$; with $xg_x + yg_y = 0$ this forces $xu'(x) = -yv'(y) = a$ constant, so $g = a\ln(x/y) + b$. That classifies the family at Putnam 2019 B4 and gives a second reason the answer is independent of f. The manufacturing step is symmetric functions of commuting operators: $2DE = (D+E)^2 - D^2 - E^2$, so knowing a sum and a sum of squares is knowing a product, which is why two symmetric equations determine the cross derivative.

- **WHEN** additivity $g(x+y) = g(x) + g(y)$ appears **DO** check whether continuity, monotonicity or boundedness is given; if it is, conclude $g(x) = cx$; if it is not, do not **BECAUSE** without regularity there are Hamel basis solutions with dense graph, so the conclusion is false as stated (Putnam 2022 B6, 2008 A1)
- **WHEN** a midpoint value is determined by the endpoint values **DO** run the dyadic subdivision and extend by continuity **BECAUSE** the equation determines the function on a dense set, and continuity does the rest (Putnam 2012 A6)
- **WHEN** $f(A+B) + f(A-B) = 2f(A)f(B)$, or a recursion transports an addition law $\theta_{n+1} = \theta_n + \theta_{n-1}$ **DO** quote d'Alembert and take the cosine branch or the hyperbolic cosine branch **BECAUSE** those are the only continuous solutions, so the substitution is forced and the two branches are the bounded and unbounded cases (Putnam 2018 B4)
- **WHEN** every window of fixed width has the same integral, or every sphere of fixed radius has the same mean **DO** convert to $G(z+T) = G(z) + c$ and expect a periodic kernel, so expect the answer no **BECAUSE** a single fixed width gives one shift equation with a large solution space; only a varying family forces rigidity (Putnam 2019 A4, contrast 2012 A6)
- **WHEN** the unknown function appears at an argument that depends on the unknown **DO** name it an iterative differential equation and stop looking for a closed form **BECAUSE** no standard existence or uniqueness theorem applies, so the contradiction must be asymptotic (Putnam 2010 B5)
- **WHEN** a functional equation is tested with $e^{\lambda x}$ and yields a transcendental characteristic equation **DO** look for a second instance of the hypothesis and intersect the root sets **BECAUSE** one delay equation has infinitely many characteristic roots; two independent ones can have only the trivial common root (Putnam 2010 A2)

7.18 The recursion file

Every recursive shape in this chapter reduces to one of the following. Learn the general form, the general solution and the recognition cue; then any particular instance is a substitution.

1. Constant coefficient linear recurrences and the shift operator. $\ker(E - tI)$ is spanned by t^k, and $\ker(E-I)^{d+1}$ is the polynomial sequences of degree at most d. In general u satisfies a constant coefficient recurrence with characteristic polynomial $\prod_i (x-\lambda_i)^{m_i}$ if and only if $u_k = \sum_i p_i(k)\lambda_i^k$ with $\deg p_i < m_i$. Cue: any relation between u_k and finitely many earlier terms with constant coefficients. Parent of the difference operator results, the Chebyshev recursion (characteristic equation $\lambda^2 - 2y\lambda + 1 = 0$, roots $e^{\pm i\theta}$) and the power sum recursion.

2. The power sum recursion. $s_j = e_1 s_{j-1} - e_2 s_{j-2} + \cdots$ is instance 1 with the polynomial itself as characteristic polynomial. Run it forwards for power sums from coefficients, Newton's first family for the reverse. For two quantities of known sum and product, $T_{k+1} = (D+E)T_k - DE\,T_{k-1}$; running it once in reverse recovers the product from T_1, T_2, the move at Putnam 2019 B4.

3. The elementary symmetric recursion. $e_k^{(n)} = e_k^{(n-1)} + u_n e_{k-1}^{(n-1)}$ with $e_0 = 1$, closed form $\prod_j (1 + xu_j) = \sum_k e_k x^k$. Cue: any symmetric function computation. Related: $R_n = \prod_{k=1}^{n} k(-k)$ satisfies $R_n = -n^2 R_{n-1}$, the source of the factorial at Putnam 2023 A2.

4. Difference annihilation and Newton's forward formula. $\Delta^m P = 0$ if and only if $\deg P < m$, and $P(x) = \sum_j \binom{x}{j} (\Delta^j P)(0)$. Instance 1 with root 1 of multiplicity m. Cue: block sums with digit weights, shift relations, interpolation at consecutive integers.

5. First order linear recursions. $u_{k+1} = Au_k + B$ solves as $u_k = A^{k-1}u_1 + B(1 + A + \cdots + A^{k-2})$, a fixed point plus a geometric term. Cue: anything of that shape, including $u_{k+1} = 1 + p'u_k$ with $u_1 = 1$, whose closed form governs iterates modulo the fixed point polynomial (Putnam 2024 A2), and the geometric sum $S_{m+1} = wS_m + 1$.

6. First order coefficient recursions become differential relations. $c_k = \lambda(k)c_{k-1}$ converts into a first order differential relation on the generating polynomial, the truncation supplying the error term: $c_k = c_{k-1}/k$ gives $e' = e$ minus the top term, $c_k = kc_{k-1}$ gives $x^2f' + (x-1)f + 1 = 0$ (Putnam 2011 B6).

7. Telescoping and summation by parts. Any sum whose terms are $g(n+1) - g(n)$ collapses. Polynomial form: $(z-1)\sum c_k z^k = \sum(c_{k-1} - c_k)z^k$ plus boundary terms (Putnam 2018 B2). Trigonometric form: the product to sum identities (Putnam 2020 A6). Orbit form: $f(x) - f(y) = (x-y)D(x,y)$ iterates to $b_{m+n} - b_n = (b_m - b_0)\prod_{i<n} D(b_{m+i}, b_i)$, the discrete derivative of an iterate (Putnam 2025 A6).

8. Chain rule along an orbit. $(f^{\circ n})'(x_0) = \prod_{i<n} f'(x_i)$ with $x_i = f^{\circ i}(x_0)$; for repeated squaring, $g'_{k+1}(0) = g'_k(g_k(0))g'_k(0)$. Item 7 in the limit $y \to x$ (Putnam 2025 A6).

9. Shortcut chord reduction. If a three term cyclic relation holds for all triples, the n term cyclic sum vanishes for every $n \geq 2$: apply the base identity to the last two edges to replace them by one chord, so $S(n) = S(n-1)$, and induct down to $S(2) = 0$ by antisymmetry. Cue: promoting any local three term relation along an arbitrarily long path (Putnam 2008 A1).

10. Dyadic subdivision. An equation giving the midpoint value from the endpoints determines the function on all dyadic points, and continuity extends it. Cue: Jensen, Cauchy, midpoint convexity (Putnam 2012 A6).

11. Shift equations. $G(z+T) = G(z) + c$ solves as cz/T plus a T periodic function; the discrete twin $u_{n+T} = u_n + c$ as cn/T plus a T periodic sequence. Cue: sliding window integral conditions (Putnam 2019 A4).

12. Refinement recursions. $\frac{1}{m}\mathbb{Z} \subseteq \frac{1}{jm}\mathbb{Z}$: a property holding at every level of a refinement passes to the limit object, here from mesh secant slopes to the derivative (Putnam 2008 B5).

13. Annihilating polynomial as a linear recurrence. If u equals one of $v_0, \ldots, v_{k-1}$ for each parameter value, then $\prod_j (u - v_j) = 0$ identically, so $u^k = \sum_{i<k}(-1)^{k-i-1}e_{k-i}(v)u^i$: a recurrence of order k on the powers of u, so every u^N with $N \geq k$ lies in the span of $1, u, \ldots, u^{k-1}$. Cue: a case analysis of variable length. That is all of Putnam 2007 B5, where $u = \lfloor n/k \rfloor$ takes one of the k values $(n-j)/k$ and the case analysis becomes one identity uniform in k.

14. Threshold families from truncation depth. Retaining m leading terms in the Cauchy bound proof gives the threshold polynomial $t^m + t - 1$ with positive root $\tau_m \uparrow 1$ and bound $|\alpha| < 1/\tau_m$, so one recursion in m generates the whole family (Putnam 2021 A6).

15. Conjugate pair doubling. Each further conjugate pair in a sum of two squares target doubles the number of representations, $c_{n+1} = 2c_n$ with $c_1 = 4$, since each pair is an independent binary choice; in general the count is $2\prod_j(e_j + 1)$ (Putnam 2007 B4).

16. Coefficient convolution triangularity. $b_k = 2a_0a_k + \sum_{i=1}^{k-1} a_i a_{k-i}$ determines a_k from $a_0, \ldots, a_{k-1}$ whenever $a_0 \neq 0$. Cue: any statement about the coefficients of p^2 (Putnam 2022 A2, answer $2n-2$, where $b_0 = a_0^2$, $b_1 = 2a_0a_1$, $b_{2n-1} = 2a_{n-1}a_n$, $b_{2n} = a_n^2$ carry the forced signs).

17. Degree growth from a dominant term. If one term of a recurrence has strictly larger degree than the other at every step, the degree grows by exactly that term's degree with no cancellation to check: $P_n = xP_{n-1} - P_{n-2}$ with $\deg P_0 = 0$, $\deg P_1 = 1$ gives $\deg P_n = n$ (Putnam 2017 A2).

7.19 What is rare, and what to learn last

Priority is not a mood, it is a count. Order your drilling by how often the corpus actually uses a thing.

Learn first, because they appear constantly: the identity theorem and the root bound; degree arguments in $k[x]$; Vieta at the two ends and elementary symmetric functions in general; the finite difference operator and its annihilation law; reversal; the factor theorem with root counting; Newton's identities; the formal derivative with its characteristic p caveat; Chebyshev; roots of unity orthogonality and the filter; the substitution repertoire for functional equations.

Learn second: Cauchy's root bound and its sharpening; Enestrom and Kakeya; the left half plane positivity lemma; Newton's forward formula and Lagrange interpolation; the Taylor congruence for integer polynomials; Frobenius and freshman's dream; Mobius transformations and their fixed point classification; the complex packaging moves; Cauchy's and Jensen's equations with the regularity discussion.

Learn last, and only once the above are automatic. Gauss and Lucas appeared twice, both times in an alternative solution (Putnam 2018 B2). Descartes' rule of signs appeared once, at Putnam 2022 A2, and only to be discarded. Hutchinson's log concavity threshold appeared once, at Putnam 2014 B4, though it was decisive there. The hypocycloid description of the set of sums of unimodular numbers with prescribed product appeared once, in the published second solution to Putnam 2020 B5. Theta functions and the Jacobi transformation appeared once, at Putnam 2024 B6, and there the useful content is negative: knowing the transformation exists is what stops you hunting for a closed form for a one sided weighted partial theta series, which has none, so the growth must be found by peak analysis.

And be honest about what is not there at all. In the 2006 to 2025 corpus there is no use of Eisenstein's criterion, no use of irreducibility by reduction modulo p, no use of Sturm sequences, no use of Rouche's theorem, no use of Chevalley and Warning, no use of Legendre or Hermite polynomials, and no use of interlacing. Every one of those is a real theorem and you should be able to state it if asked. None of them is worth an hour of drilling before the items above are reflexive. Knowledge is not the goal; the circuit firing in the arena is the goal, and a circuit that never fires is dead weight.

Chapter 8

Integrals, inequalities and optimisation

8.1 What this chapter is for

A-level gave you integration as a procedure: recognise the form, apply the rule, evaluate at the limits. Further Maths added a few named substitutions, reduction formulae for $\int \sin^n$, volumes of revolution about the two coordinate axes, and Maclaurin series. That is a toolkit for integrals that were designed to be done. The Putnam does not set integrals that were designed to be done. It sets integrals that were designed to be *rerouted*, and the reroute is almost always one of six structural moves: exploit a symmetry of the domain, introduce a parameter and differentiate with respect to it, turn a single integral into a double integral and swap the order, turn a sum into an integral or an integral into a sum, change variables and pay the Jacobian, or integrate by parts in the direction that kills a boundary term exactly.

On the other side, A-level gave you exactly one inequality with a name attached: the discriminant condition. Everything else you were allowed to do was completing the square. The Putnam expects a ladder of named inequalities, each with its equality case, because on the Putnam the equality case is usually where the marks are: the question asks for the *best* constant, and a bound without its equality case is half a solution.

By the end of this chapter you will read a definite integral and know within fifteen seconds which of the six moves it wants, you will be able to prove every inequality in the standard ladder in a paragraph each rather than quoting them as folklore, and you will know how to *design* the weights in a weighted inequality instead of guessing them, which is the difference between a bound and a sharp bound.

Two standing warnings. On the Putnam, "evaluate this integral" is rare; "prove this bound" and "find the largest constant such that" are common, so the two halves of this chapter are one subject, joined at the section on integral analogues. And every technique here has hypotheses: a solution that differentiates under the integral sign without saying why it may gets marked down, and you are not writing for partial credit.

8.2 Symmetry: the first thing you try on any definite integral

The reflection substitution

Lemma 8.2.1 (Reflection)**.** *For any integrable f on $[a, b]$,*

$$\int_a^b f(x)\,dx = \int_a^b f(a+b-x)\,dx.$$

Proof. Substitute $u = a + b - x$, so $du = -dx$, and the limits swap from $x = a, b$ to $u = b, a$. The two sign changes cancel. □

One line, and the single most productive move in Putnam integration, because of the corollary you actually use.

Corollary 8.2.2. $2\displaystyle\int_a^b f(x)\,dx = \int_a^b \big(f(x) + f(a+b-x)\big)\,dx.$

The point of interest falls out of the definition here in the same way the centre falls out of the definition of a circle. A definite integral is a number attached to a *set*, not to a formula, and the set $[a,b]$ has a symmetry: reflection in its midpoint. Any operation that preserves the set is an operation the integral cannot see. So if the integrand is ugly but the *sum* of the integrand and its reflection is clean, you never need the integrand.

Example 8.2.3 (Putnam 2016 A3)**.** f satisfies $f(x) + f\big(1 - \frac{1}{x}\big) = \arctan x$ for all real $x \neq 0$, and you want $\int_0^1 f$. The map $\sigma(x) = 1 - 1/x$ has $\sigma^3 = \mathrm{id}$, so writing the hypothesis at x, at $\sigma(x)$ and at $\sigma^2(x)$ gives three linear equations in the three unknowns $f(x), f(\sigma x), f(\sigma^2 x)$. Solving,

$$2f(x) = \arctan x + \arctan\frac{1}{1-x} - \arctan\frac{x-1}{x}.$$

Now do *not* integrate that. Add the reflected copy at $1-x$ and use $\arctan t + \arctan(1/t) = \pi/2$ for $t > 0$ three times, on the pairs $\{x, 1/x\}$, $\{1-x, 1/(1-x)\}$ and $\{(1-x)/x,\ x/(1-x)\}$:

$$2\big(f(x) + f(1-x)\big) = \tfrac{\pi}{2} + \tfrac{\pi}{2} + \tfrac{\pi}{2} = \tfrac{3\pi}{2}, \qquad \text{so} \qquad f(x) + f(1-x) = \tfrac{3\pi}{4}.$$

Hence $2\int_0^1 f = \int_0^1 \frac{3\pi}{4}\,dx$ and $\int_0^1 f = \frac{3\pi}{8}$. You never learned what f is at a single point, and you did not need to.

Symmetrisation over a group

Reflection is the case $|G| = 2$ of a general statement, and stating the general case is what lets you recognise the move when the symmetry is not a reflection.

Proposition 8.2.4 (Symmetrisation)**.** *Let a finite group G act on a region $D \subseteq \mathbb{R}^n$ by measure preserving bijections. For integrable F,*

$$\int_D F = \int_D F^G, \qquad F^G := \frac{1}{|G|}\sum_{\sigma\in G} F\circ\sigma.$$

Proof. Each σ is measure preserving and maps D onto D, so $\int_D F\circ\sigma = \int_D F$ by the change of variables formula with $|\det| = 1$. Average over $\sigma \in G$. □

This is the licence to replace an asymmetric integrand by its average whenever the *domain* is symmetric. Putnam 2021 A4 is the clean case: the region is a disc (symmetric in $x \leftrightarrow y$), the denominator $1 + x^4 + 6x^2y^2 + y^4$ is already symmetric, and the two numerators $1 + 2x^2$ and $1 + y^2$ are not. Symmetrising turns them into $1 + x^2 + y^2$ and $1 + \frac{1}{2}(x^2 + y^2)$, at which point the two integrands are visibly images of one another under a single linear map. That is the whole problem: the asymmetry was decoration.

Example 8.2.5 (Folding to remove an absolute value, Putnam 2006 A6)**.** If $g(u,v) = g(v,u)$ and h is odd, then over a domain symmetric in $u \leftrightarrow v$,

$$\iint g(u,v)\,|h(u-v)|\,du\,dv = 2\iint_{v<u} g(u,v)\,h(u-v)\,du\,dv.$$

The modulus is not removed by a case analysis; it is removed by folding the domain in half along the diagonal, where the two halves contribute equally because g is symmetric and $|h|$ is even. Case analysis on $\operatorname{sgn}(u-v)$ gets the same answer and costs three times the ink.

- **WHEN** you see a definite integral over $[a,b]$ whose integrand resists antidifferentiation **DO** write down $\int_a^b f(a+b-x)\,dx$ and add it to the original **BECAUSE** the integral depends on the set, not the formula, and the set is symmetric about its midpoint; if $f(x)+f(a+b-x)$ collapses you never need an antiderivative
- **WHEN** the integrand carries an absolute value of an odd function of $u-v$ over a symmetric domain **DO** fold the domain along $u=v$ and double **BECAUSE** the two halves are exchanged by a measure preserving involution under which the integrand is invariant, so a sign case split is doing work the symmetry already did
- **WHEN** a functional equation relates $f(x)$ to $f(\sigma(x))$ for a map σ of finite order **DO** write the equation at $x,\sigma x,\dots,\sigma^{k-1}x$ and solve the linear system **BECAUSE** a finite order map generates a finite cyclic group, so the orbit is finite and the system is square; then symmetrise before integrating
- **WHEN** the domain of a multiple integral is symmetric under swapping two variables but the integrand is not **DO** replace the integrand by its average under the swap **BECAUSE** the swap is a measure preserving bijection of the domain, so the integral cannot distinguish a function from its symmetrisation

8.3 Parameters: differentiating under the integral sign

The second move. You are handed a definite integral with no parameter, and you *insert* one, so that differentiating in the parameter produces an integral you can do, and integrating the answer back recovers what you wanted. This is Feynman's trick, and the reason it is not a trick is that it is the only systematic way to convert a hard integral into a differential equation, which is a strictly easier object.

Theorem 8.3.1 (Leibniz rule for differentiating under the integral sign)**.** *Let $\phi(x,t)$ and $\partial\phi/\partial t$ be continuous on $[a,b]\times[c,d]$. Then $F(t)=\int_a^b \phi(x,t)\,dx$ is differentiable on $[c,d]$ and*

$$F'(t)=\int_a^b \frac{\partial\phi}{\partial t}(x,t)\,dx.$$

Proof. For $h\neq 0$, the mean value theorem in t gives, for each x, a $\theta(x,h)$ between t and $t+h$ with $\big(\phi(x,t+h)-\phi(x,t)\big)/h=\phi_t(x,\theta)$. Hence

$$\left|\frac{F(t+h)-F(t)}{h}-\int_a^b \phi_t(x,t)\,dx\right|\le\int_a^b \big|\phi_t(x,\theta)-\phi_t(x,t)\big|\,dx.$$

The rectangle $[a,b]\times[c,d]$ is compact, so ϕ_t is uniformly continuous on it; given $\varepsilon>0$ choose δ so that $|\theta-t|<\delta$ forces the integrand below $\varepsilon/(b-a)$. Since $|\theta-t|\le|h|$, any $|h|<\delta$ works. □

Two things to notice, because they are exactly what a marker checks. The hypothesis is on ϕ_t, not on ϕ, and it is uniform continuity on a *compact* set that does the work. So on a compact interval with a continuous ϕ_t you may differentiate under the sign without comment. On an infinite interval you may not: there you need a dominating function, $|\phi_t(x,t)|\le g(x)$ with $\int_a^\infty g<\infty$, uniformly in t near the point of interest. Say which one you are using.

Theorem 8.3.2 (Variable limits)**.** *If in addition a,b are differentiable functions of t,*

$$\frac{d}{dt}\int_{a(t)}^{b(t)}\phi(x,t)\,dx=\phi(b(t),t)\,b'(t)-\phi(a(t),t)\,a'(t)+\int_{a(t)}^{b(t)}\phi_t(x,t)\,dx.$$

Proof. Apply the chain rule to $G(u,v,t)=\int_u^v\phi(x,t)\,dx$, whose partials in u and v are $-\phi(u,t)$ and $\phi(v,t)$ by the fundamental theorem of calculus, and whose partial in t is the previous theorem. □

The special case with no t inside is worth its own line because it fires constantly: for continuous g,

$$\frac{d}{ds}\int_s^{s+1} g(x)\,dx = g(s+1) - g(s),$$

so a sliding window integral is increasing exactly when g is increasing over the window. Putnam 2019 B4 uses precisely this to prove that $\Delta(s) = \frac{1}{2} + \int_s^{s+1} \ln x\,dx$ is strictly increasing on $s \geq 1$, which reduces a minimisation over a half line to an evaluation at the endpoint.

Putnam 2024 A5 is the plain use: a quantity $P(c) = \arccos\frac{c-1}{9} - \arccos\frac{c+1}{9}$ has to be shown monotone, and

$$P'(c) = \frac{1}{\sqrt{81-(c+1)^2}} - \frac{1}{\sqrt{81-(c-1)^2}},$$

whose sign is the sign of c, because $\frac{d}{dx}\arccos x = -1/\sqrt{1-x^2}$ and the larger denominator wins. Putnam 2009 A6 uses the theorem to justify that each of two integral equations is separately solvable, which is what locates the content of that problem in the *simultaneity*, not in either equation.

Introducing a parameter that was not there

Example 8.3.3 (The parameter is chosen so that the derivative is elementary)**.** To evaluate $\int_0^1 \frac{x^b-1}{\ln x}\,dx$ for $b > -1$, set $F(b)$ to be that integral. Then $\phi_b(x,b) = x^b$, continuous on $[0,1]\times[c,d]$ for $-1 < c \leq b \leq d$ after removing the removable singularity at $x = 1$, so

$$F'(b) = \int_0^1 x^b\,dx = \frac{1}{b+1}, \qquad F(0) = 0, \qquad F(b) = \ln(b+1).$$

The parameter was inserted at the exponent because that is the only place where differentiating destroys the $\ln x$ in the denominator: $\partial_b x^b = x^b \ln x$.

That last sentence is the reusable decision rule. You do not choose the parameter and hope. You look at the obstruction, ask which derivative annihilates it, and put the parameter where that derivative will act.

- **WHEN** a definite integral contains a factor you cannot antidifferentiate, typically $\ln x$, $1/x$, arctan, or an exponent you would like to vary **DO** insert a parameter at the position whose t derivative cancels that factor, differentiate under the sign, solve the resulting first order ODE, and fix the constant at a value of t where the integral is obvious **BECAUSE** differentiation in a parameter commutes with integration in x under continuity of ϕ_t on a compact rectangle, and an ODE is strictly easier than an integral
- **WHEN** the interval of integration is infinite and you want to differentiate under the sign **DO** produce an explicit dominating function g with $|\phi_t| \leq g$ and $\int g < \infty$ near the parameter value, and say so **BECAUSE** on a non compact domain uniform continuity is unavailable, and without domination the interchange can be false
- **WHEN** a quantity is defined as an integral over a window whose endpoints move **DO** differentiate with the full Leibniz formula and read the sign **BECAUSE** a monotone quantity attains its extremum at an endpoint of the parameter range, which converts an optimisation into an evaluation

8.4 The classical integrals you must own outright

These are not exercises. They are vocabulary. If you cannot produce any of the following in under two minutes in the arena, that is a knowledge gap, and knowledge gaps are named, not disguised.

The Gaussian integral, by three routes

Theorem 8.4.1. $\displaystyle\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}.$

Route 1, polar coordinates. Let I be the integral. By Tonelli (the integrand is positive, so no absolute integrability check is needed),

$$I^2 = \iint_{\mathbb{R}^2} e^{-(x^2+y^2)}\,dA = \int_0^{2\pi}\int_0^{\infty} e^{-r^2} r\,dr\,d\theta = 2\pi\cdot\tfrac{1}{2} = \pi.$$

The move that makes it work is not "use polar". It is: a product of two one dimensional integrals is a two dimensional integral, and a two dimensional integral is entitled to a change of variables that a one dimensional one is not. The factor r in $r\,dr\,d\theta$ is exactly what makes $e^{-r^2}r$ an exact derivative.

Route 2, a parameter and a differential equation. Put

$$F(t) = \left(\int_0^t e^{-x^2}dx\right)^2 + \int_0^1 \frac{e^{-t^2(1+u^2)}}{1+u^2}\,du.$$

Differentiating (the first term by the chain rule, the second under the integral sign, legitimate since the u interval is compact and the t derivative is continuous),

$$F'(t) = 2e^{-t^2}\int_0^t e^{-x^2}dx - 2t\,e^{-t^2}\int_0^1 e^{-t^2u^2}\,du = 0,$$

the second term becoming the first after $x = tu$. So F is constant. At $t = 0$, $F = \int_0^1 du/(1+u^2) = \pi/4$. As $t\to\infty$ the second term vanishes, so $(I/2)^2 = \pi/4$.

Route 3, Gamma. Substituting $s = x^2$,

$$\int_0^{\infty} e^{-x^2}dx = \tfrac{1}{2}\int_0^{\infty} s^{-1/2}e^{-s}\,ds = \tfrac{1}{2}\Gamma\left(\tfrac{1}{2}\right),$$

and $\Gamma(1/2) = \sqrt{\pi}$ is proved in the next section from the Beta integral. Learn all three: Route 1 when the problem is already two dimensional, Route 2 when the problem already has a parameter, Route 3 when the exponent is not 2.

Putnam 2024 B6 is the Gaussian appearing in disguise. Completing the square in an exponent,

$$bn - \frac{n^2}{t} = \frac{b^2t}{4} - \frac{1}{t}\left(n - \frac{bt}{2}\right)^2,$$

so $e^{bn}x^{n^2}$ with $\ln x \approx -1/t$ is a Gaussian in n, centred at $bt/2$, of height $e^{b^2t/4}$ and width $\sqrt{t}$. Once you see it as a Gaussian bump you know the sum over n is the height times the width times a constant, which is the entire asymptotic.

Dirichlet's integral

Theorem 8.4.2. $\displaystyle\int_0^{\infty}\frac{\sin x}{x}\,dx = \frac{\pi}{2}$, *the integral being improper at infinity and conditionally convergent.*

Proof. For $t \geq 0$ set $F(t) = \int_0^{\infty} e^{-tx}\dfrac{\sin x}{x}\,dx$. For $t > 0$ the t derivative $-e^{-tx}\sin x$ is dominated on $[t_0,\infty)$ by e^{-t_0x}, which is integrable, so

$$F'(t) = -\int_0^{\infty} e^{-tx}\sin x\,dx = -\frac{1}{1+t^2}$$

by two integrations by parts. Hence $F(t) = C - \arctan t$. Since $|F(t)| \leq \int_0^{\infty} e^{-tx}dx = 1/t \to 0$, we get $C = \pi/2$. Finally F is continuous at $t = 0$: the integral $\int_0^{\infty} e^{-tx}\frac{\sin x}{x}dx$ converges uniformly for $t \geq 0$ by Dirichlet's test, because $\int_0^R \sin x\,dx$ is bounded and e^{-tx}/x decreases to 0 uniformly in t. So $F(0) = \pi/2$. □

Do not skip the uniform convergence sentence. Without it the argument proves nothing about $t = 0$, which is the only value you wanted.

Frullani's integral

Theorem 8.4.3 (Frullani)**.** *Let f be continuous on $[0,\infty)$ with $L = \lim_{x\to\infty} f(x)$ existing and finite, and let $a, b > 0$. Then*

$$\int_0^\infty \frac{f(ax) - f(bx)}{x}\,dx = \big(f(0) - L\big)\ln\frac{b}{a}.$$

Proof. For $0 < \varepsilon < R$, substituting $u = ax$ and $u = bx$,

$$\int_\varepsilon^R \frac{f(ax) - f(bx)}{x}dx = \int_{a\varepsilon}^{aR} \frac{f(u)}{u}du - \int_{b\varepsilon}^{bR} \frac{f(u)}{u}du = \int_{a\varepsilon}^{b\varepsilon} \frac{f(u)}{u}du - \int_{aR}^{bR} \frac{f(u)}{u}du.$$

By the mean value theorem for integrals with the positive weight $1/u$, the first term equals $f(\xi_\varepsilon)\ln(b/a)$ with ξ_ε between $a\varepsilon$ and $b\varepsilon$, hence tends to $f(0)\ln(b/a)$; the second tends to $L\ln(b/a)$ because $f(u) \to L$. □

Frullani has not appeared in the filtered corpus. Learn it anyway, but learn it last among the classical integrals, and learn it for its *shape*: any integrand of the form $\big(f(ax) - f(bx)\big)/x$ is a Frullani cue, and the answer depends only on the two endpoint values of f, never on f in between. That last clause is the part worth carrying: a scale invariant difference cannot see the interior.

- **WHEN** you meet $\int_0^\infty e^{-\alpha x^2}$ or any integrand that is a Gaussian after completing the square in the exponent **DO** complete the square, read off centre, height and width, and quote $\int e^{-u^2}du = \sqrt{\pi}$ **BECAUSE** a Gaussian is characterised by centre and width, and its total mass is height times width times $\sqrt{\pi}$, so an asymptotic in a parameter is read off without any integration
- **WHEN** the integrand is $\sin x/x$, or any oscillatory kernel over a non compact interval **DO** insert the damping factor e^{-tx}, differentiate in t, and recover $t = 0$ by a uniform convergence argument **BECAUSE** damping makes the integral absolutely convergent so Leibniz applies, and Dirichlet's test transports the answer back to the undamped case
- **WHEN** the integrand is $(f(ax) - f(bx))/x$ **DO** quote Frullani **BECAUSE** the substitution $u = ax$ turns the difference into an integral over a shrinking and an expanding window, each of which sees only one endpoint value of f

8.5 Beta, Gamma and reduction formulae

Definition 8.5.1. For $s > 0$, $\Gamma(s) = \displaystyle\int_0^\infty t^{s-1}e^{-t}\,dt$. For $a, b > 0$, $B(a,b) = \displaystyle\int_0^1 t^{a-1}(1-t)^{b-1}\,dt$.

The point of Γ is not that it interpolates the factorial. It is that it is the unique natural object whose defining integral is invariant under the scaling $t \mapsto \lambda t$ up to a power of λ, which is why every integral with a power times a decaying exponential reduces to it.

Proposition 8.5.2 (Functional equation)**.** $\Gamma(s+1) = s\,\Gamma(s)$*, and* $\Gamma(n+1) = n!$ *for* $n \in \mathbb{N}$.

Proof. Integrate by parts: $\int_0^\infty t^s e^{-t}dt = \big[-t^s e^{-t}\big]_0^\infty + s\int_0^\infty t^{s-1}e^{-t}dt$, and the boundary term vanishes at both ends for $s > 0$. Then $\Gamma(1) = 1$ and induct. □

Theorem 8.5.3 (Beta and Gamma)**.** $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Proof. By Tonelli on the positive integrand,

$$\Gamma(a)\Gamma(b) = \int_0^\infty\int_0^\infty x^{a-1}y^{b-1}e^{-(x+y)}\,dx\,dy.$$

Change variables to $x = st$, $y = s(1-t)$ with $s > 0$, $t \in (0,1)$; this is a bijection onto the open quadrant with Jacobian

$$\det \begin{pmatrix} t & s \\ 1-t & -s \end{pmatrix} = -s, \qquad |\det| = s.$$

The integral becomes $\int_0^1 \int_0^\infty (st)^{a-1}(s(1-t))^{b-1} e^{-s} s \, ds \, dt = \Gamma(a+b)B(a,b)$. □

Corollary 8.5.4. $\Gamma(1/2) = \sqrt{\pi}$.

Proof. $B(1/2,1/2) = \int_0^1 t^{-1/2}(1-t)^{-1/2} dt = 2\int_0^{\pi/2} d\theta = \pi$ under $t = \sin^2\theta$, and $B(1/2,1/2) = \Gamma(1/2)^2/\Gamma(1)$. □

The substitution $t = \sin^2\theta$ deserves its own line, because it is the bridge between the Beta integral and every trigonometric power integral you will ever meet:

$$B(a,b) = 2\int_0^{\pi/2} \sin^{2a-1}\theta \, \cos^{2b-1}\theta \, d\theta.$$

Taking $b = 1/2$ gives, for $n \geq 0$,

$$\int_0^{\pi/2} \sin^n \theta \, d\theta = \frac{1}{2} B\left(\frac{n+1}{2}, \frac{1}{2}\right) = \frac{\Gamma\left(\frac{n+1}{2}\right)\sqrt{\pi}}{2\,\Gamma\left(\frac{n}{2}+1\right)}.$$

Putnam 2006 A6 needs exactly the even case, and it is worth memorising in the binomial form

$$\int_0^{\pi} \sin^{2m}\theta \, d\theta = \pi \binom{2m}{m} 4^{-m},$$

giving $3\pi/8$, $5\pi/16$, $35\pi/128$ for $m = 2,3,4$. Putnam 2009 B4 needs the two variable version: the circle moment $\frac{1}{2\pi}\int_0^{2\pi} \cos^a\theta \sin^b\theta \, d\theta$ vanishes unless a and b are both even, and the parity half of that statement is what kills every odd degree condition in that problem. Putnam 2022 A4 needs only the integer case: $\int_0^1 t(1-t)^{i-1} dt = B(2,i) = \frac{1!\,(i-1)!}{(i+1)!} = \frac{1}{i(i+1)}$.

Reduction formulae as recursions

A reduction formula is a recursion, and you should treat it as one: state it, solve it, and then recognise the family.

Proposition 8.5.5 (Wallis reduction). *$I_k = \int_0^{\pi/2} \cos^k\theta \, d\theta$ satisfies $I_k = \dfrac{k-1}{k} I_{k-2}$ for $k \geq 2$, with $I_0 = \pi/2$, $I_1 = 1$.*

Proof. Write $\cos^k = \cos^{k-1} \cdot \cos$ and integrate by parts with $u = \cos^{k-1}$: the boundary term vanishes and $I_k = (k-1)\int_0^{\pi/2} \cos^{k-2}\sin^2 = (k-1)(I_{k-2} - I_k)$. □

Solving the recursion by iterating gives the double factorial ratio, which is the binomial expression above. The general lesson, and this is the standing recursion order in action: whenever integration by parts returns the same integral with the exponent lowered by a fixed amount d, you have a linear recursion in the exponent with rational coefficients; solve it by iterating down to the d base cases, and the answer is always a ratio of products of arithmetic progressions, that is, a ratio of Gamma functions.

- **WHEN** the integrand is a power of t times a power of $1-t$ on $[0,1]$, or a product of powers of sin and cos on $[0,\pi/2]$ **DO** write it as a Beta function and convert to Gammas **BECAUSE** the Beta integral is the unique scaling invariant two parameter family containing both, so the answer is forced to be a ratio of factorials with no computation

- **WHEN** the integrand is a power times a decaying exponential on $[0,\infty)$ **DO** substitute to reach $\int t^{s-1}e^{-t}dt$ and quote Γ **BECAUSE** Γ is defined by exactly that shape, and the functional equation $\Gamma(s+1)=s\Gamma(s)$ then does all the bookkeeping
- **WHEN** integrating by parts returns the original integral with the exponent lowered **DO** stop computing, write the reduction as a recursion $I_k = r(k)I_{k-d}$, and solve it **BECAUSE** a linear recursion with rational coefficient is solved by iteration in one line, whereas continuing to integrate by parts is zombie maths

8.6 Substitutions that always terminate

Most substitutions are opportunistic. Three are algorithmic, which means you may reach for them without thinking, and that is exactly what you want under time pressure.

The Weierstrass substitution

Proposition 8.6.1. *With $t=\tan(\theta/2)$ on $\theta\in(-\pi,\pi)$,*

$$\sin\theta=\frac{2t}{1+t^2},\qquad \cos\theta=\frac{1-t^2}{1+t^2},\qquad d\theta=\frac{2\,dt}{1+t^2}.$$

Consequently every integral of a rational function of $\sin\theta$ *and* $\cos\theta$ *becomes an integral of a rational function of t, and hence terminates in partial fractions.*

Two standard evaluations follow, both of which appear in Putnam 2021 A4:

$$\int_0^{\pi/2}\frac{d\phi}{a\cos^2\phi+b\sin^2\phi}=\frac{\pi}{2\sqrt{ab}}\quad(a,b>0),\qquad \int_0^{2\pi}\frac{d\theta}{a+b\cos\theta}=\frac{2\pi}{\sqrt{a^2-b^2}}\quad(a>|b|).$$

The first is faster with $t=\tan\phi$ than with the half angle: dividing top and bottom by $\cos^2\phi$ turns the integrand into $\sec^2\phi\,d\phi/(a+b\tan^2\phi)=dt/(a+bt^2)$, and $\int_0^\infty dt/(a+bt^2)=\pi/(2\sqrt{ab})$. That is the reusable point: when the integrand is even in both sin and cos, use $t=\tan\phi$, which halves the degree; use the half angle only when odd powers are present.

The substitution $u=r^2$ and its relatives

If an integrand is $p(r^2)\,r\,dr$ for a rational p, then $u=r^2$ turns it into $\frac{1}{2}p(u)\,du$. This is the reason polar coordinates are so effective on radially even integrands: the Jacobian r is precisely the factor that the substitution consumes. Putnam 2021 A4 lands on

$$\int\frac{du}{a+cu^2}=\frac{1}{\sqrt{ac}}\arctan\!\left(u\sqrt{c/a}\right)+C,\qquad \int\frac{u\,du}{a+cu^2}=\frac{1}{2c}\ln(a+cu^2)+C$$

after exactly this move.

Recognising an exact derivative

The third algorithmic move is not a substitution but a pattern.

Lemma 8.6.2. *If $\Phi'=B$ then $B(f(x))f'(x)=\dfrac{d}{dx}\Phi(f(x))$. Hence $f'=A(x)/B(f)$ becomes $\dfrac{d}{dx}\Phi(f)=A(x)$.*

That is the chain rule read backwards, and it is what makes separable differential equations work, but its Putnam use is sharper than that. Putnam 2009 B5 has an equation that is not solvable in closed form; multiplying through by $1+f^2$ turns it into $g' = 1 - f^2/x^2$ with $g = f + f^3/3$, and the whole solution rests on that single move. The cue was that the awkward factor $1+f^2$ is Φ' for $\Phi(y) = y + y^3/3$.

The same lemma in its simplest form is worth naming separately because it appears as an endgame: $\int_a^b 2g(u)g'(u)\,du = g(b)^2 - g(a)^2$. Putnam 2015 B6 ends with

$$\int_0^1 \frac{2\arctan u}{1+u^2}\,du = \arctan(1)^2 = \frac{\pi^2}{16},$$

and recognising the pattern *in advance* explains why the answer is the square of the Leibniz constant rather than some unrelated number. Two more you should have automatic: $\int \tan u\,du = -\ln|\cos u| + C$ and $\int \cot u\,du = \ln|\sin u| + C$, both instances of $\int f'/f = \ln|f|$, and both needed in Putnam 2009 A2 to integrate $f'/f = 2\tan u + \cot u$.

- **WHEN** the integrand is a rational function of sin and cos **DO** substitute $t = \tan(\theta/2)$, or $t = \tan\theta$ if the integrand is even in both **BECAUSE** it rationalises the integrand, and every rational function has an elementary antiderivative by partial fractions, so the route is guaranteed to terminate
- **WHEN** a radial integrand carries an odd power of r times a function of r^2 **DO** substitute $u = r^2$ **BECAUSE** the polar Jacobian r is exactly $\frac{1}{2}\,du/dr$, so the substitution consumes it and leaves a rational integrand
- **WHEN** a differential equation has an awkward factor $B(f)$ multiplying f' **DO** antidifferentiate B to get Φ and rewrite the left side as $\frac{d}{dx}\Phi(f)$ **BECAUSE** the chain rule read backwards converts a nonlinear equation into an exact derivative, after which one integration finishes it
- **WHEN** you see g' multiplied by a function of g anywhere in an integrand **DO** name the antiderivative of that function and finish in one line **BECAUSE** the integrand is already an exact derivative, and continuing to look for a substitution is doing the work twice

8.7 Sums as integrals and integrals as sums

Riemann sums, used forwards and backwards

Proposition 8.7.1. *If f is Riemann integrable on $[0,1]$ then $\frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k}{n}\right) \to \int_0^1 f$.*

Forwards this is A-level. Backwards it is the move you need: a limit of a sum in which $1/n$ appears as a factor and k/n appears as an argument is an integral, and you evaluate it by evaluating the integral. The cue is the pair $(1/n,\ k/n)$, not the word "limit".

The refinement that Putnam 2009 B2 needs is the monotone version, and you should carry it separately.

Lemma 8.7.2. *If f is strictly increasing on $[a,b]$ and $a = x_0 < x_1 < \cdots < x_m = b$, then*

$$\sum_{i=1}^{m} f(x_i)(x_i - x_{i-1}) > \int_a^b f,$$

with the excess tending to 0 as the mesh tends to 0. For strictly decreasing f the right endpoint sum is a strict underestimate.

Proof. On each subinterval $f(x) < f(x_i)$ except at the right endpoint, so each term strictly exceeds the corresponding piece of the integral. The excess is bounded by $\max_i(x_i - x_{i-1}) \cdot (f(b) - f(a))$ because the excesses telescope inside a staircase of total height $f(b) - f(a)$. □

That lemma supplies both halves of a sharp constant: the strict inequality gives the bound, and the vanishing of the excess gives that the bound is approached but not attained. Putnam 2009 B2 is exactly that shape, and the equality analysis (which configurations attain the bound) is what upgrades a bound into a characterisation.

Reciprocals as integrals

This is the single highest yield conversion in the whole chapter, and it appears three times in the filtered corpus.

Lemma 8.7.3. *For $m > 0$, $\dfrac{1}{m} = \displaystyle\int_0^1 t^{m-1}\,dt$; equivalently $\dfrac{1}{t+1} = \displaystyle\int_0^1 x^t\,dx$ for $t \geq 0$.*

Why this is worth so much: a sum with denominators is not summable term by term, but once each denominator becomes an integral you may swap the sum and the integral (Fubini or monotone convergence), and the resulting sum inside the integral is a *geometric or power series*, which does have a closed form. You have traded an intractable sum for a tractable integral.

Example 8.7.4 (Putnam 2015 B6)**.** The alternating harmonic tail $T_\ell = \sum_{m\geq\ell}(-1)^{m-1}/m$ has no elementary closed form as a sum. Writing $1/m = \int_0^1 t^{m-1}dt$ and summing the geometric series inside,

$$T_\ell = (-1)^{\ell-1}\int_0^1 \frac{t^{\ell-1}}{1+t}\,dt,$$

after which the outer sum becomes $\int_0^1 C(-t)/(1+t)\,dt$ where C is the generating function of the outer weights. Changing the weights changes only C. In that problem the weights produce the Gregory series, $\sum_{\ell\geq1}(-1)^{\ell-1}t^{\ell-1}/(2\ell-1) = \arctan(\sqrt{t})/\sqrt{t}$, and the substitution $t = u^2$ turns $dt/(1+t)$ into $2u\,du/(1+u^2)$, cancelling the $1/u$ and producing the exact derivative $2\arctan u \cdot \arctan'(u)$, whose integral is $\pi^2/16$.

Putnam 2016 B6 is the same circuit with $1/(k2^n+1) = \int_0^1 x^{k2^n}dx$ and $\sum_{k\geq1}(-1)^{k-1}u^k/k = \ln(1+u)$, converting sums of reciprocals into integrals of logarithms.

- **WHEN** a limit of a sum contains a factor $1/n$ and arguments k/n **DO** read it as a Riemann sum and evaluate the integral **BECAUSE** the definition of the integral is that limit, so the recognition is the definition and not a technique
- **WHEN** a sum has denominators that are linear in the summation index **DO** replace each $1/m$ by $\int_0^1 t^{m-1}dt$ and swap sum and integral **BECAUSE** the sum inside becomes a power series with a closed form, and the denominators that blocked summation have been absorbed into the integration
- **WHEN** a right endpoint sum of a monotone function is compared with its integral **DO** use the strict overestimate and the vanishing of the excess **BECAUSE** monotonicity makes every subinterval comparison strict and makes the total excess telescope, giving the sharp constant and the fact that it is not attained

8.8 Double integrals, Fubini, and the Jacobian

Swapping the order to break a single integral

Theorem 8.8.1 (Fubini and Tonelli, the versions you need)**.** *If F is continuous on a closed rectangle $[a,b]\times[c,d]$ then the two iterated integrals exist and are equal. If $F \geq 0$ is measurable on a product of intervals then the two iterated integrals are equal in $[0,\infty]$ (Tonelli), with no integrability hypothesis. For a sign changing F on an unbounded domain you must first check $\iint |F| < \infty$, and the standard way to check it is Tonelli applied to $|F|$.*

The Putnam use is not "compute a double integral". It is: a single integral that resists evaluation becomes a double integral that does not, because in two dimensions you may change the order, and the inner integral in the other order is elementary. Two instances, both worth memorising as templates.

Example 8.8.2 (A product of integrals is a double integral, Putnam 2025 B2)**.** For continuous a, b on $[0,1]$,

$$\left(\int_0^1 a\right)\left(\int_0^1 b\right) = \iint_{[0,1]^2} a(x)b(y)\,dx\,dy.$$

The value of the right side is unchanged by renaming $x \leftrightarrow y$. Adding the two named versions symmetrises an asymmetric integrand, and the sign of a symmetric integrand is usually readable when the sign of the original is not. This is the mechanism behind Chebyshev's integral inequality, proved in the inequality half of this chapter.

Example 8.8.3 (A second difference is a double integral, Putnam 2019 B4)**.** For f of class C^2,

$$f(a+h,b+k) - f(a+h,b) - f(a,b+k) + f(a,b) = \int_a^{a+h}\int_b^{b+k} f_{xy}(x,y)\,dy\,dx.$$

Apply the fundamental theorem in y first, then in x. The consequence you carry is the characterisation: the mixed second difference vanishes identically exactly when $f(x,y) = u(x) + v(y)$, that is, exactly when f is additively separable. So whenever a problem's data are second differences, the data are really f_{xy}, and everything separable is invisible to them.

The same reading gives Putnam 2009 A6 its one sentence summary: with $a = \int_0^1 f(0,y)dy$, $b = \int_0^1 f(1,y)dy$, $c = \int_0^1 f(x,0)dx$, $d = \int_0^1 f(x,1)dx$, the fundamental theorem in one variable and Fubini in the other give

$$b - a = \iint_{[0,1]^2} f_x, \qquad d - c = \iint_{[0,1]^2} f_y.$$

The four boundary integrals are the average of the gradient. Once you have written that sentence, the question "must the gradient equal $(b-a, d-c)$ somewhere inside?" is visibly a question about whether a mean value theorem holds in two dimensions, and the answer is no, which tells you to build a counterexample rather than a proof.

Change of variables and the Jacobian

Theorem 8.8.4 (Change of variables)**.** *Let $T : U \to V$ be a C^1 diffeomorphism between open subsets of $\mathbb{R}^n$ and h integrable on V. Then*

$$\int_V h(\mathbf{y})\,d\mathbf{y} = \int_U h(T(\mathbf{u}))\,\big|\det DT(\mathbf{u})\big|\,d\mathbf{u}.$$

The determinant is not bookkeeping. $\det DT$ is by definition the factor by which T scales oriented volume to first order, so the theorem says nothing more than "volume element transforms by the local volume scaling factor", and the modulus is there because volume is unsigned. Polar coordinates are the case $T(r,\theta) = (r\cos\theta, r\sin\theta)$ with $\det DT = r$; spherical coordinates give $\rho^2 \sin\phi$.

The linear case is worth stating separately because it is the one that fires on Putnam problems.

Corollary 8.8.5. *For invertible linear T and measurable E, $\int_{T(E)} h = |\det T| \int_E h \circ T$. If $T = \lambda U$ with U orthogonal then $|T(p)| = \lambda|p|$ and $|\det T| = \lambda^n$.*

Putnam 2021 A4 is exactly this: $T(x,y) = (x+y, x-y)$ has $|\det T| = 2$ and scales lengths by $\sqrt{2}$, and the identity $(x+y)^4 + (x-y)^4 = 2(x^4 + 6x^2y^2 + y^4)$ is what makes the denominator transform. So the integral of one function over the disc of radius R equals a constant times the integral of the other over the disc of radius $R\sqrt{2}$, and the two quantities the problem compares are the same quantity at two scales.

Anchoring the polar origin cleverly is a separate decision. Putnam 2006 A6 puts the polar origin at a boundary point of the circle rather than at the centre, and the circle of radius R through the origin with centre in the direction $\theta = \pi/2$ then has the single equation $r = 2R\sin\theta$, $\theta \in [0, \pi]$. Choosing the origin so the region becomes one inequality rather than two is the same instinct as choosing coordinates so a constraint becomes an axis.

- **WHEN** a single integral resists every substitution **DO** write one of its factors as an integral (a reciprocal, a difference, a max) and swap the order of the resulting double integral **BECAUSE** Fubini makes the two orders equal, and the inner integral in the other order is usually elementary because you chose the representation for that reason
- **WHEN** you are comparing two integrals of similar shape over a symmetric region **DO** look for a linear map carrying one integrand to the other and pay $|\det T|$ **BECAUSE** a linear change of variables is always legal, and if the two integrands are related by one it converts a comparison into an identity
- **WHEN** the region is a disc or annulus, or the integrand depends on $x^2 + y^2$ **DO** go polar and choose the origin so the region becomes a single inequality **BECAUSE** the radial Jacobian r is exactly the factor a radially even integrand needs, and the choice of origin decides whether the domain description costs one line or five
- **WHEN** a hypothesis is stated as a second difference of a two variable function **DO** rewrite it as a double integral of f_{xy} **BECAUSE** the double integral form shows immediately that additively separable parts are invisible to the data, which is usually the whole content of the problem

8.9 Improper integrals, convergence, and principal values

Definition 8.9.1. $\int_a^\infty f := \lim_{R\to\infty} \int_a^R f$ when the limit exists; similarly at an endpoint singularity. The integral converges *absolutely* if $\int_a^\infty |f| < \infty$, and *conditionally* if it converges but not absolutely.

The three tests you need, with one line of proof each.

Proposition 8.9.2 (Comparison)**.** *If $0 \le f \le F$ on $[a,\infty)$ and $\int_a^\infty F$ converges, then so does $\int_a^\infty f$. The model case is $\int_1^\infty t^{-p}dt$, convergent exactly for $p > 1$, with value $1/(p-1)$; at the other end, $\int_0^1 t^{-p}dt$ converges exactly for $p < 1$.*

Proof. $R \mapsto \int_a^R f$ is nondecreasing and bounded above by $\int_a^\infty F$, so it converges. □

Putnam 2011 A5 uses only this: the hypothesis $x^2 g(x) \le 1$ gives $g(t) \le t^{-2}$, comparison makes the primitive G bounded, and a bounded primitive turns the whole real line into a bounded interval, which is what makes a pigeonhole step available. That is the general reason to care about convergence on a Putnam: convergence is rarely the question, it is the device that compactifies a non compact domain.

Proposition 8.9.3 (Dirichlet's test)**.** *If $F(R) = \int_a^R g$ is bounded and f is monotone with $f(x) \to 0$, then $\int_a^\infty fg$ converges.*

Proof. Integrate by parts: $\int_a^R fg = f(R)F(R) - \int_a^R Ff'$. The first term tends to 0; the second converges absolutely because F is bounded and $\int_a^\infty |f'| = |f(a)|$ by monotonicity. □

This is what makes $\int_0^\infty \sin x/x\,dx$ converge while $\int_0^\infty |\sin x|/x\,dx$ diverges. Conditional convergence is not a technicality: it is the reason you may not swap that integral with a sum without justification.

Definition 8.9.4 (Principal value)**.** If f is integrable off a point c but not through it,

$$\mathrm{PV}\int_a^b f := \lim_{\varepsilon\to 0^+}\left(\int_a^{c-\varepsilon} f + \int_{c+\varepsilon}^b f\right),$$

the two sides being cut symmetrically. So $\mathrm{PV}\int_{-1}^1 dx/x = 0$ although the integral does not converge.

Principal values have not appeared in the filtered corpus. Learn the definition (it costs one line) and learn the warning that goes with it: a principal value is not an integral, and you may not apply Fubini, dominated convergence or a substitution to it without checking. Learn it last.

One more warning that the corpus does supply, from Putnam 2010 A6: for $t \in [0, 1/2]$ you have $-\ln(1-t) \le t + 2t^2 \le 2t$, but *no* constant C satisfies $-\ln(1-t) \le Ct$ for all $t \in [0,1)$, because the left side blows up at 1. A bound that holds on a compact subinterval is not a bound on the half open interval, and the failure of uniformity is exactly the gap that forces a harder argument. When you write "$\ln(1-t) \approx -t$", say on what range.

- **WHEN** an integral over an infinite interval appears in a hypothesis rather than a question **DO** use it to bound a primitive and thereby compactify the domain **BECAUSE** a bounded monotone primitive maps an unbounded domain into a bounded interval, which is what makes pigeonhole, Bolzano and Weierstrass, or the intermediate value theorem available
- **WHEN** the integrand oscillates and decays too slowly for absolute convergence **DO** quote Dirichlet's test with the bounded primitive of the oscillating factor **BECAUSE** integration by parts moves the decay onto a term that is absolutely integrable, and monotonicity makes $\int |f'|$ telescope
- **WHEN** you are about to use a local approximation such as $\ln(1+u) \approx u$ inside an estimate **DO** state the interval on which your two sided bound holds and check the argument stays there **BECAUSE** such bounds fail uniformly near the boundary of their range, and a non uniform bound proves nothing about a limit that approaches the boundary

8.10 Integration by parts as a structural tool

Everyone can integrate by parts. The Putnam skill is choosing the *direction*, and there are exactly three reasons to choose one direction over the other.

Reason one: annihilate the boundary term exactly

Proposition 8.10.1. *When integrating a product by parts and one factor vanishes at an endpoint of interest, differentiate that factor rather than integrate it, so the boundary term is exactly zero rather than merely small.*

Putnam 2011 A3 is the model. With $f(r) = \int_0^{\pi/2} x^r \sin x\,dx$ and $g(r) = \int_0^{\pi/2} x^r \cos x\,dx$, integrating g by parts with $u = \cos x$, $dv = x^r dx$ gives

$$g(r) = \left[\frac{x^{r+1}}{r+1}\cos x\right]_0^{\pi/2} + \frac{1}{r+1}\int_0^{\pi/2} x^{r+1}\sin x\,dx = \frac{f(r+1)}{r+1},$$

because $\cos(\pi/2) = 0$ kills the boundary term outright. An *exact identity* between g and f replaces any need to estimate g separately. Had you integrated cos instead, you would have a boundary term of size $(\pi/2)^{r+1}$ that you then have to bound, and the whole asymptotic would become a competition between two large quantities.

Reason two: trade a rough bound for a smooth one via the primitive

Putnam 2014 B2: if you only know a rough pointwise bound on f but a good Lipschitz bound on its primitive $F(x) = \int_a^x f$, integrate by parts to move the derivative off f:

$$\int_a^b f(x)\phi(x)\,dx = \big[F\phi\big]_a^b - \int_a^b F(x)\phi'(x)\,dx.$$

The reflex: when a rough bound on f is too weak against a smooth kernel, the primitive is the object with the good bound, and parts is the only way to swap them.

Reason three: make the inequality close on itself

Putnam 2018 A5 is the sharpest example in the corpus. For u of class C^2 with $u(0) = 0$ and $x \geq 0$,

$$\int_0^x t\,u''(t)\,dt = x\,u'(x) - \int_0^x u'(t)\,dt = x\,u'(x) - u(x).$$

The two quantities appearing on the right, $xu'(x)$ and $u(x)$, are exactly the two quantities in the statement being proved. So an inequality bounding u' by a multiple of tu'' converts into an inequality relating u and u' alone, and the induction closes on itself. The base case is the crude comparison: if $u'' \geq 0$ then u' is nondecreasing, so $u(x) = \int_0^x u' \leq x\,u'(x)$, which is the constant $c_1 = 1$. The bootstrap then improves the constant to $c_k = 1/k$; the recursion is $c_{k+1} = c_k/(1 + c_k)$, which is solved in the recursion file below.

- **WHEN** you integrate a product by parts and one factor vanishes at an endpoint you care about **DO** differentiate that factor, do not integrate it **BECAUSE** the boundary term becomes exactly zero, converting an asymptotic comparison into an exact identity, and exact identities never need error terms
- **WHEN** a pointwise bound on f is too weak once f is multiplied by a smooth kernel **DO** integrate by parts to replace f by its primitive F and ϕ by ϕ' **BECAUSE** the primitive is smoother than f and typically satisfies a Lipschitz bound, and the derivative of the kernel is usually small
- **WHEN** an inequality involves both a function and its derivative **DO** integrate by parts so that the boundary term and the remaining integral are precisely the two quantities in the statement **BECAUSE** the inequality then closes on itself and becomes a one dimensional recursion in the constant, which you solve rather than iterate

8.11 Asymptotics of integrals

Two facts cover almost every Putnam asymptotic in the filtered corpus.

Lemma 8.11.1 (Endpoint concentration)**.** *Let w be continuous on $[0, b]$ with $w(b) \neq 0$. Then*

$$\int_0^b x^r w(x)\,dx \sim \frac{b^{r+1}w(b)}{r} \qquad (r \to \infty).$$

Proof. Substituting $x = bt$, the claim is $r\int_0^1 t^r w(bt)\,dt \to w(b)$. Since $r\int_0^1 t^r\,dt = r/(r+1) \to 1$, it suffices to show $r\int_0^1 t^r\big(w(bt) - w(b)\big)dt \to 0$. Given $\varepsilon > 0$ pick δ with $|w(bt) - w(b)| < \varepsilon$ on $(1-\delta, 1]$, and let $M = \max|w|$. The contribution from $[0, 1-\delta]$ is at most $2Mr(1-\delta)^{r+1}/(r+1) \to 0$, and the rest is at most $\varepsilon\,r/(r+1) \leq \varepsilon$. □

The reading is: x^r is a spike at the right endpoint of geometric sharpness, so only $w(b)$ survives. Putnam 2011 A3 uses it twice. Directly, $f(r) = \int_0^{\pi/2} x^r \sin x\,dx \sim (\pi/2)^{r+1}/r$, since $\sin(\pi/2) = 1$. And by contrast, $g(r) = \int_0^{\pi/2} x^r \cos x\,dx$ has a weight vanishing at the endpoint, so its order drops by exactly one full power of r; the exact identity $g(r) = f(r+1)/(r+1)$ from the previous section then gives

$$\frac{r\,g(r)}{f(r)} = \frac{r}{r+1} \cdot \frac{f(r+1)}{f(r)} \longrightarrow \frac{\pi}{2},$$

because the lemma gives $f(r+1)/f(r) \to \pi/2$. The recursion worth extracting: for such a family, consecutive members satisfy $f(r)/f(r+1) \to 1/b$, and a weight vanishing at b costs exactly one power of r.

Lemma 8.11.2 (Two sided bounds for the logarithm)**.** *For* $0 < u \le \frac{1}{2}$, $-u - u^2 \le \ln(1-u) \le -u$.

Proof. $\ln(1-u) = -\sum_{j\ge1} u^j/j$. The upper bound drops the terms $j \ge 2$, all negative contributions. For the lower bound, $\sum_{j\ge2} u^j/j \le \frac{1}{2}\sum_{j\ge2} u^j = \dfrac{u^2}{2(1-u)} \le u^2$ when $u \le \frac{1}{2}$. □

The second order term is not decoration. It matters exactly when it multiplies a quantity of size u^{-2}, which is what happens in Putnam 2024 B6: writing $x^{n^2} = e^{n^2 \ln x}$ with $\ln x = \ln(1-u)$ and $u \approx 1/t$, the term u^2 times $n^2 \approx t^2$ is of order one. Read that as a rule: keep one more term in the expansion than the power of the large parameter you are multiplying by.

- **WHEN** an integral $\int_0^b x^r w(x)dx$ appears with $r \to \infty$ **DO** quote endpoint concentration; if $w(b) = 0$, integrate by parts to relate it exactly to a member of the same family with r shifted **BECAUSE** the spike x^r localises at b, so a nonvanishing weight gives leading order $b^{r+1}w(b)/r$ and a vanishing weight loses exactly one power of r
- **WHEN** you expand $\ln(1-u)$ or e^{-u} inside an expression carrying a large parameter **DO** count how many powers of the large parameter multiply the discarded term before discarding it **BECAUSE** a second order term times a second power of the parameter is of order one, and dropping it silently changes the answer

8.12 Volumes, surfaces and arc lengths

A-level gave you $V = \pi \int y^2 dx$ about the two axes. Here is the rest, stated compactly, because these are formulae rather than ideas and the ideas are in the multivariable section.

Washers and shells. Revolving a planar region about an axis,

$$V = \int \pi\big(r_{\text{out}}(z)^2 - r_{\text{in}}(z)^2\big)\,dz \qquad \text{(washers)}, \qquad V = \int 2\pi\,\rho\,h(\rho)\,d\rho \qquad \text{(shells)}.$$

Washers slice perpendicular to the axis, shells parallel. Putnam 2006 A1 is a washer computation that Pappus does in one line instead.

Pappus. If a plane region of area A lies on one side of a coplanar line L at distance d from its centroid, the solid of revolution has volume $V = 2\pi dA$. The surface version: a plane curve of length L at centroid distance d sweeps area $S = 2\pi dL$. Both are one line consequences of Fubini once you write the volume as $\iint 2\pi\,\mathrm{dist}(p, L)\,dA(p)$ and recognise $\iint \mathrm{dist}\,dA = dA$ as the definition of the centroid distance.

Arc length and surface area.

$$L = \int \sqrt{1 + (y')^2}\,dx = \int \sqrt{(x')^2 + (y')^2}\,dt = \int \sqrt{r^2 + (r')^2}\,d\theta, \qquad S = 2\pi \int y\sqrt{1 + (y')^2}\,dx.$$

Surface integrals. For a surface parametrised by $S(u, v)$, the area element is $dS = |S_u \times S_v|\,du\,dv$. For the unit sphere with $S(\phi, \theta) = (\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi)$ this gives $dS = \sin\phi\,d\phi\,d\theta$, and the

substitution $t = \cos\phi$ consumes the $\sin\phi$ exactly. Putnam 2019 A4 verifies a counterexample with precisely this computation, so the whole problem turns on being fluent with one Jacobian.

Centroids by slicing. The centroid is the first moment divided by the total mass. For the region under $y = f(x)$ on $[0,1]$, $\bar{x} = \int xf / \int f$. For a solid whose cross section at abscissa x has area $A(x)$, $\bar{x} = \int xA / \int A$; for a solid of revolution $A(x) = \pi f(x)^2$, so $\bar{x} = \int xf^2 / \int f^2$ with π cancelling. Note what this says: the x centroid depends only on the area function, not on the shape of the cross sections. Putnam 2025 B2 compares exactly these two numbers, and the observation that rotation merely squares the weight is what turns a solid geometry problem into the Chebyshev inequality.

Cylindrical coordinates and rotational invariance. A region in $\mathbb{R}^3$ whose defining relation involves x, y only through $r = \sqrt{x^2+y^2}$ is invariant under rotation about the z axis, so the substitution collapses a three variable inequality into a two variable one in the meridian half plane. Putnam 2006 A1 is exactly this collapse followed by Pappus.

- **WHEN** a solid of revolution is generated by a region whose centroid you can locate, and the region does not meet the axis **DO** quote Pappus, $V = 2\pi d A$ **BECAUSE** the volume integral is $\iint 2\pi \,\mathrm{dist}\, dA$ and the centroid is defined as the average of dist, so the formula is the definition rearranged
- **WHEN** a three variable region is defined by a relation depending on x, y only through x^2+y^2 **DO** switch to cylindrical coordinates and work in the meridian half plane **BECAUSE** the defining relation is rotation invariant, so the region is a surface of revolution and one dimension carries no information
- **WHEN** two averages weighted by f and by f^2 are to be compared **DO** write both as $\int xw / \int w$ and apply an ordering inequality to the weights **BECAUSE** a centroid is a weighted mean, and comparing weighted means with different weights is exactly what Chebyshev's inequality does

8.13 Multivariable calculus at Putnam depth

Gradients and directional derivatives

Definition 8.13.1. For $h : \mathbb{R}^n \to \mathbb{R}$ differentiable and $\mathbf{v} \in \mathbb{R}^n$, the directional derivative is

$$D_{\mathbf{v}} h(\mathbf{p}) = \left.\frac{d}{dt}\right|_{t=0} h(\mathbf{p} + t\mathbf{v}) = \nabla h(\mathbf{p}) \cdot \mathbf{v}.$$

That equality is the entire content of "the gradient". The gradient is not a list of partial derivatives; it is the vector that represents the linear map $\mathbf{v} \mapsto D_{\mathbf{v}} h$, and everything you know about gradients (steepest ascent, orthogonality to level sets) is that sentence read in different directions.

Recognising a directional derivative in a printed equation is the whole entry point into Putnam 2010 A3, and the problem is worth doing in full here because it is three lines.

Example 8.13.2 (Putnam 2010 A3)**.** Suppose $h : \mathbb{R}^2 \to \mathbb{R}$ has continuous partials and satisfies $h = a\,h_x + b\,h_y$ for constants a, b, and $|h| \le M$ everywhere. Fix (x_0, y_0) and set $\phi(t) = h(x_0 + at,\ y_0 + bt)$. By the chain rule,

$$\phi'(t) = a\,h_x + b\,h_y = h(x_0 + at, y_0 + bt) = \phi(t),$$

so $\phi(t) = \phi(0)e^t$. If $\phi(0) \neq 0$ this is unbounded as $t \to \infty$ (or $t \to -\infty$ if $\phi(0)$ has the other sign of the exponent), contradicting $|h| \le M$. Hence $h(x_0, y_0) = \phi(0) = 0$, and the point was arbitrary.

That is the *method of characteristics*: for a first order linear equation $ah_x + bh_y = \lambda h$ with constant a, b, λ, restrict h to the lines in direction (a, b) and the partial differential equation becomes the ordinary equation $\phi' = \lambda\phi$, whose solution is $\phi(0)e^{\lambda t}$. The lines are the characteristics. The bifurcation point in that problem is the decision to restrict to a line instead of hunting for a two

variable solution formula: one line per point suffices, because the equation only ever compares h with its derivative in the single direction (a, b).

The variable coefficient version, needed in Putnam 2011 A5, is the same idea.

Proposition 8.13.3. *If a, b are continuous and (a, b) never vanishes, a C^1 function F satisfies $aF_u + bF_v = 0$ exactly when F is constant along the integral curves of the field (a, b). If w is a first integral with nonvanishing gradient, every solution is $F = h \circ w$. In particular $\Phi_x + \Phi_y = 0$ on a convex domain forces $\Phi(x, y) = h(x - y)$.*

The chain rule in several variables, and polar operators

For C^1 maps, $\frac{d}{dt}\Phi(\gamma(t)) = \nabla\Phi(\gamma(t)) \cdot \gamma'(t)$, and $\frac{\partial}{\partial x}F(\alpha(x), \beta(y)) = F_u(\alpha(x), \beta(y))\alpha'(x)$. The two operators worth memorising are the polar ones. With $x = s\cos t$, $y = s\sin t$,

$$\frac{\partial}{\partial t} = -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}, \qquad s\frac{\partial}{\partial s} = x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} \text{ (the Euler operator).}$$

The first has an immediate consequence with real force: for every single valued C^1 function h and every $s > 0$,

$$\int_{|p|=s} (y\, h_x - x\, h_y)\, dt = -\int_0^{2\pi} \frac{\partial h}{\partial t}\, dt = 0,$$

because h is 2π periodic in t. So every circle centred at the origin is a null set for the density $\rho = yh_x - xh_y$. That single observation is the mechanism of Putnam 2021 B3.

Vector calculus where it appears

You need three identities and two theorems. The identity $\operatorname{div}(h\mathbf{V}) = \nabla h \cdot \mathbf{V} + h \operatorname{div}\mathbf{V}$; the fact that the rotational field $\mathbf{V} = (y, -x)$ is divergence free; and $\operatorname{curl}\nabla h = 0$. The theorems are Green's theorem in the plane and the divergence theorem

$$\int_D \operatorname{div}\mathbf{F}\, dA = \int_{\partial D} \mathbf{F}\cdot\mathbf{n}\, ds.$$

Combining them gives the alternative route in Putnam 2021 B3: since $\operatorname{div}\mathbf{V} = 0$ for $\mathbf{V} = (y, -x)$, the density $\rho = yh_x - xh_y = \nabla h \cdot \mathbf{V} = \operatorname{div}(h\mathbf{V})$, so the area integral of ρ over a region equals the boundary integral of $h\,\mathbf{V}\cdot\mathbf{n}$. On the circle of radius r about $c = (d\cos p, d\sin p)$ parametrised by t, one computes $\mathbf{V}\cdot\mathbf{n} = d\sin(p - t)$. The moral: an integrand that is a total derivative in disguise is always worth spotting, because it converts an area integral into a boundary integral of one dimension less.

Differentiating a domain integral

Theorem 8.13.4 (Transport, or first variation of a domain integral)**.** *For a smooth family of domains $R(t)$ with boundary normal speed V_n, and f independent of t,*

$$\frac{d}{dt}\int_{R(t)} f = \int_{\partial R(t)} f\, V_n\, dS.$$

In one dimension this is $\frac{d}{dt}\int_{a(t)}^{b(t)} f = f(b)b' - f(a)a'$.

The reading: moving a domain only gains and loses material at the boundary, at a rate equal to the boundary's normal speed, so the derivative of an integral over a moving region is a boundary integral. Putnam 2012 A6 is three applications of this one formula to three motions of the same rectangle (two translations and one shape change), and it is the mechanism that converts an averaged hypothesis into edge information. Whenever a hypothesis says "the average of f over every rectangle of a given type is zero", differentiate in the rectangle's parameters: the averages are the data, the boundary values are what you want, and this theorem is the bridge.

Critical points, the second derivative test, Lagrange multipliers

Proposition 8.13.5. *A continuous function on a compact set attains its maximum and minimum. If an extremum occurs at an interior point where the function is differentiable, the gradient vanishes there.*

This is used constructively as a *design constraint*, not just as a theorem. In Putnam 2009 A6 you are building a counterexample with nowhere vanishing gradient, so the proposition tells you that every extreme value must sit on the boundary of the square, which is exactly why the construction uses a factor of y or a monotone factor. Normalisation helps: setting $g = f - \alpha x - \beta y$ shifts $\nabla g = \nabla f - (\alpha, \beta)$ and shifts the boundary constants by the same amounts, so choosing $(\alpha, \beta) = (b - a,\ d - c)$ reduces the whole claim to the existence of an interior critical point. The general reflex: subtract the affine function that makes the target read "gradient vanishes" rather than "gradient equals a specific vector".

Proposition 8.13.6 (Second derivative test)**.** *At a critical point of a C^2 function, if the Hessian is positive definite the point is a strict local minimum; negative definite, a strict local maximum; indefinite, a saddle. If the Hessian is singular the test is inconclusive.*

Theorem 8.13.7 (Lagrange multipliers)**.** *Let $f, g_1, \dots, g_m$ be C^1 on an open set. If f restricted to $M = \{g_1 = \cdots = g_m = 0\}$ has a local extremum at $\mathbf{p}$ and the constraint qualification holds at $\mathbf{p}$, namely that $\nabla g_1(\mathbf{p}), \dots, \nabla g_m(\mathbf{p})$ are linearly independent, then there exist λ_i with $\nabla f(\mathbf{p}) = \sum_i \lambda_i \nabla g_i(\mathbf{p})$.*

Three failure modes, all of which have cost people the marks, and one of which is directly evidenced in the corpus.

First, the constraint qualification is a hypothesis, not a formality: at a point where ∇g vanishes the conclusion is simply false, and you must check it or handle such points separately.

Second, a Lagrange stationary point is a *critical point*, not an optimum. You must still compare the values you find, and you must separately handle the boundary and any non compact directions. Putnam 2023 B4 is the corpus's warning: minimising a positive linear form on the positive part of an ellipsoid has no interior minimum at all, so the interior stationary point that multipliers hand you is the *maximum*, and the natural first attempt gives $s_k \propto 1/k$, a confidently wrong answer.

Third, inequality constraints are not equality constraints. With constraints $g_i \le 0$ you need the Karush, Kuhn and Tucker conditions, which add complementary slackness: $\lambda_i \ge 0$ and $\lambda_i g_i = 0$, so a multiplier can be nonzero only on an active constraint. In practice, on a Putnam, the cleanest replacement for KKT is the vertex argument of the next section: on a polytope, a convex objective is maximised at a vertex, and you enumerate vertices.

- **WHEN** a printed equation contains $ah_x + bh_y$ **DO** recognise a directional derivative and restrict h to the lines in direction (a, b) **BECAUSE** the chain rule turns the partial differential equation into $\phi' = \lambda\phi$ along each such line, and a one variable ordinary equation you can solve exactly
- **WHEN** you see $yh_x - xh_y$ **DO** recognise minus the angular derivative in polar coordinates **BECAUSE** it is an exact derivative in t, so its integral round any circle centred at the origin vanishes by periodicity, with no computation
- **WHEN** a hypothesis controls the average of f over a whole family of moving regions **DO** differentiate with respect to the motion parameters using the transport formula **BECAUSE** a moving domain gains material only at its boundary, so differentiating converts averaged information into boundary information, which is the information you actually want
- **WHEN** you are asked to build a counterexample whose gradient avoids a value **DO** normalise by subtracting the affine function that turns the target into "gradient vanishes", then place all extrema on the boundary **BECAUSE** interior extrema force critical points, so a nowhere critical function must attain its extremes on the boundary, and that is a design constraint you can build to

- **WHEN** you reach for Lagrange multipliers on a constrained optimisation **DO** check that the constraint gradients are independent, compare the values at all stationary points, and handle the boundary and any inequality constraints separately **BECAUSE** multipliers detect stationarity, not optimality, and with inequality constraints the optimum routinely sits at a vertex where no interior stationary point exists

8.14 The mean ladder

Now the inequality half. Everything begins with one convexity fact, so state it first and derive the ladder from it rather than memorising four separate statements.

Definition 8.14.1. F is convex on an interval I if $F(\lambda x_1 + (1-\lambda)x_2) \le \lambda F(x_1) + (1-\lambda)F(x_2)$ for all $x_1, x_2 \in I$ and $\lambda \in [0,1]$; concave if the reverse holds. For F twice differentiable, convex is equivalent to $F'' \ge 0$.

Theorem 8.14.2 (Jensen)**.** *If F is convex on I, $x_i \in I$ and $\lambda_i \ge 0$ with $\sum \lambda_i = 1$, then*

$$F\left(\sum_i \lambda_i x_i\right) \le \sum_i \lambda_i F(x_i).$$

Equality for strictly convex F forces all x_i with $\lambda_i > 0$ to be equal. The integral form is $F\big(\int u\,d\mu\big) \le \int F(u)\,d\mu$ for a probability measure μ.

Proof. Let $m = \sum \lambda_i x_i \in I$. A convex function has a supporting line at every interior point of I: there is c with $F(x) \ge F(m) + c(x-m)$ for all $x \in I$ (take c to be any value between the left and right derivatives at m, which exist and satisfy $F'_- \le F'_+$ by convexity). Multiply by λ_i, put $x = x_i$, and sum: the linear term collapses because $\sum \lambda_i (x_i - m) = 0$. Strict convexity makes the supporting line inequality strict off m, giving the equality case. The integral form is identical with $\sum \lambda_i$ replaced by $\int d\mu$. □

That proof is worth remembering as a technique in its own right: *the supporting line collapses the linear part*. That is the tangent line trick, and it appears again below on its own.

Corollary 8.14.3 (Weighted AM and GM)**.** *For $x_i > 0$ and weights $\lambda_i \ge 0$ with $\sum \lambda_i = 1$,*

$$\prod_i x_i^{\lambda_i} \le \sum_i \lambda_i x_i,$$

with equality iff all x_i are equal. The case $\lambda_i = 1/n$ is the classical statement $\sqrt[n]{x_1 \cdots x_n} \le \frac{x_1 + \cdots + x_n}{n}$.

Proof. Apply Jensen to the strictly concave ln: $\ln\big(\sum \lambda_i x_i\big) \ge \sum \lambda_i \ln x_i = \ln \prod x_i^{\lambda_i}$, and exponentiate. □

The two variable case $t + 1/t \ge 2$ for $t > 0$ is the one you will use most often, and it is worth recording separately with its equality case ($t = 1$) because Putnam 2007 A2 applies it four times, once per cyclic pair, and reads off the equality configuration $a = b = c = d$ in the same step. Similarly $1 + x^2 \ge 2|x|$, which is the same statement, gives $\left|\frac{2x}{1+x^2}\right| \le 1$ with equality only at $x = \pm 1$: that is, the derivative of $\ln(1+x^2)$ has modulus at most 1, which is the threshold in Putnam 2022 A1.

Definition 8.14.4 (Power means)**.** For $x_i > 0$, weights λ_i summing to 1, and $r \ne 0$,

$$M_r = \left(\sum_i \lambda_i x_i^r\right)^{1/r}, \qquad M_0 := \prod_i x_i^{\lambda_i}.$$

So M_{-1} is the harmonic mean, M_1 the arithmetic mean, M_2 the quadratic mean, and $M_{\pm\infty}$ the maximum and minimum.

Theorem 8.14.5 (Power mean inequality). *$M_r \le M_s$ whenever $r < s$, with equality iff all x_i are equal.*

Proof. For $0 < r < s$: the map $t \mapsto t^{s/r}$ is convex on $(0,\infty)$ since $s/r > 1$, so Jensen with $y_i = x_i^r$ gives $\left(\sum \lambda_i y_i\right)^{s/r} \le \sum \lambda_i y_i^{s/r}$, that is $M_r^s \le M_s^s$. For $r < s < 0$ apply the positive case to $1/x_i$ and reverse. For $r < 0 < s$ it suffices to prove $M_r \le M_0 \le M_s$, and $M_0 \le M_s$ is weighted AM and GM applied to x_i^s after taking s th roots, while $M_r \le M_0$ is the same applied to $1/x_i$. □

The continuous version is worth stating because it is the shape a Putnam problem takes: for a probability measure μ and bounded positive f, $M_r(f) = \left(\int f^r d\mu\right)^{1/r}$ is nondecreasing in r and tends, as $r \to 0$, to $\exp\left(\int \ln f \, d\mu\right)$, the geometric mean. Putnam 2021 A2 is exactly this: the quantity there is e times a mean of numbers in $[x, x+1]$, so the answer e is forced by a squeeze, not computed. The supporting identity in that problem is worth its own line: for $0 < a < b$ and real s,

$$b^s - a^s = s\int_a^b u^{s-1}\,du, \qquad \text{so} \qquad \frac{b^{r+1} - a^{r+1}}{r+1} = \int_a^b u^r\,du,$$

a mean of u^r over $[a,b]$, whose value lies between a^r and b^r (with the order reversed when $r < 0$). A difference of shifted powers is an integral, and an integral over an interval of length one is a mean.

- **WHEN** a sum of positive terms is to be bounded below, or a product bounded above, and the two are related **DO** apply AM and GM, and read the equality case in the same breath **BECAUSE** the equality case names the extremal configuration, and on the Putnam the extremal configuration is usually half the answer
- **WHEN** you see $t + k/t$ for $t > 0$ **DO** its minimum is $2\sqrt{k}$ at $t = \sqrt{k}$, and it is strictly convex with second derivative $2k/t^3$ **BECAUSE** this is AM and GM in two terms, and the convexity means the function meets any horizontal line at most twice, so level set counting reduces to comparing endpoint values with the minimum
- **WHEN** two different means of the same numbers appear in one statement **DO** place them on the power mean ladder $M_{-1} \le M_0 \le M_1 \le M_2$ and quote monotonicity **BECAUSE** all four are the single family M_r, so the four separate inequalities are one theorem, proved by Jensen on $t \mapsto t^{s/r}$
- **WHEN** a quantity is a difference of shifted powers, $b^s - a^s$ **DO** write it as $s\int_a^b u^{s-1}du$ and read it as a mean **BECAUSE** the integral form makes the quantity visibly lie between the two endpoint values, which is a squeeze with no algebra

8.15 Cauchy and Schwarz, and the norm family

Theorem 8.15.1 (Cauchy and Schwarz). *For real u_i, v_i, $\left(\sum u_i v_i\right)^2 \le \left(\sum u_i^2\right)\left(\sum v_i^2\right)$, with equality iff one sequence is a scalar multiple of the other (or one is zero). The integral form, with a nonnegative weight w, is*

$$\left(\int w\,uv\right)^2 \le \left(\int w\,u^2\right)\left(\int w\,v^2\right),$$

with equality iff u and v are proportional.

Proof. Lagrange's identity,

$$\left(\sum u_i^2\right)\left(\sum v_i^2\right) - \left(\sum u_i v_i\right)^2 = \sum_{i<j}(u_i v_j - u_j v_i)^2,$$

is verified by expanding both sides. It gives the inequality and the equality case in one line: the right side vanishes iff every 2×2 minor vanishes, that is iff the two vectors are proportional. Alternatively, $\sum(tu_i + v_i)^2 \ge 0$ for all t has nonpositive discriminant, which is the same statement; the integral version uses $\int w(tu+v)^2 \ge 0$. □

Three faces of the same theorem, all of which appear in the corpus.

Face one: the reciprocal form. For $u_k > 0$, $\left(\sum_{k=1}^N u_k\right)\left(\sum_{k=1}^N 1/u_k\right) \geq N^2$, with equality iff all u_k are equal. This is Cauchy and Schwarz with $\sqrt{u_k}$ and $1/\sqrt{u_k}$, and it is the same statement as the arithmetic and harmonic mean inequality. It is the *only* bridge from an upper bound on $\sum 1/u_k$ to a lower bound on $\sum u_k$, since no pointwise inequality relates a positive number to its reciprocal in that direction. Putnam 2011 B5 lives entirely on that sentence.

Face two: the Engel form. $\displaystyle\sum_i \frac{c_i^2}{u_i} \geq \frac{\left(\sum_i c_i\right)^2}{\sum_i u_i}$ for $u_i > 0$, from Cauchy and Schwarz with $c_i/\sqrt{u_i}$ and $\sqrt{u_i}$. Use it whenever squares sit over positive denominators; it collapses the whole sum in one step.

Face three: exponent arithmetic as method selection. If a sum has N terms and its reciprocal sum is at most M, the reciprocal form gives exactly N^2/M. In Putnam 2011 B5, $N = n^2$ pairs and M is of size n, so the output is n^3. Read that backwards: an n^3 target in the conclusion *signals* Cauchy and Schwarz before you write anything, and an n^4 target signals that Cauchy and Schwarz will not reach it and a counting or Jensen argument is needed. The exponents in the answer are a command term.

Example 8.15.2 (Putnam 2010 B1)**.** Is there a sequence of reals with $\sum_i a_i^m = m$ for every positive integer m? Cauchy and Schwarz with $u_i = a_i$, $v_i = a_i^2$ gives $\left(\sum a_i^3\right)^2 \leq \left(\sum a_i^2\right)\left(\sum a_i^4\right)$, that is $9 \leq 2 \cdot 4 = 8$. No such sequence exists. The whole problem is choosing which three of the infinitely many hypotheses to feed the inequality, and the answer is: the ones whose exponents are in arithmetic progression, because that is the only pattern Cauchy and Schwarz can consume.

Theorem 8.15.3 (Hölder)**.** *For conjugate exponents $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$ and nonnegative a_i, b_i,*

$$\sum_i a_i b_i \leq \left(\sum_i a_i^p\right)^{1/p} \left(\sum_i b_i^q\right)^{1/q},$$

with the corresponding integral form. The endpoint case $(p, q) = (1, \infty)$ reads $\int |uv| \leq \left(\sup |v|\right) \int |u|$.

Proof. Normalise so that both factors on the right equal 1. Weighted AM and GM with weights $1/p, 1/q$ gives $a_i b_i = (a_i^p)^{1/p}(b_i^q)^{1/q} \leq \frac{a_i^p}{p} + \frac{b_i^q}{q}$ (Young's inequality). Sum: the right side is $\frac{1}{p} + \frac{1}{q} = 1$. □

The Putnam skill is not quoting Hölder; it is *choosing the exponents*. Putnam 2012 B2 needs an inequality tying a count n to two power sums of the same sequence at degrees a and b. Take $p = b/a$ and $q = b/(b-a)$: then

$$\sum r_i^a = \sum r_i^a \cdot 1 \leq \left(\sum r_i^b\right)^{a/b} n^{1-a/b},$$

which reproduces exactly the powers demanded. So the rule is: read the target's exponents, and solve for the conjugate pair that produces them. In that problem $a = 2$, $b = 3$ gives $p = 3/2$, $q = 3$, and there is nothing arbitrary about it. Putnam 2013 B4 uses only the endpoint case, $\int f_0^2 g^2 \leq M(g)^2 \int f_0^2$, which converts a product into a constant times a single factor while leaving the other factor untouched.

Theorem 8.15.4 (Minkowski)**.** $\left(\int (u+v)^2\right)^{1/2} \leq \left(\int u^2\right)^{1/2} + \left(\int v^2\right)^{1/2}$*, and more generally with exponent $p \geq 1$.*

Proof. Expand $\int (u+v)^2 = \int u^2 + 2\int uv + \int v^2$ and bound the middle term by Cauchy and Schwarz; the right side is then the square of the sum of the two norms. □

Minkowski says the L^2 norm is a norm, and its Putnam use is that $\sqrt{\mathrm{Var}}$ is therefore a seminorm on functions modulo constants. Putnam 2013 B4 gets the strictly stronger bound $\sqrt{\mathrm{Var}(fg)} \leq$

$M(g)\sqrt{\mathrm{Var}(f)} + M(f)\sqrt{\mathrm{Var}(g)}$ from it, and the factor 2 in the printed statement is exactly the slack between the arithmetic and quadratic means of two numbers, since $(A+B)^2 \le 2A^2 + 2B^2$. That two term bound is worth its own line: it rearranges to $(A-B)^2 \ge 0$, and its m term form $\left(\sum A_i\right)^2 \le m \sum A_i^2$ is Cauchy and Schwarz against the all ones vector, while $\left(\sum A_i\right)^p \le m^{p-1} \sum A_i^p$ for $p \ge 1$ is Jensen on $t \mapsto t^p$.

Two smaller members of the family, both used in the corpus. The triangle inequality for integrals, $\left|\int f\right| \le \int |f| \le \max |f|$ over $[0,1]$, which is what eliminates the mean in Putnam 2013 B4 when M is not translation invariant. And the triangle inequality for complex numbers with its equality case: $|\sum w_k| \le \sum |w_k|$, with equality iff all w_k share an argument. Putnam 2018 B2 uses it twice, once for the bound and once for the equality case, and it is the equality case that forces $z = 1$.

- **WHEN** a sum of products, or an integral of a product, must be bounded by the two factors separately **DO** Cauchy and Schwarz, and state the equality case immediately **BECAUSE** Lagrange's identity gives both at once, and the Putnam usually asks for the extremal configuration
- **WHEN** you have an upper bound on $\sum 1/u_k$ and want a lower bound on $\sum u_k$ **DO** use $(\sum u_k)(\sum 1/u_k) \ge N^2$ **BECAUSE** no pointwise inequality goes in that direction, so the only bridge is the reciprocal form of Cauchy and Schwarz
- **WHEN** the target inequality has exponents a and b on power sums plus a count n **DO** solve $p = b/a$, $q = b/(b-a)$ and apply Hölder against the all ones vector **BECAUSE** the conjugate pair is determined by the exponents you must produce, so the choice is a calculation, not a search
- **WHEN** squares sit over positive denominators **DO** use the Engel form $\sum c_i^2/u_i \ge (\sum c_i)^2/\sum u_i$ **BECAUSE** it collapses the entire sum in one application instead of term by term estimates
- **WHEN** the answer's exponent in n is stated in the problem **DO** use it to choose the method before writing anything **BECAUSE** Cauchy and Schwarz on N terms with reciprocal sum M produces exactly N^2/M, so the exponent tells you whether that method can reach the target at all

8.16 Convexity as a working tool

Jensen was proved above. Here are the four consequences that do the work.

The tangent line inequality. A convex differentiable F lies above each of its tangents: $F(x) \ge F(a) + F'(a)(x-a)$; a concave F lies below its tangents and above its chords. Three instances you must know cold:

$$e^x \ge 1 + x, \qquad \ln(1+x) \le x \ \ (x > -1), \qquad \frac{x}{1+x} \le \ln(1+x),$$

and Bernoulli, $(1+x)^r \ge 1 + rx$ for $x \ge -1$ and $r \ge 1$ or $r \le 0$, with the inequality reversed for $0 \le r \le 1$ (in each case it is the tangent line at 0 of $t \mapsto (1+t)^r$, and the direction follows the sign of $r(r-1)$).

Putnam 2012 B4 uses $e^x \ge 1 + x$ twice with opposite signs, at $x = 1/(n+1)$ to close the induction $a_n > \ln(n+1)$ and at $x = -1/(n+1)$ to show the difference $a_n - \ln n$ is decreasing. Putnam 2016 B1 uses the same bound to certify that $e^x - x > 0$, so every logarithm in a recursion has a positive argument. Both are the same circuit: a recursion whose increment is a transcendental function is compared with a recursion whose increment is the tangent line.

Chords, for the reverse direction. If $F'' \le 0$ on $[p,q]$ then F lies *above* the chord: $F(x) \ge F(p) + \frac{F(q)-F(p)}{q-p}(x-p)$. Jordan's inequality is the case of sin on $[0, \pi/2]$: $\sin x \ge 2x/\pi$, with equality only at the endpoints. Putnam 2011 A3 uses it to pin an exact asymptotic constant, matched against the trivial upper bound $\sin x < 1$.

Example 8.16.1 (Putnam 2025 A2, both directions in four lines)**.** Find the largest a and smallest b with $a\,x(\pi-x) \le \sin x \le b\,x(\pi-x)$ on $(0,\pi)$. Since $x(\pi-x) > 0$, the admissible a are exactly those at most $\inf\rho$ and the admissible b exactly those at least $\sup\rho$, where $\rho(x) = \sin x/(x(\pi-x))$. Now $\rho(x) \to 1/\pi$ as $x \to 0^+$ and $\rho(\pi/2) = 4/\pi^2$, which gives the two answers, and each needs one chord.

Lower: $\cos t \ge 1 - 2t/\pi$ on $[0,\pi/2]$ is the chord of cos; integrating from 0 to x (legitimate because integrating an inequality between derivatives of functions agreeing at 0 preserves it) gives $\sin x \ge x - x^2/\pi = x(\pi-x)/\pi$, and symmetry about $\pi/2$ finishes.

Upper: put $u = x - \pi/2$, so $\sin x = \cos u$ and $\frac{4}{\pi^2}x(\pi-x) = 1 - \frac{4}{\pi^2}u^2$. The claim becomes $1-\cos u \ge \frac{4}{\pi^2}u^2$; the half angle identity $1-\cos u = 2\sin^2(u/2)$ turns this into $\sin v \ge \frac{2\sqrt{2}}{\pi}v$ for $v = |u|/2 \in [0,\pi/4]$, which is the chord of sin on $[0,\pi/4]$, of slope $\sin(\pi/4)/(\pi/4) = 2\sqrt{2}/\pi$.

Notice what was not done: no critical point of any difference was ever located. The half angle identities $1-\cos u = 2\sin^2(u/2)$ and $1+\cos u = 2\cos^2(u/2)$ were used because they rewrite the deficit of a trigonometric function from its peak as a perfect square, which is the only form in which it can be compared with a parabola centred at that peak.

Closure rules. Convexity is preserved by pointwise suprema, by nonnegative combinations, by affine substitution inside, and by composition with a nondecreasing convex function outside. It is *not* preserved by infima. Putnam 2012 B5 is nothing but the supremum rule: each $x \mapsto g_1(s)^x g_2(s) = e^{\alpha_s x + \beta_s}$ is convex (indeed affine in the exponent), so the supremum over s is convex, and a convex function is the supremum of its affine minorants. A positive F is *log convex* if $\ln F$ is convex; log convex implies convex (compose with the increasing convex exp) but not conversely, so proving log convexity when convexity was asked is doing extra work, and noticing that is what lets you delete a stage of the proof.

Convex functions are maximised at extreme points. On a compact convex set, a convex function attains its maximum at an extreme point; for a polytope this is Jensen applied to a representation of a point as a convex combination of vertices. This is the practical replacement for KKT. Putnam 2023 B4 uses it to *prove*, not assume, that the optimal vector saturates every lower bound constraint: for $P = \{s \in \mathbb{R}^N : \sum s_k = T,\ s_k \ge \ell_k\}$ the vertices are exactly the feasible points at which all coordinates but one sit at their floors, and enumerating them reduces the problem to two explicit families. Putnam 2016 A6 uses the linear version: a linear functional on a compact convex subset of $\mathbb{R}^d$ is maximised at an extreme point, and an extreme point has at least d active constraints, which predicts *how many* contact conditions to impose and turns the equality analysis into a determinate linear problem.

- **WHEN** an inequality involves e^x, $\ln(1+x)$ or $(1+x)^r$ against a linear function **DO** quote the tangent line at 0 and check the direction from the sign of the second derivative **BECAUSE** a convex function lies above every tangent, so the comparison is a one line consequence of the definition and needs no calculus at the point of use

- **WHEN** a trigonometric function must be compared with a polynomial on a closed interval **DO** use the chord on the interval where it is concave and the tangent where it is convex, after recentring at the peak with a half angle identity **BECAUSE** concavity places the graph above its chords and below its tangents, so both bounds are structural and neither requires locating a critical point

- **WHEN** a quantity is a supremum over a parameter of a family of exponentials or affine functions **DO** quote closure of convexity under suprema **BECAUSE** the supremum of convex functions is convex directly from the definition, and no differentiability is involved

- **WHEN** you must maximise a convex objective over a polytope **DO** enumerate the vertices **BECAUSE** a convex function on a compact convex set attains its maximum at an extreme point, which converts a continuous optimisation into a finite check and sidesteps complementary

slackness

8.17 Ordering inequalities: Chebyshev, rearrangement, majorisation

These are the inequalities whose hypothesis is about the *order* of the terms rather than their size, and they are the ones A-level never hints at.

Theorem 8.17.1 (Chebyshev's sum inequality, and its integral form)**.** *If $a_1 \le \cdots \le a_n$ and $b_1 \le \cdots \le b_n$ (similarly ordered), then $n\sum a_i b_i \ge \big(\sum a_i\big)\big(\sum b_i\big)$; oppositely ordered reverses it. With a weight $w \ge 0$ and u, v monotone in the same direction on $[0,1]$,*

$$\left(\int w\right)\left(\int wuv\right) - \left(\int wu\right)\left(\int wv\right) = \tfrac{1}{2}\iint w(x)w(y)\big(u(x)-u(y)\big)\big(v(x)-v(y)\big)\,dx\,dy \;\ge 0.$$

Proof. Expand the double integral: the four terms give $\frac{1}{2}\big[2(\int w)(\int wuv) - 2(\int wu)(\int wv)\big]$. Each factor $(u(x)-u(y))(v(x)-v(y))$ is nonnegative because u and v move together, so the double integral is nonnegative, strictly so if both are strictly monotone and $w > 0$ on a set of positive measure. The discrete statement is the same computation with sums. □

That identity is the whole of Putnam 2025 B2. Taking $w = f$, $u = x$, $v = f$ on $[0,1]$ with $f > 0$ increasing,

$$\left(\int f\right)\left(\int xf^2\right) \ge \left(\int xf\right)\left(\int f^2\right) \qquad\Longrightarrow\qquad \frac{\int xf^2}{\int f^2} \;\ge\; \frac{\int xf}{\int f},$$

which says the centroid of the solid of revolution lies to the right of the centroid of the plane region. The generalisation to extract: $x_p = \int xf^p / \int f^p$ is strictly increasing in p for strictly increasing positive f, by the same identity with $w = f^p$ and $v = f^{q-p}$. A one variable proof of the same fact is worth carrying as a fallback: if $w > 0$, ϕ is increasing and $\int w\phi = 0$, then ϕ changes sign exactly once, at some x_0, so $(x - x_0)\phi(x) \ge 0$ pointwise and therefore $\int xw\phi \ge x_0 \int w\phi = 0$. Apply it with $\phi = f - c$, $c = \int f^2 / \int f$.

Theorem 8.17.2 (Rearrangement)**.** *For $a_1 \le \cdots \le a_n$ and $b_1 \le \cdots \le b_n$ and any permutation σ,*

$$\sum_i a_i b_{n+1-i} \;\le\; \sum_i a_i b_{\sigma(i)} \;\le\; \sum_i a_i b_i.$$

Proof. Exchange argument. If σ is not the identity there are $i < j$ with $\sigma(i) > \sigma(j)$; swapping their images changes the sum by $(a_j - a_i)(b_{\sigma(i)} - b_{\sigma(j)}) \ge 0$, so the sum does not decrease. Finitely many swaps sort σ. □

Chebyshev's sum inequality is rearrangement averaged over all n cyclic shifts, which is the cleanest way to remember that the two are one theorem.

Schur. For $x, y, z \ge 0$ and $t > 0$,

$$x^t(x-y)(x-z) + y^t(y-z)(y-x) + z^t(z-x)(z-y) \;\ge\; 0.$$

Proof: assume without loss $x \ge y \ge z$ by symmetry; the first two terms combine to $(x-y)\big[x^t(x-z) - y^t(y-z)\big] \ge 0$ and the third is $z^t(x-z)(y-z) \ge 0$. At $t = 1$ this expands to $x^3 + y^3 + z^3 + 3xyz \ge xy(x+y) + yz(y+z) + zx(z+x)$, which is the form that closes symmetric three variable problems that AM and GM cannot reach.

Majorisation and Muirhead. Say α majorises β if, after sorting both decreasingly, the partial sums of α dominate those of β and the totals agree. Karamata: if α majorises β and F is convex then $\sum F(\alpha_i) \ge \sum F(\beta_i)$. Muirhead: for nonnegative reals, the symmetric mean $[\alpha] = \frac{1}{n!}\sum_\sigma \prod x_{\sigma(i)}^{\alpha_i}$ satisfies $[\alpha] \ge [\beta]$ exactly when α majorises β. Both are proved by *bunching*: a single majorisation

step moves two coordinates toward each other, and the two variable case of the inequality is elementary; iterate. Neither appears in the filtered corpus, so learn them after everything above, but learn the word "majorises", because it is the correct name for the smoothing arguments that do appear.

Symmetric function inequalities. Maclaurin: for $x_i \geq 0$ with $S_k = e_k/\binom{n}{k}$, $S_1 \geq S_2^{1/2} \geq \cdots \geq S_n^{1/n}$, equality throughout iff all x_i are equal. This is the generalisation showing that *any* Vieta coefficient, not only the product, bounds the arithmetic mean of the moduli from below, which is the extension of the AM and GM step in Putnam 2019 A3. Newton: if $\sum a_k x^k$ has all real roots then $p_k^2 \geq p_{k-1}p_{k+1}$ for $p_k = a_k/\binom{n}{k}$. Putnam 2014 B4 is the warning attached: Newton's inequalities are necessary and not sufficient for real rootedness, so verifying them can never prove real rootedness. Thirty seconds on them buys a sanity check and nothing else. Know which of your tools are one way.

One more, small and constantly useful: for real t_j, $\left(\sum t_j\right)^2 = \sum t_j^2 + 2e_2(t)$, so $e_1(t) = 0$ forces $e_2(t) = -\frac{1}{2}\sum t_j^2 \leq 0$, with equality only if every $t_j = 0$. That is the contradiction closing Putnam 2020 B5.

- **WHEN** two weighted means of the same variable with different weights are compared **DO** apply Chebyshev's inequality via the symmetrised double integral identity **BECAUSE** the symmetrised form has a manifestly signed integrand whenever the two functions are monotone in the same direction, so the sign is read rather than computed
- **WHEN** a sum $\sum a_i b_{\sigma(i)}$ is to be optimised over permutations **DO** sort both sequences the same way for the maximum and oppositely for the minimum **BECAUSE** any inversion can be swapped out at no cost by the exchange argument, so the sorted arrangement is optimal by a finite descent
- **WHEN** a symmetric three variable inequality resists AM and GM, especially with equality at both $x = y = z$ and at a boundary point **DO** try Schur **BECAUSE** Schur's equality cases are exactly $x = y = z$ and two equal with the third zero, so it is the only standard inequality with the right equality set
- **WHEN** a symmetric sum inequality has both sides of the same degree with different exponent patterns **DO** compare the exponent vectors for majorisation and quote Muirhead **BECAUSE** Muirhead is exactly the statement that the symmetric mean is monotone under majorisation, so the check is a partial sum comparison and not an algebraic search

8.18 Structural moves: SOS, normalisation, smoothing, extremal principle

Sums of squares. The certificate $x^2 + y^2 + z^2 - xy - yz - zx = \frac{1}{2}\left[(x-y)^2 + (y-z)^2 + (z-x)^2\right]$ is the model: a positive semidefinite quadratic form is an explicit sum of squares, and its zero set is the common zero set of the squares, which hands you the equality case for free. Putnam 2006 B1 uses precisely this to identify a unique real point. The general SOS method for symmetric inequalities is to write the difference as $S_a(b-c)^2 + S_b(c-a)^2 + S_c(a-b)^2$ and then argue about the signs of the coefficients S_a, S_b, S_c.

Positivity certificates on a box. To prove a polynomial P is nonnegative on $[-1,1]^2$, exhibit it as a positive combination of products of squares with the atoms $1-z$, $1+z$, $1-y$, $1+y$. Putnam 2013 A6 is exactly this, and the case split there is forced: zy is the one product of atoms whose sign the box does not determine, so it is the only thing you may have to case on.

Completing the square with a parameter. For $a \neq 0$, $by - ay^2 = \frac{b^2}{4a} - a\left(y - \frac{b}{2a}\right)^2$. Putnam 2006 B5 is one line of this. Maximise $I(f) - J(f) = \int_0^1 \left(x^2 f(x) - x f(x)^2\right)dx$: pointwise,

$$x^2 f - x f^2 = x\left(\tfrac{x^2}{4} - \left(f - \tfrac{x}{2}\right)^2\right) \leq \tfrac{x^3}{4},$$

so $I - J \leq \int_0^1 \frac{x^3}{4} dx = \frac{1}{16}$, attained at $f(x) = x/2$. The licence for that move is the pointwise maximisation principle: if $F(f) = \int \Phi(x, f(x)) dx$ contains no derivative of f, then $F(f) \leq \int \sup_y \Phi(x, y)\, dx$, with equality whenever the pointwise argmax is single valued and admissible. That principle replaces the entire calculus of variations here; recognising that you may decline the Euler and Lagrange equation $\frac{d}{dx}\Phi_{f'} = \Phi_f$ is the point, and you can only decline what you can name.

Normalisation and homogenisation. If an inequality is homogeneous of degree 0 under $x_i \mapsto \lambda x_i$, impose a normalisation ($\sum x_i = 1$, or $\max x_i = 1$) and you have removed a dimension for free. If it is not homogeneous, use the constraint to homogenise it: multiply the lower degree side by the constraint raised to the power that equalises degrees. Both directions are the same observation, that a constraint and a scaling are interchangeable.

Smoothing and the exchange argument. Fix all but two coordinates, and study the objective as a function of one of them with their sum fixed. For an objective $\sum f(x_i)$ under a linear constraint, this reduces local optimality to the sign of the constant curvature of $g(y) = f(y) + f(s - y)$: positive curvature forces the optimum to a boundary of the two variable slice, negative curvature forces the two coordinates equal. Putnam 2018 A3 is exactly this, and the reason it is needed there is that global convexity fails, so a *local* pairwise test replaces it. Two companion habits from that problem: when a reduction step produces a smaller instance of the same extremal problem, restate the claim for every relevant size before inducting; and before assuming that more resource helps, check the monotonicity, which there is $\partial_b W = 8a^3/b^3 > 0$.

The extremal principle. Before optimising, prove a maximiser exists: a continuous function on a compact set attains its bounds. Then argue about the maximiser's properties. Without the existence step, "let x be the maximiser" is an unlicensed assumption; with it, every structural property you derive is a genuine constraint. When the domain is not compact, either compactify it (bound the variables using the constraint) or replace the maximum by a supremum and work with an approximating sequence.

Best constants as extremal ratios. For continuous $F > 0$ and G on an open interval, the largest a with $aF \leq G$ is $\inf(G/F)$ and the smallest b with $G \leq bF$ is $\sup(G/F)$, where the infimum and supremum must include the limits at endpoints where F vanishes, and the extremal constant is admissible even when the infimum is a limit that is never attained. This converts "find the largest a" into two facts about one function, as in Putnam 2025 A2. The companion protocol, from Putnam 2023 A3: an extremal existence question has exactly two halves, a construction attaining the claimed value and a proof that nothing smaller works, and you should allocate effort by noticing that the construction is usually one line and the lower bound carries the marks.

Optimising a free scale parameter. A bound of the form $A/L + BL/2$ over $L > 0$ is minimised at $L = \sqrt{2A/B}$ with value $\sqrt{2AB}$; more generally $A/L^a + BL^b$ is minimised at $L \propto (aA/(bB))^{1/(a+b)}$ with value proportional to $A^{b/(a+b)} B^{a/(a+b)}$. This is AM and GM in disguise and it is how Putnam 2019 A6 converts a Taylor bound into a scale free estimate: the optimisation is what tells you the correct window length is proportional to x^r and not to x. Whenever you introduce an arbitrary length, radius, or cutoff, do not choose it, carry it and optimise at the end.

Two more that cost one line each. The averaging bound: if nonnegative reals $g_1, \ldots, g_k$ sum to T then $\min_i g_i \leq T/k$ (Putnam 2006 B2). And the epsilon squeeze: if $X \leq T + \varepsilon$ and $X \geq T - \varepsilon$ for every $\varepsilon > 0$ then $X = T$, and the two bounds may come from entirely different constructions, which is what fuses two unrelated estimates into an equality in Putnam 2016 B5.

- **WHEN** a symmetric expression must be shown nonnegative with a known equality point **DO** write it as an explicit sum of squares vanishing exactly there **BECAUSE** the zero set of a sum of squares is the common zero set of the squares, so the certificate proves the inequality and identifies the equality case simultaneously
- **WHEN** an integral functional contains no derivative of the unknown function **DO** maximise the integrand pointwise in $f(x)$ for each fixed x and check the argmax is admissible **BECAUSE**

the integral of a pointwise bound is a bound on the integral, and equality holds as soon as the pointwise optimiser is an admissible function, so the calculus of variations is unnecessary

- **WHEN** an inequality is homogeneous **DO** normalise, typically $\sum x_i = 1$ **BECAUSE** scaling and constraint are interchangeable, so normalising removes one variable at no cost
- **WHEN** an extremal problem resists a global convexity argument **DO** fix all but two variables with their sum constant and test the curvature of the two variable slice **BECAUSE** local optimality at a maximiser is a genuine constraint, and the sign of the slice curvature classifies each pair as equal or at a boundary
- **WHEN** you introduce an arbitrary window length, cutoff or radius **DO** carry it as a free parameter and optimise at the end by AM and GM **BECAUSE** the optimal scale balances the two competing error terms, and choosing the scale in advance guarantees a suboptimal exponent
- **WHEN** the question asks for the largest or smallest constant **DO** write the constant as inf or sup of a ratio, including endpoint limits, and then supply a construction and a bound **BECAUSE** that is the definition of the best constant, and an extremal question is by definition two claims, not one

8.19 Integral analogues and functional optimisation

Every inequality above has an integral form, obtained by replacing $\sum \lambda_i$ with $\int d\mu$ for a probability measure. The discrete case is the integral case for a measure on finitely many points, so each family is one theorem in two notations. Putnam problems here usually state the integral form and are solved by a discrete instinct.

Three functional optimisation principles that are integral specific.

The bathtub principle. For a fixed measure m and a decreasing kernel ϕ, the integral $\int_S \phi$ over sets S of measure m is maximised by taking S to be an initial segment of the domain and minimised by a terminal segment. What depends on the kernel is the numeric bound; the optimal *set* depends only on the constraint and the direction of monotonicity, so if the kernel is replaced by another decreasing one the optimal set is unchanged, and if the monotonicity flips so does the set.

The bang bang principle. A linear functional $f \mapsto \int f\phi$, maximised over $|f| \le 1$ subject to one linear integral constraint, attains its maximum at an extreme point of that constraint set, and the extreme points are exactly the functions taking values ± 1 almost everywhere. So the search may be restricted to sign valued step functions, and this is a proof, not a convenience.

Example 8.19.1 (Putnam 2014 B2)**.** $-1 \le f \le 1$ on $[1,3]$ with $\int_1^3 f = 0$; maximise $\int_1^3 f(x)/x\,dx$. Bang bang plus bathtub name the candidate before any verification: $g = +1$ on $[1,2]$ and -1 on $(2,3]$, the switch at 2 being forced by $\int g = 0$. To verify, use the mean zero hypothesis by insertion: since $\int (f-g) = 0$, for any constant λ,

$$\int_1^3 \frac{f-g}{x}\,dx = \int_1^3 (f-g)\left(\frac{1}{x} - \lambda\right) dx.$$

Choose $\lambda = \frac{1}{2}$, the value of $1/x$ at the switch. On $[1,2]$, $\frac{1}{x} - \frac{1}{2} \ge 0$ and $f - g = f - 1 \le 0$; on $[2,3]$ both signs reverse. So the integrand is nonpositive throughout and $\int f/x \le \int g/x = \ln 2 - \ln \frac{3}{2} = \ln \frac{4}{3}$.

That insertion step is a circuit in its own right, and Putnam 2017 A3 is the same move: if $\int (f-g) = 0$ then $\int \phi(f-g) = \int (\phi - c)(f-g)$ for every constant c, and choosing c well converts an unsigned integrand into a manifestly signed one. There, with $I_n = \int f^{n+1}/g^n$ and $f > g > 0$ of

equal integral,

$$I_{n+1} - I_n = \int \frac{f^{n+1}}{g^{n+1}}(f-g) = \int \left(\frac{f^{n+1}}{g^{n+1}} - 1\right)(f-g) > 0,$$

since both factors are positive. The subtraction of 1 is not cosmetic; it is what makes the sign visible. Learn to look for the constant that centres the integrand.

The same problem carries a reusable log convexity circuit: applying Cauchy and Schwarz to the square roots of f^{n+1}/g^n and f^{n+3}/g^{n+2} gives $I_n I_{n+2} \geq I_{n+1}^2$, so $\ln I_n$ is convex in n and the ratios I_{n+1}/I_n are nondecreasing. Whenever a sequence is an integral of a fixed factor times a quantity raised to the n th power, Cauchy and Schwarz between exponents n and $n+2$ against the midpoint always yields log convexity. That is a general route to monotonicity and growth rates for families of integrals.

- **WHEN** a hypothesis says an integral of a difference is zero **DO** insert a constant, $\int \phi(f-g) = \int (\phi - c)(f-g)$, and choose c so that the two factors change sign at the same point **BECAUSE** the zero integral makes the insertion free, and a product of two factors with a common sign change is single signed, which turns an estimate into an observation
- **WHEN** you must optimise a linear functional of f under a pointwise bound and one integral constraint **DO** restrict to ± 1 valued functions and place the $+1$ region where the kernel is largest **BECAUSE** the maximum of a linear functional on a convex set is at an extreme point, and the extreme points of that constraint set are exactly the sign valued functions
- **WHEN** a family of integrals is indexed by an exponent n **DO** apply Cauchy and Schwarz between exponents n and $n+2$ to get $I_n I_{n+2} \geq I_{n+1}^2$ **BECAUSE** that is log convexity of the sequence, which immediately gives monotonicity of consecutive ratios and hence growth rates

8.20 The recursion file

Recursion is a standing order, so here is every recursive shape that appears in the calculus and inequality items, stated in general form with its solution method and its recognition cue. Read this section as a list of families, not of problems.

1. Reduction formulae are linear recursions in the exponent. If integration by parts returns $I_k = r(k) I_{k-d}$ with r rational, iterate down to the d base cases; the answer is a ratio of products of arithmetic progressions, hence a ratio of Gamma functions. Wallis, $I_k = \frac{k-1}{k} I_{k-2}$, is the model (Putnam 2006 A6, 2009 B4).

2. A constant recursion of the form $c_{k+1} = c_k/(a + bc_k)$ is linearised by reciprocals. Taking reciprocals gives $1/c_{k+1} = a/c_k + b$, an affine recursion, solved in closed form. Putnam 2018 A5 produces $c_{k+1} = c_k/(1 + c_k)$ with $c_1 = 1$, hence $1/c_{k+1} = 1/c_k + 1$ and $c_k = 1/k$. Recognition cue: the constant in a bootstrap improves by a fixed rational function of itself at each round.

3. An operator applied n times to a fixed seed preserves a small family. Find the smallest finite dimensional family of functions the operator preserves, calibrate the coefficients on the first explicit case, and propagate by induction using the operator's own defining identity. In Putnam 2008 B2 the operator is integration from 0, the seed is $\ln x$, the preserved family is spanned by $x^n \ln x$ and x^n, and the answer is $F_n(x) = (\ln x - H_n)x^n/n!$. The same playbook covers iterated derivatives, iterated finite differences, and linear recurrences solved in an eigenbasis of characteristic roots.

4. A two term inequality between consecutive members of a family is a convexity recursion. Putnam 2019 A6 is the sharpest instance: applying Landau's inequality $\sup|f'| \leq 2\sqrt{\sup|f| \cdot \sup|f''|}$ to the $(k-1)$ th derivative gives $2L_k \leq L_{k-1} + L_{k+1} + \ln 4$ for $L_k = \ln \sup|f^{(k)}|$, which is midpoint convexity up to a constant. Solve it by interpolating between the endpoints L_0 and L_n, and you have Kolmogorov's interpolation inequality. General lesson: whenever consecutive members satisfy $2a_k \leq a_{k-1} + a_{k+1} + C$, interpolate rather than iterate.

5. Log convexity from Cauchy and Schwarz. $I_n I_{n+2} \geq I_{n+1}^2$ for any family $I_n = \int \psi \chi^n$, giving monotone consecutive ratios (Putnam 2017 A3).

6. Monotone transfer along a recursion, and lower solutions. If $a_{n+1} = f(a_n)$ with f increasing on the relevant interval, and c_n satisfies $f(c_n) \geq c_{n+1}$, then $a_0 > c_0$ forces $a_n > c_n$ for all n. The comparison sequence c_n is called a lower solution, and the entire content of such an induction is finding one. In Putnam 2012 B4 the recursion is $a_{n+1} = a_n + e^{-a_n}$, the comparison is $c_n = \ln(n+1)$, and the tangent line inequality supplies $f(c_n) \geq c_{n+1}$. Companion habit from Putnam 2020 A3: when the true asymptotic is $x_n \sim An^{-1/p}$ and the orbit approaches from below, an induction proving $x_n \geq Cn^{-1/p}$ can only close if $C < A$; the surplus is what survives after the leading terms cancel.

7. Telescoping by design. To bound $\sum_n w_n (a_1 \cdots a_n)^{1/n}$ by $C \sum_k a_k$, insert positive weights c_k before applying AM and GM, exchange the order of summation, and *choose* the c_k so that the coefficient of a_k is independent of k. The resulting constant is then sharp, because the same choice makes AM and GM tight along $a_k \propto 1/c_k$. Putnam 2021 B2, with $w_n = n2^{-n}$: taking $c_k = 4^k$ gives $\prod_{j \leq n} c_j = 4^{n(n+1)/2}$, so $(a_1 \cdots a_n)^{1/n} \leq 2^{-(n+1)} \frac{1}{n} \sum_{k \leq n} 4^k a_k$, and

$$\sum_{n \geq 1} \frac{n}{2^n} (a_1 \cdots a_n)^{1/n} \leq \sum_k 4^k a_k \sum_{n \geq k} 2^{-2n-1} = \sum_k 4^k a_k \cdot \tfrac{2}{3} 4^{-k} = \tfrac{2}{3} \sum_k a_k.$$

Equality needs $4^k a_k$ constant, so $a_k = 3 \cdot 4^{-k}$ under $\sum a_k = 1$, and the maximum is $2/3$. The parent theorem is Carleman's inequality, $\sum_n (a_1 \cdots a_n)^{1/n} \leq e \sum_k a_k$, proved the same way with $c_k = (k+1)^k / k^{k-1}$ so that $\prod_{j \leq n} c_j = (n+1)^n$ telescopes, giving the coefficient $c_k \sum_{n \geq k} \frac{1}{n(n+1)} = c_k / k = (1 + 1/k)^k < e$. Knowing the *proof* of Carleman turns the Putnam problem into a five minute exercise; quoting only its statement gives the far weaker bound $e/2$. That is the strongest argument in this chapter for learning proofs rather than statements.

8. Flows and iterations are the same recursion. The flow of a vector field satisfies $\Psi_{s+t} = \Psi_s \circ \Psi_t$, the continuous analogue of $f^{m+1} = f \circ f^m$. In both cases the method is to find coordinates in which the map is a translation: solve $\phi' \cdot (\text{field}) = 1$ for a flow, or Abel's equation $\phi(f(u)) = \phi(u) + 1$ for an iteration, after which $\phi(f^m(u)) = \phi(u) + m$. Putnam 2011 A5 is the flow version; Putnam 2008 A4 is the iteration version, where $\frac{d}{dx} \ln^k x = (x \ln^1 x \cdots \ln^{k-1} x)^{-1}$ makes $1/h$ an exact derivative on each escape time block. The convolution semigroup $P_s * P_t = P_{s+t}$ for the Cauchy kernel $P_t(x) = t/(\pi(t^2 + x^2))$ in Putnam 2011 B5 is the same relation for kernels; the Gaussian analogue is that variances add.

9. Continuous shadows of discrete recursions. A recursion $y_{n+1} = y_n + g(y_n)$ with a small autonomous increment has the continuous shadow $y' = g(y)$, which separates as $\int dy/g(y) = x + C$. This predicts the growth rate, and more usefully names the substitution $F(y) = \int dy/g(y)$ in which the flow has unit speed. In Putnam 2006 B6, separating $y' = y^{-1/k}$ produces the exponent $(k+1)/k$, which is both the shape of the answer and the substitution used in the proof. Recognition cue: an increment that depends only on the current value.

10. Self-similar bounds along an orbit. If $g(\psi(n)) \leq g(n)^K$ for a polynomial ψ, iterate along $N_j = \psi^{\circ j}(n)$ to get polynomial growth of g on that subsequence, contradicting superpolynomial growth. Putnam 2025 B6 uses $\psi(n) = n^4$; the official alternative uses $\psi(n) = 2^n$. Recognition cue: an inequality relating g at n to g at a composite argument.

11. Asymptotic recursions in a continuous index. For $f(r) = \int_0^b x^r w(x) dx$, consecutive members satisfy $f(r)/f(r+1) \to 1/b$, and integration by parts gives exact relations between f and g at shifted indices (Putnam 2011 A3). Treat a continuous parameter exactly as you treat n.

12. Telescoping increments with index dependent steps. In Putnam 2023 B4 the heights satisfy $h_m = h_{m-1} + \frac{m}{2} s_m^2$, a first order recursion whose increment depends only on the current index and step, so it telescopes to $h_0 + \sum_k$, turning the objective into a separable sum. Whenever the increment does not involve the accumulated value, do not iterate: telescope.

13. Chains carried whole through an induction. To show a nonlinear recursion escapes zero, carry the entire chain of previous inequalities as the induction hypothesis, bounding the product term below

and the subtracted term above by the same quantity. Putnam 2018 B4: from $1 = |x_0| \le |x_1| \le \cdots \le |x_n|$ and $x_{n+1} = 2x_n x_{n-1} - x_{n-2}$ one gets $|x_{n+1}| \ge 2|x_n||x_{n-1}| - |x_{n-2}| \ge |x_n|(2|x_{n-1}| - 1) \ge |x_n|$, extending the chain.

14. Self-similar block decompositions. If a set meets every window of g consecutive integers and contains 1, then for $0 < r < 1$, $\sum_{n \in S} r^n \ge r + r^{g+1}/(1 - r^g)$, because the tail after each block of length g is a scaled copy of the whole (Putnam 2015 A4). Recognition cue: a constraint that repeats identically on every window.

15. Products of many factors. $(f_1 f_2 \cdots f_n)'$ is the n term sum in which each summand replaces one factor by its derivative. Whenever a system of n coupled equations, each multiplied by the complementary product of the other $n - 1$ functions, sums to that expansion, the full product decouples the system (Putnam 2009 A2). The multiplicative analogue is the logarithmic derivative: $(\ln f)' = \sum f_m'/f_m$ is additive, and $f''/f = (\ln f)'' + ((\ln f)')^2$ (Putnam 2023 A1).

- **WHEN** a bootstrap improves a constant by a fixed rational function of itself **DO** write the constant recursion and linearise it, usually by reciprocals **BECAUSE** an affine recursion has a closed form, so you get the limiting constant in one line instead of iterating and guessing
- **WHEN** an operator is applied n times to a fixed seed **DO** find the smallest family the operator preserves and induct with the operator's defining identity **BECAUSE** the family is finite dimensional, so the whole orbit is described by finitely many coefficient sequences, each satisfying a simple recursion
- **WHEN** you must prove $a_n > c_n$ for a recursively defined a_n **DO** check f is increasing and verify only $f(c_n) \ge c_{n+1}$ **BECAUSE** monotone transfer carries the inequality forward automatically, so the entire induction reduces to one inequality about the comparison sequence
- **WHEN** you are inserting weights before an inequality **DO** leave them symbolic, impose that the final coefficient be independent of the index, and solve for them **BECAUSE** that condition is exactly the condition for the equality case to be attainable simultaneously at every index, which is what makes the constant sharp
- **WHEN** a recursion has an increment depending only on the current value **DO** write the separable differential equation $y' = g(y)$ as its continuous shadow **BECAUSE** the shadow predicts the growth exponent and names the substitution in which the increment becomes constant

8.21 What is rare, and what to learn last

Order your revision by frequency, and be honest about the tail.

Learn first, because they fire in almost every paper: reflection and symmetrisation; AM and GM with its equality case; Cauchy and Schwarz in all three faces; tangent line and chord bounds for e^x, $\ln(1 + x)$ and sin; Jensen; differentiating under the integral sign; Fubini and the Jacobian; parts chosen to kill a boundary term. Learn second: Hölder with matched exponents, Minkowski, power means, Chebyshev, Beta and Gamma, reduction formulae, endpoint concentration, the transport formula, the extreme point principle, the reciprocal as an integral.

Learn last, and know that you are learning them for completeness rather than for expected value. Frullani's integral and principal values do not appear anywhere in the filtered corpus; they are standard vocabulary and cost one page each, but no Putnam problem in the harvested set needed them. Muirhead and Karamata do not appear either, although the smoothing arguments they formalise do, so learn the vocabulary and the bunching proof rather than a list of exponent patterns. Schur's inequality does not appear in the corpus; it is cheap insurance for symmetric three variable problems and nothing more. Linear programming duality and complementary slackness appear once, in Putnam 2014 A4, and the vertex argument of Putnam 2023 B4 covers most of what

they would give you. Quadrature rules and degrees of exactness appear once, in Putnam 2016 A6, where the whole problem is choosing the rule: the moment equations $\sum_i w_i x_i^m = 1/(m+1)$ for $m = 0, \ldots, n$ define the search space, symmetric node sets gain one degree of exactness for free (which is why Simpson is exact on cubics), and the sharp constant is $1 - w_0$ where w_0 is the weight at the prescribed root. Newton's inequalities appear once, in Putnam 2014 B4, purely as a route to be declined. The Euler and Lagrange equation appears once, in Putnam 2006 B5, and there too the correct move is to recognise it in order to decline it.

That last pattern is worth naming, because it recurs: a substantial fraction of the value of an advanced technique on the Putnam is the ability to identify, in ten seconds, that it is the nominal route and that a shorter structural route exists. You cannot decline a method you cannot name. That is why the exotica are in this chapter at all.

Chapter 9

Geometry and vectors

9.1 What this chapter is for

A-level and Further Maths hand you vectors as a notation. You get column vectors in two and three dimensions, the scalar product as a formula with a $\cos\theta$ attached, the vector product as a determinant mnemonic you were told to memorise, the vector and Cartesian equations of lines and planes, and a small menu of distance formulas. What you were never given is the thing that actually decides Putnam problems: the *definitions* these objects satisfy, stated so that the properties fall out of them rather than being memorised alongside them, and stated in a way that does not break when the ambient dimension stops being three.

That is the gap, and here is the test. Why does $\mathbf{u} \cdot \mathbf{v}$, computed as a sum of coordinate products, not depend on the orthonormal basis you used? Why does the cross product exist in three dimensions and in no other except seven, and why does it fail to be associative? What replaces det when you have k vectors in $\mathbb{R}^n$ with $k < n$? If any of those made you pause, the circuits below are not installed, and a Putnam geometry problem will find the gap in ten minutes.

By the end you will hold the dot and cross products from their definitions, volume in every dimension, lines and hyperplanes with every distance derived rather than recalled, the full transformation group, the classical plane geometry the Putnam still tests, conics, convexity up to Helly and extreme points, the regular solids with coordinates, integral geometry, the calculus of curves and surfaces, and the geometry of numbers.

Geometry is the fourth largest field in the archive. It is not a corner of the paper you can skip. Between 2006 and 2025 it carries A1 and B1 openers you must not drop (Putnam 2006 A1, Putnam 2006 B1, Putnam 2009 A1, Putnam 2012 A1, Putnam 2017 B1, Putnam 2021 A1), and it carries some of the hardest problems on the paper (Putnam 2007 A6, Putnam 2013 A5, Putnam 2018 A6, Putnam 2023 A4). The openers are won by having the definition to hand. The hard ones are won by knowing which structural object the configuration secretly is.

9.2 Vectors, and the three faces of the dot product

Vectors as objects, not as columns

The column of numbers is not the vector. It is what the vector looks like after you have chosen a basis, and every time you write one you have spent a choice. That matters because half of the geometry problems are solved by choosing the basis last, or by never choosing one.

Two roles get confused because they wear the same notation. A *point* is a location; a *displacement* is a difference of two points. You may subtract points, and add displacements to points, but you may not add two points. The one apparent exception, $\lambda A + (1-\lambda)B$, is legitimate precisely because it rewrites as $A + (1-\lambda)(B-A)$, a point plus a displacement. That is the whole of barycentric coordinates, and it is why centroids, incentres and convex combinations are one species of object.

Definition 9.2.1 (Affine combination and convex combination)**.** Given points $P_1, \dots, P_m$ and reals $\lambda_1, \dots, \lambda_m$ with $\sum \lambda_i = 1$, the point $\sum \lambda_i P_i$ is well defined independently of the origin and is called an affine combination. If in addition every $\lambda_i \geq 0$ it is a convex combination.

Why it is well defined. Moving the origin by $\mathbf{d}$ replaces each P_i by $P_i - \mathbf{d}$, and $\sum \lambda_i (P_i - \mathbf{d}) = \sum \lambda_i P_i - \mathbf{d}$, the same point described from the new origin. The condition $\sum \lambda_i = 1$ is exactly what makes the correction one copy of $\mathbf{d}$. □

- **WHEN** a problem gives you points and a linear relation among their position vectors **DO** check whether the coefficients sum to 1 before you do anything else **BECAUSE** if they do, the relation is origin free and you may place the origin wherever it kills the most algebra; if they do not, the relation is secretly about displacements and the origin is already pinned
- **WHEN** you are asked to prove a statement about an arbitrary triangle or tetrahedron **DO** put the origin at the centroid, at a vertex, or at the circumcentre, whichever makes one of the hypotheses read $\sum P_i = 0$ or $|P_i| = R$ **BECAUSE** a coordinate choice is free, and every free choice you decline to make costs you ink later

The dot product, three characterisations

Here are the three definitions you will meet. They are all the same function.

Definition 9.2.2 (Dot product)**.** For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$:

(A) *Algebraic.* $\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} u_i v_i$, the coordinates taken in an orthonormal basis.

(B) *Trigonometric.* $\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}|\,|\mathbf{v}| \cos\theta$, where $\theta \in [0, \pi]$ is the angle between them.

(C) *Metric, or polarisation.* $\mathbf{u} \cdot \mathbf{v} = \frac{1}{2}\big(|\mathbf{u}|^2 + |\mathbf{v}|^2 - |\mathbf{u} - \mathbf{v}|^2\big)$.

Proposition 9.2.3. *(A), (B) and (C) define the same function.*

Proof. (A) implies (C): expand $|\mathbf{u} - \mathbf{v}|^2 = \sum (u_i - v_i)^2 = |\mathbf{u}|^2 - 2\sum u_i v_i + |\mathbf{v}|^2$ and rearrange. (C) implies (B): the cosine rule applied to the triangle with sides $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{u} - \mathbf{v}$ says $|\mathbf{u} - \mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2 - 2|\mathbf{u}||\mathbf{v}| \cos\theta$; substitute and rearrange. The cosine rule is Euclidean geometry you already own, so nothing is circular here provided you take (C) as the definition and (A) as a computation. □

Remark 9.2.4. Take (C) as the definition and one irritation disappears. Formula (A) looks basis dependent; formula (C) mentions only lengths; since they agree, (A) cannot depend on the orthonormal basis chosen. That is the reason orthogonal matrices preserve dot products, and it is worth repeating: *the dot product is a function of distances alone*. Whenever a problem gives distances and asks for angles, or the reverse, polarisation is the bridge, and it is the only bridge you need.

Putnam 2021 A3 puts points on a sphere with all pairwise distances equal. Polarisation converts that into equal dot products in one line: $|P_i - P_j|^2 = 2R^2 - 2P_i \cdot P_j$, so the edge length is constant exactly when $P_i \cdot P_j$ is constant, and the problem becomes a question about a Gram matrix.

- **WHEN** a hypothesis is about squared distances or about equal lengths **DO** expand every squared distance by polarisation into norms and dot products **BECAUSE** squared distance is a quadratic form and dot products are its bilinear shadow; the expansion turns metric hypotheses into linear algebra, which has determinants and rank and therefore has leverage
- **WHEN** a hypothesis is about perpendicularity **DO** write it as a vanishing dot product and then, if it helps, as an equality of two squared distances via polarisation **BECAUSE** $\mathbf{u} \cdot \mathbf{v} = 0$ and $|\mathbf{u} - \mathbf{v}| = |\mathbf{u} + \mathbf{v}|$ are the same statement, and one of the two forms is always the cheaper one to combine with the rest of the hypotheses

Projection

Definition 9.2.5 (Orthogonal projection onto a line)**.** For $\mathbf{v} \neq \mathbf{0}$, the projection of $\mathbf{u}$ onto the line $\mathbb{R}\mathbf{v}$ is

$$\operatorname{proj}_{\mathbf{v}}(\mathbf{u}) = \frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{v}|^2}\,\mathbf{v}.$$

The number $\mathbf{u}\cdot\mathbf{v}/|\mathbf{v}|$ is the signed length of the shadow of $\mathbf{u}$ on the direction of $\mathbf{v}$. The defining property, the one to go back to, is that $\mathbf{u} - \operatorname{proj}_{\mathbf{v}}(\mathbf{u})$ is orthogonal to $\mathbf{v}$: verify by dotting with $\mathbf{v}$ and watching the two terms cancel. Everything follows from that one property. Splitting $\mathbf{u} - t\mathbf{v}$ into the orthogonal leftover plus a multiple of $\mathbf{v}$ and applying Pythagoras shows at once that $|\mathbf{u} - t\mathbf{v}|$ is minimised exactly at $t = \mathbf{u}\cdot\mathbf{v}/|\mathbf{v}|^2$, so the projection is the nearest point of the line.

Projection onto a subspace W with orthonormal basis $\mathbf{w}_1, \dots, \mathbf{w}_k$ is $P_W\mathbf{u} = \sum_i (\mathbf{u}\cdot\mathbf{w}_i)\mathbf{w}_i$, and the same orthogonality argument gives the same minimising property. Projection onto the hyperplane $\mathbf{u}^\perp$ is $\mathbf{x} \mapsto \mathbf{x} - (\mathbf{x}\cdot\mathbf{u})\mathbf{u}$ for unit $\mathbf{u}$; this is the map that appears in Putnam 2013 A5.

Cauchy and Schwarz as a statement about angles

Theorem 9.2.6 (Cauchy and Schwarz)**.** $|\mathbf{u}\cdot\mathbf{v}| \le |\mathbf{u}|\,|\mathbf{v}|$*, with equality exactly when* $\mathbf{u}$ *and* $\mathbf{v}$ *are linearly dependent.*

Proof. If $\mathbf{v} = \mathbf{0}$ both sides vanish. Otherwise set $t_0 = \mathbf{u}\cdot\mathbf{v}/|\mathbf{v}|^2$ and expand

$$0 \le |\mathbf{u} - t_0\mathbf{v}|^2 = |\mathbf{u}|^2 - \frac{(\mathbf{u}\cdot\mathbf{v})^2}{|\mathbf{v}|^2},$$

which rearranges to the inequality. Equality forces $\mathbf{u} = t_0\mathbf{v}$. □

The proof *is* the projection: Cauchy and Schwarz asserts that the leftover after projecting has non negative length. Geometrically it says $\cos\theta \in [-1, 1]$, which is the entire content. The triangle inequality follows by squaring and the reverse triangle inequality $\big||\mathbf{u}| - |\mathbf{v}|\big| \le |\mathbf{u} - \mathbf{v}|$ by substitution. That reverse form with its equality case is the engine of Putnam 2010 B2: $\big|\,|AC| - |BC|\,\big| \le |AB|$ with equality exactly when A, B, C are collinear, so strict inequality is noncollinearity and a distance bound converts into an equality of side lengths.

- **WHEN** a hypothesis says three points satisfy a distance equation with no slack **DO** test it against the reverse triangle inequality **BECAUSE** equality in the reverse triangle inequality is exactly collinearity, so a tight distance hypothesis is a hidden collinearity hypothesis

9.3 The cross product, and why it lives only in three dimensions

The definition that explains everything

Do not define the cross product by the mnemonic with $\mathbf{i}, \mathbf{j}, \mathbf{k}$ in the top row. That is a memory aid, not a definition, and it explains none of the properties. Define it by what it is for.

Definition 9.3.1 (Cross product)**.** For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$, the vector $\mathbf{u} \times \mathbf{v}$ is the unique vector satisfying

$$(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} = \det\begin{bmatrix}\mathbf{u} & \mathbf{v} & \mathbf{w}\end{bmatrix} \qquad \text{for every } \mathbf{w} \in \mathbb{R}^3,$$

where the determinant is of the matrix with those three columns.

Existence and uniqueness need one sentence: $\mathbf{w} \mapsto \det[\mathbf{u}\ \mathbf{v}\ \mathbf{w}]$ is linear, and every linear functional on $\mathbb{R}^n$ is $\mathbf{w} \mapsto \mathbf{a}\cdot\mathbf{w}$ for exactly one $\mathbf{a}$ (take a_i to be its value on $\mathbf{e}_i$). So $\mathbf{u} \times \mathbf{v}$ represents the functional "signed volume of the box obtained by adjoining $\mathbf{w}$", and everything falls out of that.

Proposition 9.3.2. *For all* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ *and* $\lambda \in \mathbb{R}$*:*

1. $\mathbf{u} \times \mathbf{v}$ *is orthogonal to both* $\mathbf{u}$ *and* $\mathbf{v}$*;*
2. *the product is bilinear and antisymmetric, so* $\mathbf{v} \times \mathbf{u} = -\mathbf{u} \times \mathbf{v}$ *and* $\mathbf{u} \times \mathbf{u} = \mathbf{0}$*;*
3. $\mathbf{u} \times \mathbf{v} = \mathbf{0}$ *exactly when* $\mathbf{u}, \mathbf{v}$ *are linearly dependent;*
4. $|\mathbf{u} \times \mathbf{v}|^2 = |\mathbf{u}|^2|\mathbf{v}|^2 - (\mathbf{u} \cdot \mathbf{v})^2$*, hence* $|\mathbf{u} \times \mathbf{v}| = |\mathbf{u}||\mathbf{v}| \sin\theta$*;*
5. *in coordinates* $\mathbf{u} \times \mathbf{v} = (u_2v_3 - u_3v_2,\ u_3v_1 - u_1v_3,\ u_1v_2 - u_2v_1)$.

Proof. (1) Put $\mathbf{w} = \mathbf{u}$: the determinant has a repeated column, so it vanishes, so $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{u} = 0$. Same with $\mathbf{v}$. (2) The determinant is multilinear and alternating in its columns, and the correspondence functional to vector is linear, so these properties transfer. (3) If $\mathbf{u}, \mathbf{v}$ are dependent every such determinant vanishes; if they are independent, extend to a basis $\mathbf{u}, \mathbf{v}, \mathbf{w}$ and the determinant is nonzero, so the vector is nonzero. (5) Read off the coordinates: the ith coordinate of $\mathbf{u} \times \mathbf{v}$ is $\det[\mathbf{u}\ \mathbf{v}\ \mathbf{e}_i]$, which is the stated 2×2 cofactor. (4) Verify the identity in coordinates using (5): both sides expand to $\sum_{i<j}(u_iv_j - u_jv_i)^2$, which is Lagrange's identity. Then substitute $\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}||\mathbf{v}| \cos\theta$ and use $1 - \cos^2\theta = \sin^2\theta$ with $\sin\theta \geq 0$ on $[0, \pi]$. □

Property (4) is the norm law and it is load bearing. In Putnam 2022 B2 it manufactures a geometric progression of norms while orthogonality of the output makes every later sine factor 1, and finiteness kills the progression. In Putnam 2010 A5 orthogonality and non vanishing manufacture a mutually orthogonal triple $\mathbf{b}, \mathbf{c}, \mathbf{b} \times \mathbf{c}$ inside a set only assumed closed under an operation.

Area and orientation

Since $|\mathbf{u} \times \mathbf{v}| = |\mathbf{u}||\mathbf{v}| \sin\theta$, it is the area of the parallelogram spanned by $\mathbf{u}$ and $\mathbf{v}$, and half of it is the area of the triangle.

Definition 9.3.3 (Vector area)**.** For $P, Q, R \in \mathbb{R}^3$ the vector area of the triangle PQR is $\mathbf{V} = \frac{1}{2}(Q - P) \times (R - P)$.

Its magnitude is the area and it vanishes exactly when the points are collinear, so the degenerate case needs no separate handling. Its direction is the normal, oriented by the vertex order. Vector areas add over a triangulated surface and sum to $\mathbf{0}$ over any closed surface, because interior edges are traversed twice with opposite orientations. This is the object Putnam 2013 A5 is really about; writing area as $|\mathbf{V}|$ rather than as a positive square root of a symmetric expression is what removes the sign ambiguity of the unit normal.

Proposition 9.3.4 (Area of a shadow)**.** *Let* π *be orthogonal projection of* $\mathbb{R}^3$ *onto the plane* $\mathbf{u}^\perp$ *for a unit vector* $\mathbf{u}$*. Then for any triangle* T *with vector area* $\mathbf{V}$*,*

$$\operatorname{Area}(\pi(T)) = |\mathbf{V} \cdot \mathbf{u}| = \operatorname{Area}(T)\,|\mathbf{n} \cdot \mathbf{u}|,$$

where $\mathbf{n}$ *is either unit normal of* T*.*

Proof. Write $\mathbf{a} = Q - P$, $\mathbf{b} = R - P$, so $\pi(\mathbf{a}) = \mathbf{a} - (\mathbf{a} \cdot \mathbf{u})\mathbf{u}$ and likewise for $\mathbf{b}$. Expanding and using $\mathbf{u} \times \mathbf{u} = \mathbf{0}$,

$$\pi(\mathbf{a}) \times \pi(\mathbf{b}) = \mathbf{a} \times \mathbf{b} - (\mathbf{b} \cdot \mathbf{u})(\mathbf{a} \times \mathbf{u}) - (\mathbf{a} \cdot \mathbf{u})(\mathbf{u} \times \mathbf{b}).$$

Both $\pi(\mathbf{a})$ and $\pi(\mathbf{b})$ lie in $\mathbf{u}^\perp$, so their cross product is parallel to $\mathbf{u}$; dot the display with $\mathbf{u}$ and note that the last two terms die because $(\mathbf{a} \times \mathbf{u}) \cdot \mathbf{u} = 0$. Hence $\pi(\mathbf{a}) \times \pi(\mathbf{b}) = \big((\mathbf{a} \times \mathbf{b}) \cdot \mathbf{u}\big)\mathbf{u}$ and the projected area is $\frac{1}{2}|(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{u}| = |\mathbf{V} \cdot \mathbf{u}|$. □

- **WHEN** a problem mentions shadows, projections or the area of a figure seen from a direction **DO** write the figure's vector area $\mathbf{V}$ and the shadow area as $|\mathbf{V} \cdot \mathbf{u}|$ **BECAUSE** projection is linear, so the only thing that survives it is the component of the vector area along the viewing direction; every other route drags in an unsigned normal and forces case analysis

- **WHEN** you must prove an inequality about areas of triangles in space **DO** convert each area to $|\mathbf{V}_i|$ and look for a way to linearise the norm **BECAUSE** norms are not polynomial in the data, and the averaging identity of Section 11 is the standard device for turning $|\mathbf{V}|$ into an average of the linear quantities $\mathbf{V} \cdot \mathbf{u}$

Non associativity, and the vector triple product

The cross product is not associative, and that is a structural fact to exploit rather than a defect to apologise for. With $\mathbf{i}, \mathbf{j}, \mathbf{k}$ the standard basis, $(\mathbf{i} \times \mathbf{i}) \times \mathbf{j} = \mathbf{0} \times \mathbf{j} = \mathbf{0}$, whereas $\mathbf{i} \times (\mathbf{i} \times \mathbf{j}) = \mathbf{i} \times \mathbf{k} = -\mathbf{j}$. The two bracketings differ.

The exact failure is measured by the vector triple product.

Theorem 9.3.5 (Vector triple product)**.** *For all* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$,

$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w})\,\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\,\mathbf{w}.$$

Proof. Two vectors are equal exactly when their dot products with every $\mathbf{x}$ agree, so test against an arbitrary $\mathbf{x}$. Using the definition and then the cyclic symmetry of the determinant,

$$\big(\mathbf{u} \times (\mathbf{v} \times \mathbf{w})\big) \cdot \mathbf{x} = \det[\mathbf{u}\ \ \mathbf{v} \times \mathbf{w}\ \ \mathbf{x}] = (\mathbf{v} \times \mathbf{w}) \cdot (\mathbf{x} \times \mathbf{u}).$$

Now apply the Binet and Cauchy identity $(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c})$, which is Lagrange's identity polarised and is checked by expanding coordinates. With $\mathbf{a} = \mathbf{v}$, $\mathbf{b} = \mathbf{w}$, $\mathbf{c} = \mathbf{x}$, $\mathbf{d} = \mathbf{u}$ this equals $(\mathbf{v} \cdot \mathbf{x})(\mathbf{w} \cdot \mathbf{u}) - (\mathbf{v} \cdot \mathbf{u})(\mathbf{w} \cdot \mathbf{x})$, which is exactly $\big((\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w}\big) \cdot \mathbf{x}$. □

Read it: $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ lies in the span of $\mathbf{v}, \mathbf{w}$ while $(\mathbf{u} \times \mathbf{v}) \times \mathbf{w}$ lies in the span of $\mathbf{u}, \mathbf{v}$. Different planes, which is exactly why associativity fails. The substitute is the Jacobi identity

$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) + \mathbf{v} \times (\mathbf{w} \times \mathbf{u}) + \mathbf{w} \times (\mathbf{u} \times \mathbf{v}) = \mathbf{0},$$

which follows by writing each term with the triple product expansion and cancelling in pairs. In other words $(\mathbb{R}^3, \times)$ is a Lie algebra, not an associative algebra: no identity element, and every nonzero vector annihilates itself, so every nonzero vector is a zero divisor.

Putnam 2010 A5 is built precisely on that clash. A group law is associative and has an identity; the cross product has neither. The specialisation you need there is the one that follows by setting $\mathbf{v} = \mathbf{u}$ with $\mathbf{u} \cdot \mathbf{w} = 0$:

$$\mathbf{u} \times (\mathbf{u} \times \mathbf{w}) = -|\mathbf{u}|^2 \mathbf{w}.$$

Carry that as a picture: on $\mathbf{u}^\perp$ the map $L_{\mathbf{u}}(\mathbf{x}) = \mathbf{u} \times \mathbf{x}$ is a quarter turn about the axis $\mathbf{u}$ scaled by $|\mathbf{u}|$, since its square is $-|\mathbf{u}|^2$ times the identity, and its kernel is the line $\mathbb{R}\mathbf{u}$. That is what Putnam 2022 B2 and Putnam 2010 A5 use, and why iterating $\mathbf{w} \mapsto \mathbf{u} \times \mathbf{w}$ has period four up to scaling.

- **WHEN** a problem writes a triple cross product, or asks you to compare $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$ with $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ **DO** expand both with the triple product formula and read off which plane each lies in **BECAUSE** the expansion is the only identity in the subject that converts cross products back into dot products, and dot products are where all the hypotheses live
- **WHEN** a set is claimed to be closed under an operation that agrees with $\times$ **DO** look for the associativity clash and for the zero divisors **BECAUSE** $\times$ is a Lie bracket, so any group or ring structure imposed on it can only survive where the bracket degenerates, which is a very small set
- **WHEN** you see the iteration $\mathbf{w}_{k+1} = \mathbf{u} \times \mathbf{w}_k$ **DO** restrict to $\mathbf{u}^\perp$ and use $\mathbf{w}_{k+2} = -|\mathbf{u}|^2 \mathbf{w}_k$ **BECAUSE** the map is a scaled quarter turn on that plane, so the recursion has characteristic roots $\pm i|\mathbf{u}|$ and is periodic of period four up to the scale factor

Why three dimensions, and only three

Here is the honest reason, and it is visible in the definition. In $\mathbb{R}^n$, the map

$$\mathbf{w} \mapsto \det[\mathbf{v}_1 \;\; \mathbf{v}_2 \;\; \cdots \;\; \mathbf{v}_{n-1} \;\; \mathbf{w}]$$

is linear, so there is always a canonical product, but it takes $n-1$ inputs. The generalised cross product exists in every dimension as an $(n-1)$-ary operation. A *binary* product with these properties therefore requires $n-1=2$, that is $n=3$. In $\mathbb{R}^2$ the analogue is unary ($\mathbf{v} \mapsto$ the quarter turn of $\mathbf{v}$), and in $\mathbb{R}^4$ it takes three vectors.

One exception exists and you should know it by name. Requiring only bilinearity, orthogonality of the output to both inputs and the norm law, such a product exists exactly in dimensions 3 and 7; the seven dimensional one comes from imaginary octonion multiplication and fails the Jacobi identity. That is a standard classification result to quote, not reprove. The three dimensional product is the imaginary part of quaternion multiplication: for purely imaginary quaternions, $\mathbf{uv} = -(\mathbf{u}\cdot\mathbf{v}) + \mathbf{u}\times\mathbf{v}$, which packages both products into one associative operation and explains the norm law as multiplicativity of the quaternion norm. That is the background to Putnam 2022 B2.

- **WHEN** a problem in $\mathbb{R}^n$ with $n > 3$ tempts you to write a cross product **DO** stop and use either the wedge or, in practice, the Gram determinant of Section 4 **BECAUSE** no binary cross product exists there; the object you actually want is k dimensional volume, and the Gram determinant computes it in every dimension

9.4 Volume: triple products, Gram determinants, Cayley and Menger

The scalar triple product

By the definition of the cross product, $(\mathbf{u}\times\mathbf{v})\cdot\mathbf{w} = \det[\mathbf{u}\;\mathbf{v}\;\mathbf{w}]$, so the scalar triple product is the signed volume of the parallelepiped on $\mathbf{u}, \mathbf{v}, \mathbf{w}$. It is cyclic, alternating, and vanishes exactly when the three vectors are coplanar. The tetrahedron on P, Q, R, S has volume $\frac{1}{6}\left|\det[Q-P \;\; R-P \;\; S-P]\right|$, and in the plane the triangle on A, $A+\mathbf{u}$, $A+\mathbf{v}$ has area $\frac{1}{2}|\det[\mathbf{u}\;\mathbf{v}]|$.

The last statement is the one to reach for whenever the data are polynomial. Putnam 2018 A6 turns on exactly this: base times height and $\frac{1}{2}ab\sin C$ both introduce lengths and sines, which are not polynomial in squared distances, whereas the determinant is bilinear and alternating and therefore stays inside the ring you are working in.

- **WHEN** an area or volume appears in a problem whose hypotheses are algebraic (rationality, integrality, polynomial identities) **DO** write the area or volume as a determinant, never as base times height **BECAUSE** determinants are polynomial in the coordinates and multiplicative under composition, so they respect every algebraic hypothesis; trigonometric formulas do not

Gram determinants, and volume in n dimensions

Definition 9.4.1 (Gram matrix)**.** For $\mathbf{v}_1, \ldots, \mathbf{v}_k$ in a real inner product space, the Gram matrix is G with $G_{ij} = \mathbf{v}_i \cdot \mathbf{v}_j$. Its determinant is the Gram determinant.

Theorem 9.4.2. *Let M be the $n \times k$ matrix with columns $\mathbf{v}_i$. Then $G = M^{\mathsf{T}}M$, so G is symmetric positive semidefinite,* $\det G \geq 0$, *and* $\det G = 0$ *exactly when the $\mathbf{v}_i$ are linearly dependent. The k dimensional volume of the parallelepiped they span is $\sqrt{\det G}$, and the volume of the simplex on $\mathbf{0}, \mathbf{v}_1, \ldots, \mathbf{v}_k$ is $\sqrt{\det G}/k!$.*

Proof. $(M^{\mathsf{T}}M)_{ij} = \mathbf{v}_i \cdot \mathbf{v}_j$ by definition of matrix product, and $\mathbf{x}^{\mathsf{T}}G\mathbf{x} = |M\mathbf{x}|^2 \geq 0$ gives positive semidefiniteness, with equality exactly when $M\mathbf{x} = \mathbf{0}$, that is when the columns are dependent. For the volume, run Gram and Schmidt: replacing $\mathbf{v}_j$ by $\mathbf{v}_j$ minus a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}$ changes neither the volume (shearing) nor $\det G$ (the same column and row operations on G). After

the process G is diagonal with entries $|\mathbf{v}_j^*|^2$, so $\det G = \prod |\mathbf{v}_j^*|^2$, and the volume is $\prod |\mathbf{v}_j^*|$, base times height iterated k times. □

When $k = n$ this collapses to $\det G = (\det M)^2$, so a Gram determinant is the correct dimension free replacement for "the square of the determinant". This is why Gram matrices are all over the archive: Putnam 2018 A6, Putnam 2021 A3, Putnam 2011 A4 and Putnam 2011 B4 all reduce to a statement about $\det G$ or $\operatorname{rank} G$. For integer vectors G has integer entries and, in the square case, $\det G$ is a perfect square, which is a number theoretic hook that appears in Putnam 2021 A3.

One further identity, the mixed Gram determinant, is worth memorising because it is the load bearing step of Putnam 2018 A6: for $\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z} \in \mathbb{R}^2$,

$$\det[\mathbf{u}\ \mathbf{v}] \cdot \det[\mathbf{w}\ \mathbf{z}] = \det \begin{pmatrix} \mathbf{u}\cdot\mathbf{w} & \mathbf{u}\cdot\mathbf{z} \\ \mathbf{v}\cdot\mathbf{w} & \mathbf{v}\cdot\mathbf{z} \end{pmatrix},$$

because the right hand side is $\det\big([\mathbf{u}\ \mathbf{v}]^{\mathsf{T}}[\mathbf{w}\ \mathbf{z}]\big)$ and det is multiplicative. Note what this buys: a product of two *unknown* determinants becomes a determinant of *known* dot products. It holds only in the square case; for non square matrices the substitute is the Cauchy and Binet formula, which writes $\det(M^{\mathsf{T}}M')$ as a sum over k element row subsets of products of minors. That failure in the non square case is exactly why the planarity hypothesis in Putnam 2018 A6 cannot be dropped.

- **WHEN** a hypothesis says "all pairwise (squared) distances are rational or integral" **DO** translate it into "the Gram matrix of the difference vectors is rational or integral" and then use $\det G$ and $\operatorname{rank} G$ **BECAUSE** squared distances and dot products are polarisation images of each other, and the Gram matrix is the canonical object holding all of them at once
- **WHEN** you need the volume of a k dimensional object sitting inside $\mathbb{R}^n$ with $k < n$ **DO** compute $\sqrt{\det G}/k!$ rather than hunting for a normal vector **BECAUSE** the Gram determinant is defined without reference to the ambient dimension, so it works where cross products do not
- **WHEN** a configuration is claimed to be degenerate (points on a common sphere, in a common flat, or a regular simplex) **DO** compute $\det G$ and use the vanishing criterion **BECAUSE** $\det G = 0$ is precisely linear dependence, so degeneracy questions become one determinant evaluation

The Cayley and Menger determinant

When all you are given is the distances, this is the determinant that converts them into volume.

Theorem 9.4.3 (Cayley and Menger)**.** *For $P_0, \dots, P_k \in \mathbb{R}^k$ with $d_{ij} = |P_iP_j|$, the k dimensional volume V of the simplex they span satisfies*

$$V^2 = \frac{(-1)^{k+1}}{2^k (k!)^2} \det \begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & d_{01}^2 & \cdots & d_{0k}^2 \\ 1 & d_{10}^2 & 0 & \cdots & d_{1k}^2 \\ \vdots & & & \ddots & \vdots \\ 1 & d_{k0}^2 & d_{k1}^2 & \cdots & 0 \end{pmatrix}.$$

The determinant vanishes exactly when the points lie in a flat of dimension less than k.

For $k = 2$ this is Heron's formula as $16K^2 = -\det$, and for $k = 3$ the identity of Piero della Francesca, $288V^2 = \det$. Do not plan to reconstruct the general proof in the arena: state it, recognise when it applies, and derive the $k = 2$ case from the Gram determinant if you need it. Carry it because it is *the* canonical function of squared distance data, so a problem phrasing its hypothesis in squares of distances (Putnam 2018 A6) is signalling this object.

- **WHEN** the hypotheses are entirely about squared distances and the conclusion is about area or volume **DO** name the Cayley and Menger determinant and either use it or use the Gram determinant it is equivalent to **BECAUSE** the setter phrased the hypothesis in squares because squares are what the determinant eats; that is a planted guiding light, not a coincidence

9.5 Lines, planes and hyperplanes

The two forms, and when to use which

A line in $\mathbb{R}^n$ is $\{A + t\mathbf{d}\}$ with $\mathbf{d} \neq \mathbf{0}$: a point plus a direction. A hyperplane is $\{\mathbf{x} : \mathbf{x} \cdot \mathbf{n} = p\}$ with $\mathbf{n} \neq \mathbf{0}$: the set on which one linear functional is constant, equivalently a translate of the orthogonal complement $\mathbf{n}^\perp$ of dimension $n - 1$. In $\mathbb{R}^2$ it is a line, in $\mathbb{R}^3$ a plane; the word exists so you stop rewriting the same lemma once per dimension.

The rule for choosing between the forms is mechanical. Parametric form is for *producing* points: substitute and you generate the object. Implicit form is for *testing* points: substitute and you get a number whose sign and size mean something. Most stalled geometry problems are stalled because the solver is testing with a parametric form or producing with an implicit one.

- **WHEN** you must decide whether a given point lies on, above or below a line or plane **DO** use the implicit form $\mathbf{x} \cdot \mathbf{n} - p$ and read the sign **BECAUSE** the implicit form is a single scalar function whose sign is the side and whose size, divided by $|\mathbf{n}|$, is the distance; the parametric form gives you neither without solving equations
- **WHEN** you must intersect an object with a line, or substitute a curve into a condition **DO** use the parametric form and reduce to one variable **BECAUSE** this is what turns a two dimensional incidence question into a one variable polynomial, which is the backbone of Putnam 2007 A1

Distances and angles, all derived from projection

Every distance formula here is one application of projection, so do not memorise a list.

The distance from P to the hyperplane $\mathbf{x} \cdot \mathbf{n} = p$ is $|P \cdot \mathbf{n} - p|/|\mathbf{n}|$: write $P = Q + t\mathbf{n}$ with Q on the hyperplane, dot with $\mathbf{n}$ to get $t|\mathbf{n}|^2 = P \cdot \mathbf{n} - p$, and the distance is $|t||\mathbf{n}|$.

The distance from P to the line $A + t\mathbf{d}$ is $|\mathbf{w} - \operatorname{proj}_{\mathbf{d}} \mathbf{w}|$ where $\mathbf{w} = P - A$, which equals $\sqrt{|\mathbf{w}|^2 - (\mathbf{w} \cdot \hat{\mathbf{d}})^2}$ in every dimension and equals $|\mathbf{w} \times \mathbf{d}|/|\mathbf{d}|$ in $\mathbb{R}^3$ by the norm law.

Two lines in $\mathbb{R}^3$ that are neither parallel nor intersecting are *skew*, and the distance between $A_1 + s\mathbf{d}_1$ and $A_2 + t\mathbf{d}_2$ is

$$\frac{\left|(A_2 - A_1) \cdot (\mathbf{d}_1 \times \mathbf{d}_2)\right|}{|\mathbf{d}_1 \times \mathbf{d}_2|},$$

because $\mathbf{d}_1 \times \mathbf{d}_2$ is the unique direction orthogonal to both, so the shortest connecting segment is parallel to it and its length is the projection of any connecting vector on that direction. Notice this is the parallelepiped volume divided by its base area, that is the height: the formula is not a new fact. Angles follow the same pattern, one dot product each: between two lines take their directions, between two hyperplanes their normals, between a line and a hyperplane the complement of the angle between direction and normal.

Incidence, and the arithmetic of intersections

Two distinct lines in the plane either are parallel and disjoint, or meet in exactly one point; equivalently their direction vectors are dependent or independent. Putnam 2017 B1 is phrased so that both directions of the required proof are translations between "intersect" and "not parallel", so the dichotomy is used explicitly rather than assumed.

The arithmetic version matters more than it looks. Over any field K, two distinct K points determine a line $ax + by = c$ with coefficients in K, and two non parallel K lines meet at a K point by Cramer's rule. With the perpendicular bisector that gives the load bearing lemma of Putnam 2008 B1.

Proposition 9.5.1 (Rational rigidity of a circle's centre)**.** *If three noncollinear rational points lie on a circle, the centre is rational.*

Proof. The centre is equidistant from all three by the definition of a circle, so it lies on the perpendicular bisector of each of the two chords. The perpendicular bisector of PQ is the locus $\{X : |XP| = |XQ|\}$, which expands to the linear equation $2X \cdot (Q - P) = |Q|^2 - |P|^2$; its coefficients are built from the coordinates of P, Q by addition, subtraction and multiplication, so they are rational. The two bisectors are non parallel because P, Q, R are noncollinear, so they meet at exactly one point, which Cramer's rule places in $\mathbb{Q}^2$. □

Corollary 9.5.2. *A circle whose centre is not rational contains at most two rational points.*

Dimension free version: in $\mathbb{R}^n$, if $n + 1$ affinely independent rational points lie on a sphere the centre is rational, because the n bisector hyperplanes are rational and independent. Affine independence is not decoration; without it the points lie in a lower flat and the centre is not pinned.

The perpendicular bisector deserves its own line in the Circuit Bank because it is the definition of a circle in disguise. A circle is the set of points at fixed distance from a centre, so *the centre is the point of interest* whenever a circle is mentioned, and the bisector is the machine that produces it. Putnam 2024 B2 uses this twice: the centre O is the meet of the perpendicular bisectors of the two diagonals, and $OB = OE$ then puts O on the perpendicular bisector of the new diagonal. Putnam 2025 B1 uses the uniqueness half: three points of a circle are noncollinear (a line meets a circle at most twice) and three noncollinear points lie on exactly one circle, so the circumcentre of any three points of Γ is the centre of Γ.

State these cold. Two circles with centres at distance $d > 0$ and radii ρ, σ meet twice when $|\rho - \sigma| < d < \rho + \sigma$, once at either equality, never otherwise, and a point is determined by its distances to the two centres up to reflection in the line of centres: that is the counting engine of Putnam 2025 B1. A sphere of radius r meets a plane at distance δ from its centre in a circle of radius $\sqrt{r^2 - \delta^2}$, so a cross section radius never exceeds r: the bound driving Putnam 2012 B2.

Finally, a parametrisation that is dimension free and therefore under used: any circle of radius r centred at P in $\mathbb{R}^n$ is $\{P + \mathbf{v}_1 \cos\theta + \mathbf{v}_2 \sin\theta\}$ with $\mathbf{v}_1 \perp \mathbf{v}_2$ and $|\mathbf{v}_1| = |\mathbf{v}_2| = r$. Putnam 2008 B3 starts here, and the ambient dimension never enters.

- **WHEN** a circle appears anywhere in the hypotheses **DO** name its centre and write the perpendicular bisector conditions **BECAUSE** a circle is by definition a set of points equidistant from a centre, so the centre is the only point that all the hypotheses can talk about at once
- **WHEN** the problem has an arithmetic hypothesis (rational, integral) attached to a geometric configuration **DO** check which constructions stay inside the field: midpoints, perpendicular bisectors, quarter turns of vectors and intersections of non parallel lines all do; taking square roots does not **BECAUSE** the field is closed under the four operations, so any construction assembled from them is rationality preserving, and the ones that are not are exactly where the counterexamples live
- **WHEN** you need a circle inside $\mathbb{R}^n$ for $n > 3$ **DO** use $P + \mathbf{v}_1 \cos\theta + \mathbf{v}_2 \sin\theta$ with $\mathbf{v}_1 \perp \mathbf{v}_2$ of equal length **BECAUSE** this is the definition of a circle as a round orbit in a 2 plane and it does not mention the ambient dimension at all

9.6 Transformations: isometries, similarities, affine and projective

Isometries are affine, and why

Theorem 9.6.1 (Classification of isometries). *If $f : \mathbb{R}^n \to \mathbb{R}^n$ satisfies $|f(\mathbf{x}) - f(\mathbf{y})| = |\mathbf{x} - \mathbf{y}|$ for all $\mathbf{x}, \mathbf{y}$, then $f(\mathbf{x}) = Q\mathbf{x} + \mathbf{b}$ with Q orthogonal.*

Proof. Set $\mathbf{b} = f(\mathbf{0})$ and $g = f - \mathbf{b}$, so g is an isometry with $g(\mathbf{0}) = \mathbf{0}$; hence $|g(\mathbf{x})| = |\mathbf{x}|$. By polarisation,

$$g(\mathbf{x}) \cdot g(\mathbf{y}) = \tfrac{1}{2}\big(|g\mathbf{x}|^2 + |g\mathbf{y}|^2 - |g\mathbf{x} - g\mathbf{y}|^2\big) = \tfrac{1}{2}\big(|\mathbf{x}|^2 + |\mathbf{y}|^2 - |\mathbf{x} - \mathbf{y}|^2\big) = \mathbf{x} \cdot \mathbf{y}.$$

So g preserves dot products. Applying this to an orthonormal basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ shows $g(\mathbf{e}_1), \dots, g(\mathbf{e}_n)$ is again orthonormal, hence a basis, and then for any $\mathbf{x}$ the coordinates of $g(\mathbf{x})$ in that basis are $g(\mathbf{x}) \cdot g(\mathbf{e}_i) = \mathbf{x} \cdot \mathbf{e}_i = x_i$. Therefore $g(\mathbf{x}) = \sum_i x_i\, g(\mathbf{e}_i)$, which is linear, with matrix Q whose columns are orthonormal. □

Reconstruct that proof once by hand: *distance preservation forces dot product preservation by polarisation, and dot product preservation forces linearity*. No continuity, no hand waving about rigid motions.

An orthogonal matrix satisfies $Q^{\mathsf{T}}Q = I$, has orthonormal columns and $\det Q = \pm 1$; determinant 1 is a rotation, $SO(n)$. All eigenvalues have modulus 1 and the non real ones come in conjugate pairs, so if n is odd and $\det Q = 1$ then 1 is an eigenvalue and a rotation of $\mathbb{R}^3$ has an axis (Euler's rotation theorem), while if 1 is not an eigenvalue the multiplicity of -1 has the parity of n. Putnam 2019 B3 is built on that parity statement.

In $\mathbb{R}^3$ the rotation by θ about a unit axis $\mathbf{k}$ is Rodrigues' formula

$$R\mathbf{v} = \mathbf{v}\cos\theta + (\mathbf{k} \times \mathbf{v})\sin\theta + \mathbf{k}(\mathbf{k} \cdot \mathbf{v})(1 - \cos\theta),$$

which is only the decomposition of $\mathbf{v}$ into its fixed component along $\mathbf{k}$ and its rotated component in $\mathbf{k}^\perp$, using that $\mathbf{x} \mapsto \mathbf{k} \times \mathbf{x}$ is the quarter turn of that plane. Its trace is $1 + 2\cos\theta$, which you need in Section 13.

Reflections

The reflection in the hyperplane $\mathbf{u}^\perp$, for a unit vector $\mathbf{u}$, is the Householder matrix

$$P = I - 2\mathbf{u}\mathbf{u}^{\mathsf{T}}, \qquad P\mathbf{w} = \mathbf{w} - 2(\mathbf{u} \cdot \mathbf{w})\mathbf{u}.$$

It is symmetric, orthogonal, an involution, fixes $\mathbf{u}^\perp$ pointwise, sends $\mathbf{u}$ to $-\mathbf{u}$ and has determinant -1. Every orthogonal matrix is a product of at most n Householder reflections, the minimum being $n - \dim\ker(Q - I)$. Putnam 2019 B3 uses all of this: $P^{-1} = P$ turns $PQ\mathbf{w} = \mathbf{w}$ into $Q\mathbf{w} = P\mathbf{w}$, and $P\mathbf{w} - \mathbf{w} \in \operatorname{span}(\mathbf{u})$ forces the eigenvector to be $(Q - I)^{-1}\mathbf{u}$.

In the plane, reflection in a line is the same object one dimension down, and on the Putnam it is almost always an invariance argument rather than a computation. Putnam 2024 B2: the reflection in a line ℓ fixes ℓ pointwise, preserves all distances and maps each circle centred on ℓ to itself, which delivers $CE = AD$, $AE = CD$ and $OE = OD$ in one sentence. Putnam 2007 A1: if two curves are exchanged by a reflection, their union and their intersection set are invariant, so intersection points either lie on the axis or come in mirror pairs, which organises the entire search. Putnam 2015 A1: an isometry multiplies every area by 1, so if it fixes a curve setwise and exchanges two regions bounded by that curve, those regions have equal area with no integration at all.

- **WHEN** a configuration has a mirror symmetry, or you can normalise it into one **DO** use invariance instead of computing **BECAUSE** an isometry preserves lengths, angles and areas exactly, so any quantity built from those is forced to agree across the mirror; computing both sides is the zombie route

Homotheties, similarities and affine maps

A *dilation* or *homothety* with centre P and ratio $\lambda \neq 0$ is $h(X) = P + \lambda(X - P)$: the vector from P to $h(X)$ is λ times the vector from P to X. It sends any line not through P to a parallel line, fixes lines through P setwise, multiplies lengths by $|\lambda|$ and areas by λ^2, and homotheties with a common centre compose by multiplying ratios, so the inverse is the ratio $1/\lambda$ map at the same centre. Putnam 2017 B1 states exactly that defining equation as its hypothesis, so the whole problem becomes a statement about how one dilation interacts with two lines, and the inverse map recovers A_1 from A_2. Putnam 2016 B3 uses the size adjustment: a homothety of ratio slightly above 1 about an interior point of a convex region produces a strictly larger similar region containing it, upgrading "area at most 4" into a covering triangle of area exactly 4.

A *similarity* of ratio $k > 0$ satisfies $|f(\mathbf{x}) - f(\mathbf{y})| = k|\mathbf{x} - \mathbf{y}|$; dividing by k reduces it to an isometry, so every similarity is a homothety composed with an isometry.

An *affine map* is $\mathbf{x} \mapsto M\mathbf{x} + \mathbf{b}$ with M invertible. It preserves collinearity, parallelism, midpoints, ratios along a line, centroids and convexity, and multiplies every area by $|\det M|$, so it preserves *ratios* of areas even while destroying the areas. This is the most under used tool in the chapter.

- **WHEN** a statement to be proved involves only ratios of lengths along lines, ratios of areas, parallelism, midpoints, centroids or convexity **DO** apply an affine map that sends the configuration to the most convenient one (any triangle to the equilateral triangle, any ellipse to the unit circle, any parallelogram to the unit square) **BECAUSE** all of those notions are affine invariants and any triangle can be sent to any other by an affine map, so proving the special case proves the general one; this is the cheapest reroute in geometry and it costs almost no ink
- **WHEN** the problem involves an ellipse **DO** map it to a circle by the affine map $(x, y) \mapsto (x/a, y/b)$ and solve the circle problem **BECAUSE** an ellipse is by definition an affine image of a circle, and the quantities you are asked about are usually affine invariant

Putnam 2015 A1 uses the area preserving members of this family explicitly: the squeeze maps $(x, y) \mapsto (\lambda x, y/\lambda)$ have determinant 1 and fix each hyperbola $xy = c$ setwise, and together with the reflection $(x, y) \mapsto (y, x)$ they let you move two points of one branch into mirror position across $y = x$. That normalisation is what converts a two point problem into a one point problem, and the equal area conclusion then follows from the reflection alone.

Projective and Möbius maps, and the Cayley transform

A projective map of the line is $x \mapsto \frac{ax+b}{cx+d}$ with $ad - bc \neq 0$. Composition corresponds to multiplication of the matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ up to a scalar, so the order of the map equals the order of the matrix in PGL_2. That is an explanation rather than a verification, and it is what Putnam 2016 A3 needs: for $\sigma(x) = 1 - 1/x = \frac{x-1}{x}$ the matrix is $\left(\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ with trace 1 and determinant 1, so its characteristic equation is $\lambda^2 - \lambda + 1 = 0$ whose roots are primitive sixth roots of unity; the matrix therefore has order 6, its cube is $-I$, and the map has order 3. The six functions

$$x, \quad \frac{1}{x}, \quad 1 - x, \quad \frac{1}{1-x}, \quad \frac{x-1}{x}, \quad \frac{x}{x-1}$$

form the anharmonic group of order six under composition, isomorphic to the symmetric group on three letters, generated by $x \mapsto 1/x$ and $x \mapsto 1 - x$, and permuting a cross ratio. Its unique subgroup of order three, x, $\frac{x-1}{x}$, $\frac{1}{1-x}$, is the one in Putnam 2016 A3.

The *Cayley transform* is the bridge between rotations and skew symmetric matrices, and it is the matrix version of the tangent half angle substitution you already use for integrals.

Proposition 9.6.2 (Cayley transform). *If A is real skew symmetric ($A^{\mathsf{T}} = -A$) then $I + A$ is invertible and $Q = (I - A)(I + A)^{-1}$ is orthogonal with $\det Q = 1$ and without eigenvalue -1. Conversely every such Q arises this way from $A = (I - Q)(I + Q)^{-1}$.*

Proof. If $A\mathbf{v} = -\mathbf{v}$ then $\mathbf{v}^{\mathsf{T}}A\mathbf{v} = -|\mathbf{v}|^2$, but skewness forces $\mathbf{v}^{\mathsf{T}}A\mathbf{v} = 0$, so $\mathbf{v} = \mathbf{0}$ and $I + A$ is invertible. Since $I - A$ and $(I + A)^{-1}$ commute (they are polynomials in A and its inverse),

$$Q^{\mathsf{T}}Q = (I - A)^{-1}(I + A)(I - A)(I + A)^{-1} = (I - A)^{-1}(I - A)(I + A)(I + A)^{-1} = I.$$

Also $\det(I - A) = \det\big((I - A)^{\mathsf{T}}\big) = \det(I + A)$, so $\det Q = 1$; and $Q + I = 2(I + A)^{-1}$ is invertible, so -1 is not an eigenvalue. □

In two dimensions, taking $A = \left(\begin{smallmatrix} 0 & -t \\ t & 0 \end{smallmatrix}\right)$ gives

$$Q = \frac{1}{1+t^2}\begin{pmatrix} 1-t^2 & 2t \\ -2t & 1-t^2 \end{pmatrix},$$

which is the rotation matrix with $\cos\theta = \frac{1-t^2}{1+t^2}$ and $\sin\theta = \frac{2t}{1+t^2}$, that is with $t = \tan(\theta/2)$. So the Cayley transform is the Weierstrass substitution wearing a matrix.

- **WHEN** a problem asks you to parametrise rotations, or to produce rotations with rational entries **DO** use the Cayley transform on a skew symmetric matrix with rational entries **BECAUSE** it is a rational bijection between skew symmetric matrices and rotations without eigenvalue -1, so rationality passes straight through, exactly as $t = \tan(\theta/2)$ rationalises the circle
- **WHEN** a substitution in a functional equation or an integral has small order (applying it two or three times returns the identity) **DO** write its 2×2 matrix and read the order off the characteristic polynomial **BECAUSE** order of the map equals the order of the matrix in PGL_2, so this replaces a verification by composition with a two line eigenvalue calculation and tells you the order in advance

Inversion

Inversion in the circle of centre A and radius ρ is $X \mapsto X^* = A + \rho^2\frac{X-A}{|X-A|^2}$: an involution on the plane minus A mapping circles and lines to circles and lines, and circles through A to lines not through A. The formula to carry is the distance law

$$|X^*Y^*| = \frac{\rho^2\,|XY|}{|AX|\,|AY|},$$

proved by noting $|AX|\,|AX^*| = |AY|\,|AY^*| = \rho^2$, so triangles AXY and AY^*X^* share the angle at A and have proportional adjacent sides, hence are similar. Inversion is the fastest honest proof of Ptolemy, below.

9.7 The classical plane geometry the Putnam still uses

Triangle centres and barycentric weights

Write $a = |BC|$, $b = |CA|$, $c = |AB|$, $s = (a+b+c)/2$, K the area, r the inradius, R the circumradius, and carry these as one block:

$$G = \tfrac{1}{3}(A+B+C), \qquad I = \frac{aA+bB+cC}{a+b+c}, \qquad K = rs, \qquad K^2 = s(s-a)(s-b)(s-c), \qquad \frac{a}{\sin\alpha} = 2R = \frac{abc}{2K}.$$

The incentre formula is not separate memory: the bisector from A meets BC at D with $|BD|/|DC| = c/b$, so $D = \frac{bB+cC}{b+c}$, and I divides AD in ratio $(b+c) : a$ from A, which assembles to the displayed weights.

The circumcentre O is the meet of the perpendicular bisectors. Placing O at the origin, define $H = A + B + C$. Then $H - A = B + C$ and $(B + C) \cdot (B - C) = |B|^2 - |C|^2 = 0$, so $AH \perp BC$: H lies on the altitude from A, and by symmetry on all three. Hence H is the orthocentre, and since $G = \frac{1}{3}H$, the points O, G, H are collinear with $|OG| : |GH| = 1 : 2$: the Euler line, in four lines from the definition of the circumcentre. The nine point circle, centre the midpoint of OH and radius $R/2$, is the companion fact; state it by name.

The structural statement worth the most is the one about weights.

Proposition 9.7.1 (Parallelism through barycentric weights)**.** *If $P = \alpha A + \beta B + \gamma C$ with $\alpha + \beta + \gamma = 1$, then the signed distance from P to the line AB is γh_c, where h_c is the distance from C to AB. Consequently PQ is parallel to AB exactly when P and Q have the same C weight.*

Proof. The signed distance to a fixed line is an affine function of the point; it vanishes at A and B and equals h_c at C, so on affine combinations it returns the C coefficient times h_c. □

Putnam 2019 A2 is exactly this. "IG parallel to AB" says the C weights of I and G agree, that is $\frac{c}{a+b+c} = \frac{1}{3}$, hence $a + b = 2c$; it also says $r = h_c/3$, since G sits at one third of the height. The rest is incircle bookkeeping: the tangent lengths are $s - a$, $s - b$, $s - c$; Heron with $K = rs$ gives $r^2 = \frac{(s-a)(s-b)(s-c)}{s}$; the half angle formula is $\tan(\alpha/2) = \frac{r}{s-a}$; and multiplying two of those,

$$\tan\tfrac{\alpha}{2}\tan\tfrac{\beta}{2} = \frac{r^2}{(s-a)(s-b)} = \frac{s-c}{s},$$

with $\sum \tan\frac{\alpha}{2}\tan\frac{\beta}{2} = 1$ over the three cyclic pairs. No step in that problem is anything but one of these lines.

- **WHEN** a problem names two triangle centres and a parallel or a ratio **DO** write both centres in normalised barycentric weights and compare the coefficient belonging to the opposite vertex **BECAUSE** signed distance to a side is an affine function returning that coefficient, so parallelism becomes one scalar equation among the side lengths with no coordinates
- **WHEN** a problem mentions the incircle, the inradius, or tangent lengths **DO** write down s, the tangent lengths $s - a, s - b, s - c$, $K = rs$ and Heron at once **BECAUSE** these four facts close the system: any two of r, K, s and the side lengths determine the rest, so the problem reduces to symmetric functions of a, b, c

Areas, cuts and the extremal triangle

Three area facts run through the archive.

First, $\text{area}(XQR) = \frac{1}{2}|QR| \cdot d(X, \text{line } QR)$, so for a fixed base, comparing areas is comparing distances to a line. Putnam 2016 B3 converts each maximality inequality $\text{area}(PBC) \le \text{area}(ABC)$ into a half plane membership this way, then uses the general statement worth carrying: *a convex polygon is the intersection of the half planes bounded by its edge lines*, so several half plane memberships collapse into one.

Second, the corner cut ratio. If a line meets the two sides through V at distances x, y from V, those sides having lengths u, v, then corner and whole share the included angle at V, so

$$\frac{\text{area of corner}}{\text{area of whole}} = \frac{xy}{uv}.$$

Equal areas is therefore $xy = uv/2$, trigonometry gone. The cut segment lies on the boundary of both pieces, so equal perimeters is $x + y = s$, independent of the cut length and of every angle. Putnam 2017 B5 rests on that sum and product pair: x, y are the roots of $T^2 - sT + \frac{uv}{2} = 0$.

Third, the medial triangle. If A, B, C are the midpoints of the sides of $A'B'C'$ then ABC is similar to $A'B'C'$ with ratio $1/2$ and $\text{area}(A'B'C') = 4\,\text{area}(ABC)$; equivalently each side of $A'B'C'$ runs

through a vertex of ABC parallel to the opposite side. That manufactures the covering triangle of area exactly 4 in Putnam 2016 B3.

A related identity for Putnam 2017 B5: a line bisecting both area and perimeter passes through the incentre. Decompose the cut off region into triangles with apex I: its area is $\frac{r}{2}(x+y) \pm \frac{d}{2} \cdot (\text{cut length})$ with d the distance from I to the line, and both bisection conditions force $d = 0$. Conversely a line through I bisects the area exactly when it bisects the perimeter, which is why such lines are rigid and why rotating a line about I proves existence at once.

- **WHEN** a problem compares areas of triangles that share a side **DO** convert to a comparison of distances from the third vertices to that side's line **BECAUSE** the shared base cancels, so the whole inequality becomes a half plane membership, and half plane memberships intersect into polygon memberships
- **WHEN** a line cuts a corner off a triangle and areas or perimeters are being bisected **DO** write x, y for the two cut distances and use $xy = uv/2$ and $x + y = s$ **BECAUSE** the included angle cancels in the area ratio and the cut length cancels in the perimeter comparison, so both conditions are pure relations among side lengths
- **WHEN** an extremal principle is available (largest area triangle, largest distance pair) in a finite configuration **DO** take the extremal object and draw through each of its vertices the line parallel to the opposite side **BECAUSE** maximality then says every other point lies on the near side of each of those three lines, which is exactly membership in the quadrupled triangle; this is the standard covering argument and it generalises to k dimensions with factor k^k

Circles: inscribed angles, power of a point, cyclic quadrilaterals

Theorem 9.7.2 (Inscribed angle)**.** *If B, C lie on a circle with centre O and A is another point of the circle on the major arc, then $\angle BAC = \frac{1}{2}\angle BOC$. Consequently all points on one arc see the chord under the same angle, and a quadrilateral $ABCD$ is cyclic exactly when its opposite angles sum to π.*

Proof. Take O inside the angle at A. Triangles OAB and OAC are isosceles since $|OA| = |OB| = |OC| = R$, so the exterior angle of OAB at O is $2\angle OAB$ and likewise for OAC; adding gives $\angle BOC = 2\angle BAC$. The other configurations are the same computation with a subtraction, which is why practitioners use directed angles modulo π and get every case at once. □

Definition 9.7.3 (Power of a point)**.** For a point P and a circle with centre O and radius R, the power of P is $\operatorname{pow}(P) = |PO|^2 - R^2$.

Proposition 9.7.4. *If a line through P with unit direction $\mathbf{d}$ meets the circle at $P + t_1\mathbf{d}$ and $P + t_2\mathbf{d}$, then $t_1 t_2 = \operatorname{pow}(P)$. Hence the signed product $\overline{PA} \cdot \overline{PB}$ is the same for every line through P, it equals the square of the tangent length when P is outside, and the locus of points of equal power with respect to two circles is a line, the radical axis.*

Proof. Substitute: $|P + t\mathbf{d} - O|^2 = R^2$ expands to $t^2 + 2t\,\mathbf{d} \cdot (P - O) + |PO|^2 - R^2 = 0$, a monic quadratic in t whose product of roots is the constant term. For the radical axis, subtract the two equations $|X|^2 - 2X \cdot O_i + |O_i|^2 - R_i^2$; the quadratic terms cancel and a linear equation in X remains. □

Theorem 9.7.5 (Ptolemy)**.** *For any four points A, B, C, D in the plane, $|AC|\,|BD| \le |AB|\,|CD| + |BC|\,|AD|$, with equality exactly when $ABCD$ are concyclic or collinear in that cyclic order.*

Proof. Invert at A with radius ρ. The distance law gives $|B^*C^*| = \frac{\rho^2 |BC|}{|AB||AC|}$ and likewise for the other pairs, so the triangle inequality $|B^*D^*| \le |B^*C^*| + |C^*D^*|$, multiplied through by $\frac{|AB||AC||AD|}{\rho^2}$, is exactly the stated inequality. Equality means B^*, C^*, D^* collinear with C^* between, and a line not through A is the inverse image of a circle through A, which is concyclicity in the right order. □

For quadrilateral area, Bretschneider's formula covers the general case: sides a, b, c, d in order, semiperimeter s, opposite angles α, γ,

$$K^2 = (s-a)(s-b)(s-c)(s-d) - abcd\cos^2\Big(\frac{\alpha+\gamma}{2}\Big),$$

and Brahmagupta's formula is the cyclic case $\alpha + \gamma = \pi$. Putnam 2024 B2 uses this family, together with the fact that with the sides fixed in order the quadrilateral is determined by one diagonal and the angles opposite that diagonal increase with it. It also needs the cosine rule in central angle form: for P on the circle of radius r_1 about O and Q on the circle of radius r_2, $|PQ|^2 = r_1^2 + r_2^2 - 2r_1r_2\cos\angle POQ$, so a known distance determines the central angle in $[0, \pi]$ and the polar angle difference up to sign, which makes the configuration count finite. Two further facts: the diagonals of a convex quadrilateral cross inside, hence are not parallel, hence their perpendicular bisectors meet in exactly one point (this is where that problem spends convexity, and it fails for self intersecting quadrilaterals); and an arc of angle $\theta \in [0, \pi]$ on a circle of radius R has chord $2R\sin(\theta/2)$, increasing in θ, so among minor arcs a longer chord means a longer arc, which gives the strip lemma of Putnam 2024 A5: the arc cut by a strip of width $w < 2R$ is shortest exactly when the strip is symmetric about the parallel diameter.

- **WHEN** two chords, secants or tangents through a common point appear **DO** write the power of the point and equate the products **BECAUSE** power of a point is a single number attached to P and the circle, independent of the line chosen, so every line through P gives the same equation and you may pick the cheapest
- **WHEN** a configuration asks you to prove four points are concyclic **DO** use equal inscribed angles (directed modulo π), or equal powers, or the equality case of Ptolemy **BECAUSE** concyclicity has three equivalent tests and the hypotheses will match exactly one of them cheaply

Ceva, Menelaus, shoelace, Pick

Theorem 9.7.6 (Ceva)**.** *Cevians AD, BE, CF of triangle ABC are concurrent if and only if $\frac{\overline{BD}}{\overline{DC}} \cdot \frac{\overline{CE}}{\overline{EA}} \cdot \frac{\overline{AF}}{\overline{FB}} = 1$ with signed ratios.*

Proof. Let P be the common point. Triangles ABD and ACD share the height from A, so $\frac{\overline{BD}}{\overline{DC}} = \frac{[ABD]}{[ACD]}$, and with P for A, $= \frac{[PBD]}{[PCD]}$. Subtracting numerators and denominators of equal ratios preserves the ratio, so it equals $\frac{[ABP]}{[ACP]}$, and the three cyclic versions multiply to a telescoping product equal to 1. The converse holds because the ratio determines D on line BC. □

Theorem 9.7.7 (Menelaus)**.** *Points D, E, F on lines BC, CA, AB are collinear if and only if the same signed product equals -1.*

Proof. Let ℓ be the transversal and let $\delta(X)$ be the signed distance from X to ℓ. The point D on segment BC with $D \in \ell$ satisfies $\frac{\overline{BD}}{\overline{DC}} = -\frac{\delta(B)}{\delta(C)}$, because δ is affine along the line and vanishes at D. Multiplying the three cyclic versions telescopes to -1. □

The sign is the entire difference: $+1$ for concurrence, -1 for collinearity. That is the projective duality between points and lines, not decoration.

The shoelace formula: for a simple polygon with vertices $(x_1, y_1), \ldots, (x_m, y_m)$ in order,

$$\text{Area} = \frac{1}{2}\Big|\sum_{i=1}^{m}(x_iy_{i+1} - x_{i+1}y_i)\Big|, \qquad \text{indices modulo } m.$$

It is the fan decomposition from the origin: each term is twice the signed area of triangle OP_iP_{i+1}, and the signed areas add to the polygon's area with the outside cancelling. Putnam 2007 A2 uses the

fan form directly: place the origin inside the hull, order the points by angle, sum $\frac{1}{2}|x_i y_{i+1} - x_{i+1} y_i|$. To place the origin inside, use the angular criterion: if every consecutive angular gap around O is less than π, no line through O has all the points strictly on one side, so O is inside the hull. Do not instead assume that four points taken one from each of four angular sectors form a convex quadrilateral in that order; they may not, and the true hull may be a triangle. That failure is why the fan argument is used.

Theorem 9.7.8 (Pick). *A simple polygon whose vertices are lattice points has area $I + \frac{B}{2} - 1$, where I is the number of interior lattice points and B the number on the boundary.*

Two moves. First, $I + \frac{B}{2} - 1$ is additive when two lattice polygons are glued along a common lattice edge, because shared boundary points are counted once instead of twice and the two subtracted 1s become one. Second, every lattice polygon triangulates into *primitive* lattice triangles (no lattice points beyond the three vertices), and a primitive triangle has area exactly $\frac{1}{2}$, for which $I = 0$, $B = 3$ and the formula returns $\frac{1}{2}$. The primitive area statement is the lattice index fact of Section 13. Reconstruct the additivity by hand; cite the triangulation step by name.

The corollary you use more often than Pick: by the shoelace determinant a lattice triangle has area a half integer. Putnam 2010 B2 concludes from this that any lattice triangle with integer sides is Heronian, and Putnam 2012 A4 uses the index q refinement: three noncollinear points of a sublattice of index q span a triangle of area at least $q/2$.

- **WHEN** three cevians are claimed concurrent, or three points on the sides are claimed collinear **DO** write the signed ratio product and check whether the target is $+1$ or -1 **BECAUSE** concurrence and collinearity are the same identity with opposite sign, so one memorised product covers both, and the areas proof reconstructs it in two lines if you forget
- **WHEN** a polygon has explicit coordinates and you need its area **DO** shoelace, and if the coordinates are integers immediately conclude the area is a half integer **BECAUSE** the determinant form is polynomial in the coordinates, so arithmetic hypotheses survive it; the half integer conclusion is often the whole content of the problem
- **WHEN** a lattice polygon's area and its lattice point count both appear **DO** use Pick to convert one into the other **BECAUSE** Pick is the unique linear relation between area and lattice counts, so it is the only bridge between a continuous and a discrete hypothesis in the same figure

9.8 Conics

One definition, three curves

Definition 9.8.1 (Focus and directrix). Given a point F, a line ℓ not through F and a constant $e > 0$, the conic with focus F, directrix ℓ and eccentricity e is $\{P : |PF| = e\, d(P, \ell)\}$. It is an ellipse if $e < 1$, a parabola if $e = 1$ and a hyperbola if $e > 1$.

Taking F at the origin and ℓ the line $x = -d$, the definition reads $r = e(d + r\cos\theta)$ in polar coordinates, which rearranges to the form you should carry:

$$r = \frac{\ell_0}{1 + e\cos\theta}, \qquad \ell_0 = ed.$$

That single equation contains all three curves and is the right starting point for any problem about focal chords or orbits, because the focus is at the origin where the hypotheses live. The two focus characterisations ($|PF_1| + |PF_2| = 2a$ for the ellipse, $\big||PF_1| - |PF_2|\big| = 2a$ for the hyperbola) are equivalent to the focus and directrix definition and are usually the cheaper form when the hypothesis is about a sum or difference of distances; Putnam 2019 A2 uses the ellipse version to see

that $|CA| + |CB| = 2|AB|$ places C on an ellipse with foci A and B, so that fixing the angle at B pins C to a ray meeting the ellipse once and the triangle is unique up to similarity.

Parametrisations, in the form you should write without thinking: ellipse $(a\cos t, b\sin t)$; parabola $(at^2, 2at)$; hyperbola $(a\cosh t, b\sinh t)$ on one branch, or $(a\sec t, b\tan t)$; and the rectangular hyperbola $xy = k$ with $k > 0$ parametrised bijectively on the branch $x > 0$ by $t \mapsto (t, k/t)$. That last one is the one Putnam 2015 A1 needs: three points on a branch become three positive reals a, t, b instead of six coordinates tied by a constraint. That reduction, six unknowns and three constraints down to three unknowns, is the entire reason to parametrise.

Intersections and tangency

Two conics in the plane, neither a multiple of the other, meet in at most four points counted with multiplicity: eliminating one variable produces a quartic. This is the plane case of Bézout's theorem, and you should quote it by name.

Tangency has an algebraic test which is the workhorse of Putnam 2007 A1. If C_1 is the graph $y = f(x)$ and C_2 is another curve, substitute to get $H(x) = 0$; the common points are the roots of H, and where C_2 has no vertical tangent, the curves are tangent at x_0 exactly when x_0 is a repeated root, that is $H(x_0) = H'(x_0) = 0$. The two dimensional geometric condition becomes a one variable repeated root condition, and repeated roots are detected by discriminants and resultants, which are algebra you can actually finish.

- **WHEN** a problem says two curves are tangent **DO** substitute one into the other and impose a repeated root **BECAUSE** tangency is first order contact, which is exactly a double root of the elimination polynomial; every other formulation drags in derivatives of implicit functions and costs more ink
- **WHEN** a conic appears with its focus in the hypotheses **DO** use the polar form $r = \ell_0/(1 + e\cos\theta)$ with the focus at the origin **BECAUSE** focal distances are then a single algebraic expression in θ, and the sum or difference conditions collapse

9.9 Convexity

Hulls, half spaces and supporting hyperplanes

A set is convex if it contains the segment between any two of its points. The convex hull $\operatorname{conv}(S)$ is the intersection of all convex sets containing S, equivalently the set of convex combinations of points of S. The second description is what you compute with, the first what you prove with, and it gives the most used fact in the archive: *a convex set containing finitely many points contains their entire hull*. Putnam 2007 A2 converts "a convex set meets four disjoint branches" into "a convex set contains this four point configuration", and every bound for the hull transfers to the set.

Dually, a closed convex set is the intersection of the closed half spaces containing it. For a polygon or polytope those half spaces are the ones bounded by the facet hyperplanes, which is the fact used in Putnam 2016 B3 to collapse three half plane memberships into membership in a triangle.

Theorem 9.9.1 (Supporting hyperplane)**.** *At every boundary point p of a closed convex set $C \subseteq \mathbb{R}^n$ with nonempty interior there is a nonzero $\mathbf{n}$ with $\mathbf{x}\cdot\mathbf{n} \le p\cdot\mathbf{n}$ for all $\mathbf{x} \in C$. Two disjoint convex sets, one compact and one closed, can be strictly separated by a hyperplane.*

Its one dimensional shadow is the supporting line theorem, used in Putnam 2012 B5: a finite convex $F : \mathbb{R} \to \mathbb{R}$ has at every t a slope m with $F(x) \ge F(t) + m(x - t)$, so F is the pointwise supremum of its affine minorants and the supremum is attained. Convexity survives pointwise suprema, non negative combinations, affine substitution inside and composition with a nondecreasing convex function outside, but not infima.

Carathéodory, Radon, Helly

Theorem 9.9.2 (Carathéodory)**.** *Every point of* $\operatorname{conv}(S)$ *for* $S \subseteq \mathbb{R}^d$ *is a convex combination of at most* $d+1$ *points of* S*.*

Proof. Take a representation $x = \sum_{i=1}^{m} \lambda_i x_i$ with m minimal and all $\lambda_i > 0$. If $m > d+1$ the points $x_2 - x_1, \dots, x_m - x_1$ are $m-1 > d$ vectors in $\mathbb{R}^d$, hence dependent: there are μ_i, not all zero, with $\sum \mu_i x_i = 0$ and $\sum \mu_i = 0$. Then $x = \sum(\lambda_i - t\mu_i)x_i$ for every t; increase t until some coefficient first hits zero, which is possible since some $\mu_i > 0$. That contradicts minimality. □

Theorem 9.9.3 (Radon)**.** *Any* $d+2$ *points in* $\mathbb{R}^d$ *can be split into two sets whose convex hulls intersect.*

Proof. As above there are μ_i, not all zero, with $\sum \mu_i x_i = 0$ and $\sum \mu_i = 0$. Let $I = \{i : \mu_i > 0\}$ and J its complement, and put $M = \sum_{i \in I} \mu_i > 0$. Then $\sum_{i \in I} \frac{\mu_i}{M} x_i = \sum_{j \in J} \frac{-\mu_j}{M} x_j$ and both sides are convex combinations. □

In the plane this is exactly the classification that Putnam 2006 A6 needs: four points with no three collinear either form a convex quadrilateral or have exactly one point inside the triangle of the other three, never both. That dichotomy turns the convex position probability into four copies of one "point inside a triangle" probability.

Theorem 9.9.4 (Helly)**.** *If* $C_1, \dots, C_m$ *are convex sets in* $\mathbb{R}^d$ *with* $m \geq d+1$ *and every* $d+1$ *of them have a common point, then all* m *have a common point.*

Proof. Induct on m; the base $m = d+1$ is the hypothesis. For $m > d+1$, by induction pick for each i a point x_i lying in every C_j with $j \neq i$. Apply Radon to $x_1, \dots, x_{d+2}$ (after discarding surplus points): there is a partition $I \sqcup J$ and a point z in $\operatorname{conv}\{x_i : i \in I\} \cap \operatorname{conv}\{x_j : j \in J\}$. For any index k, either $k \notin I$ or $k \notin J$; in the first case every x_i with $i \in I$ lies in C_k, so $z \in C_k$ by convexity, and similarly in the second. Hence z lies in every C_k. □

Extreme points and Minkowski sums

A point of a convex set is *extreme* if it is not the midpoint of a segment inside the set with distinct endpoints. The finite dimensional Krein and Milman statement is that a compact convex set is the hull of its extreme points, and for a polytope those are the vertices. Two consequences. A *linear* functional on a compact convex set attains its maximum at an extreme point, and an extreme point of a set cut out by inequalities has at least d active constraints; Putnam 2016 A6 uses both halves, existence and a prediction of how many contact conditions the optimum satisfies. A *convex* functional also attains its maximum at an extreme point, by Jensen; Putnam 2023 B4 uses this on the polytope $\{s : \sum s_k = T,\ s_k \geq \ell_k\}$, whose vertices are the feasible points with all coordinates but one at their floors.

The Minkowski sum is $A + B = \{a + b : a \in A, b \in B\}$. It is convex if A and B are, and for convex A one has $A + A = 2A$, which fails badly for non convex sets and is therefore a convexity test in disguise. For intervals it is simply $[p, q] + [p', q'] = [p + p', q + q']$, which is the form Putnam 2016 B5 uses to propagate reachable displacements.

- **WHEN** a convex set is known to contain some points **DO** immediately replace it by the hull of those points and prove everything for the hull **BECAUSE** convexity is exactly closure under convex combinations, so the hull is free information and every bound for it transfers upward
- **WHEN** you must maximise or minimise a linear or convex quantity over a compact convex set or a polytope **DO** go straight to the extreme points and count active constraints **BECAUSE** the maximum is attained there, so a continuous optimisation collapses into a finite check, and the constraint count tells you in advance how many equations the optimum satisfies

- **WHEN** you need one point common to many convex sets, and you only know about small subfamilies **DO** invoke Helly with $d+1$ **BECAUSE** Helly is the only theorem that upgrades local intersection to global intersection, and its hypothesis is exactly what such problems supply
- **WHEN** you must show a point is inside a hull **DO** either exhibit it as a convex combination, or show every line through it has points on both sides (the angular gap criterion in the plane) **BECAUSE** these are the two dual certificates, and one of them is always cheap: the first is algebra, the second is a sorting argument

9.10 Solid geometry, polyhedra and the regular solids

For a convex polyhedron, $V - E + F = 2$. If every face is a triangle then $3F = 2E$, and since the degrees of the vertices sum to $2E$, a solid all of whose vertices look alike has every vertex of degree $2E/V$. Putnam 2013 A1 needs exactly this chain: the icosahedron has $V = 12$, $E = 30$, $F = 20$, hence every vertex meets $2E/V = 5$ faces, which sets the local bound the problem turns on.

The classification of the regular solids is the inequality $\frac{1}{p} + \frac{1}{q} > \frac{1}{2}$ for p sided faces with q at each vertex, whose five solutions are $(3,3)$, $(3,4)$, $(4,3)$, $(3,5)$, $(5,3)$. Carry the coordinates, because a problem that needs them and does not supply them is a problem you lose:

the cube is $\{\pm 1\}^3$; the octahedron is the six points $(\pm 1, 0, 0)$ and cyclic permutations; the tetrahedron is the four alternate vertices of the cube, for instance $(1,1,1)$, $(1,-1,-1)$, $(-1,1,-1)$, $(-1,-1,1)$; the icosahedron is the twelve cyclic permutations of $(\pm 1, \pm\varphi, 0)$ with φ the golden ratio, that is the corners of three mutually perpendicular $2 \times 2\varphi$ rectangles; the dodecahedron is $\{\pm 1\}^3$ together with the cyclic permutations of $(0, \pm\varphi^{-1}, \pm\varphi)$.

Check the icosahedron once by hand, because Putnam 2023 A4 depends on it. All twelve points have norm $\sqrt{1+\varphi^2}$; $(1, \varphi, 0)$ is at distance 2 from $(-1, \varphi, 0)$, and from $(0, 1, \varphi)$ at squared distance $1 + (\varphi - 1)^2 + \varphi^2 = 4$ using $\varphi^2 = \varphi + 1$. So the edge is 2 and each vertex has exactly five others at distance 2. Adding and subtracting vertices produces the axis vectors 2 and 2φ, which is where that problem's arithmetic comes from.

For the regular tetrahedron of edge a: circumradius $a\sqrt{6}/4$, volume $a^3/(6\sqrt{2})$, circumcentre equal to centroid, and from the centre the vertex vectors satisfy $P_i \cdot P_j = -R^2/3$. That last identity is one line: $\sum P_i = \mathbf{0}$, so $0 = |\sum P_i|^2 = 4R^2 + 12\, P_i \cdot P_j$. Putnam 2021 A3 recovers exactly this fact algebraically from the vanishing of a 4×4 Gram determinant, and its construction for $N = 3m^2$ is the alternate vertices of the cube $[-m, m]^3$.

- **WHEN** a problem names a Platonic solid **DO** write its coordinates immediately and convert every geometric hypothesis into dot products of those vectors **BECAUSE** the solids are defined by their symmetry, and coordinates are the only form in which a symmetry hypothesis becomes arithmetic you can finish in the arena
- **WHEN** a problem counts faces, edges or vertices, or bounds a local configuration on a polyhedron **DO** use $V - E + F = 2$ together with the degree sum $\sum \deg = 2E$ and the face size relation **BECAUSE** these are three linear equations in the counts, which usually pin the local degree exactly, and the local degree is what the bound is really about

9.11 Geometric probability and integral geometry

The base identity, and Crofton

For U uniform on a region K and measurable $A \subseteq K$, $\mathrm{P}(U \in A) = |A|/|K|$; and if A is itself random and independent of U then $\mathrm{P}(U \in A) = \mathrm{E}\,|A|/|K|$ by conditioning and Fubini. That single line is

what removes probability from a geometric probability problem and leaves a mean value. It is the first step in Putnam 2006 A6, where $\mathrm{P}(S \in \triangle PQR)$ becomes $\mathrm{E}[T]/\pi$.

Parametrise the lines of the plane by $x\cos\theta + y\sin\theta = p$ with $\theta \in [0,\pi)$ and $p \in \mathbb{R}$, and give the set of lines the measure $dp\,d\theta$. This measure is invariant under all rigid motions, which is why it is the right one.

Theorem 9.11.1 (Crofton). *For a rectifiable plane curve γ of length L, $\int n(\ell)\,dp\,d\theta = 2L$, where $n(\ell)$ is the number of intersections of the line ℓ with γ.*

Proof for a segment, which is all you need. For a segment of length L making angle α with the x axis, a line of direction parameter θ meets it exactly for p in an interval of length $L|\sin(\theta-\alpha)|$, the length of the projection of the segment onto the normal direction. Hence the integral is $\int_0^\pi L|\sin(\theta-\alpha)|\,d\theta = L\int_0^\pi |\sin u|\,du = 2L$. Both sides are additive over concatenation and continuous under refinement, so the general rectifiable case follows by approximating γ with polygons. □

Two payoffs. First, Buffon's needle: drop a needle of length $\lambda \le d$ on a plane ruled with parallel lines a distance d apart. With θ uniform on $[0,\pi)$ and the offset u of the centre uniform on $[0,d)$, the needle crosses exactly when $u < \frac{\lambda}{2}\sin\theta$ or $u > d - \frac{\lambda}{2}\sin\theta$, an event of conditional probability $\lambda\sin\theta/d$, so

$$\mathrm{P}(\text{cross}) = \frac{1}{\pi}\int_0^\pi \frac{\lambda\sin\theta}{d}\,d\theta = \frac{2\lambda}{\pi d}.$$

Second, the same computation with the crossing count in place of the indicator gives the *expected* number of crossings for any curve of length λ, needle or not, since expectation is linear and a curve is a union of small segments; the needle is the special case where the count is 0 or 1.

That linearity argument is the lattice argument of Putnam 2021 B1: if a measurable set K is translated by a vector uniform modulo the integer lattice, the expected number of lattice points in the translate equals the area of K, because the pieces $(C-P)\cap K$ over lattice points P reassemble K when C is a unit cell. There the shape is a unit square, the count lies in $\{0,1,2\}$ (two lattice points inside a rotated unit square must differ by a unit coordinate vector), and $\mathrm{E}[N] = 1$ forces $\mathrm{P}(N=0) = \mathrm{P}(N=2)$ with no integration at all.

Averaging over directions

The three dimensional analogue of Crofton is the Cauchy projection formula, and Putnam 2013 A5 is built on it.

Theorem 9.11.2 (Cauchy projection formula, plate case). *For $\mathbf{V} \in \mathbb{R}^3$, $\int_{S^2} |\mathbf{V}\cdot\mathbf{u}|\,d\sigma(\mathbf{u}) = 2\pi|\mathbf{V}|$. Equivalently the mean of $|\mathbf{n}\cdot\mathbf{u}|$ over uniform directions is $\frac{1}{2}$, so the average shadow of a flat plate is half its area; for a convex body the average shadow area is a quarter of the surface area.*

Proof. By rotation invariance take $\mathbf{V} = |\mathbf{V}|\mathbf{e}_3$. In spherical coordinates the integral is $\int_0^{2\pi}\int_0^\pi |\mathbf{V}||\cos\phi|\sin\phi\,d\phi\,d\theta =$ $2\pi|\mathbf{V}|\int_0^\pi |\cos\phi|\sin\phi\,d\phi = 2\pi|\mathbf{V}|$. □

Structurally the identity writes the norm, a nonlinear function, as an average of absolute values of *linear* functionals; norms with such a representation are called zonoid norms, and this linearisation is the exact obstruction that kills the algebraic approach. The constant is forced before you compute it: $I(\mathbf{V}) = \int_{S^2}|\mathbf{V}\cdot\mathbf{u}|d\sigma$ is rotation invariant and homogeneous of degree one, and $SO(3)$ acts transitively on spheres with a unique invariant measure up to scale, so $I(\mathbf{V}) = c|\mathbf{V}|$. Existence of c is the proof; the value 2π is a convenience.

Random points, random chords

Sylvester's four point problem: for bounded convex K, $\mathrm{P}(\text{four uniform points in convex position}) = 1 - 4\,\mathrm{E}[T]/|K|$ with T the area of the triangle on three uniform points. Values: $2/3$ for a triangle, $25/36$ for a square, $1 - \frac{35}{12\pi^2}$ for a disc; by Blaschke's theorem the minimum over convex K is $2/3$ at triangles and the maximum is the disc value at ellipses. The generalisation is Efron's identity: for n uniform points, with V_n the number of hull vertices and A_{n-1} the hull area of $n-1$ of them,

$$\mathrm{E}[V_n] = n\Big(1 - \frac{\mathrm{E}[A_{n-1}]}{|K|}\Big),$$

proved by writing V_n as a sum of indicators of the events "point i lies outside the hull of the others". Putnam 2006 A6 is the case $n = 4$ and the identity is the independent cross check on the answer.

Random chords carry a warning. Bertrand's paradox is that "a random chord" is undefined until you name the model: uniform endpoints, uniform midpoint and uniform offset along a uniform direction give three different distributions. Putnam 2024 A5 uses uniform endpoints, where the chord between $Re^{i\alpha}$ and $Re^{i\beta}$ has the law of a uniformly rotated vertical chord $x = R\cos\psi$ with ψ uniform and independent of the rotation, so the signed offset has the arcsine density $\frac{1}{\pi\sqrt{R^2-t^2}}$. Name the mechanism behind that decoupling: sum and difference of independent uniform angles are independent uniform angles, because an integer matrix of nonzero determinant induces a measure preserving map of the torus.

- **WHEN** a probability is asked about a uniformly random point in a region **DO** convert it to a ratio of areas, and if the target region is itself random, to a ratio of an expectation to an area **BECAUSE** this is the definition of the uniform distribution; the conversion removes probability from the problem and leaves a mean value you can integrate
- **WHEN** a quantity is an average over all directions or over all lines **DO** use invariance to force the answer to be a constant times the natural geometric quantity, then evaluate the constant on one convenient example **BECAUSE** the invariant measure is unique up to scale, so the shape of the answer is determined before any integration; only one number remains
- **WHEN** an expected count of hits, crossings or lattice points appears **DO** write the count as a sum of indicators and use linearity, never the distribution of the count **BECAUSE** linearity does not care about dependence, which is why Buffon, Barbier, Efron and the lattice identity are all one line
- **WHEN** the phrase "a random chord" or "a random line" appears **DO** state the model explicitly before computing **BECAUSE** Bertrand's paradox is that the answer changes with the model, so an unstated model is an unfinished problem

9.12 Calculus applied to geometry

Arc length of $\mathbf{r} : [a,b] \to \mathbb{R}^n$ is $\int_a^b |\mathbf{r}'(t)|\,dt$, and for $y = f(x)$ it is $\int \sqrt{1+f'^2}\,dx$. Curvature is

$$\kappa = \frac{|\mathbf{r}' \times \mathbf{r}''|}{|\mathbf{r}'|^3} \quad \text{in } \mathbb{R}^3, \qquad \kappa = \frac{|y''|}{(1+y'^2)^{3/2}} \quad \text{for a plane graph,}$$

and $1/\kappa$ is the radius of the osculating circle, the circle agreeing with the curve to second order. The surface of revolution of $y = f(x) \geq 0$ about the x axis has area $2\pi \int f\sqrt{1+f'^2}\,dx = 2\pi \int f\,ds$, and volumes of revolution come from discs, washers or shells according to which one avoids solving for the inverse function.

The two Pappus centroid theorems package all of that.

Theorem 9.12.1 (Pappus). *If a plane region R of area A lies on one side of a coplanar line L whose distance from the centroid of R is d, the solid of revolution has volume $V = 2\pi dA$. If a plane curve of length s lies on one side of L at centroid distance d, the surface of revolution has area $S = 2\pi ds$.*

Putnam 2006 A1 is one line once you see it. The region is defined by a relation depending on x, y only through $r = \sqrt{x^2 + y^2}$, so it is invariant under rotation about the z axis and the whole problem collapses to the meridian half plane; completing the square shows the meridian region is a disc $(u-a)^2 + (v-b)^2 \le \rho^2$ whose centre is further than ρ from the axis, which is exactly the hypothesis Pappus needs; the centroid of a disc is its centre; done.

Archimedes' hat box theorem is the surface analogue you should also carry: the zone of a sphere of radius R between two parallel planes a distance h apart has area $2\pi Rh$ regardless of position. Equivalently the surface measure pushes forward under the height map to $2\pi R$ times Lebesgue measure, which is why a function of one coordinate turns a spherical integral into an undistorted one dimensional integral; Putnam 2019 A4 exploits exactly that to search a cheap family for a counterexample.

Finally the isoperimetric inequality: a simple closed plane curve of length L enclosing area A satisfies $L^2 \ge 4\pi A$, with equality only for circles; in $\mathbb{R}^3$, $S^3 \ge 36\pi V^2$. The standard proof (Hurwitz) parametrises by arc length, normalises $L = 2\pi$, expands the coordinates in Fourier series, writes A and L in terms of the coefficients, and applies Wirtinger's inequality; equality forces all modes above the first to vanish, which is a circle. Know that route by name and be able to state the inequality without hesitation.

- **WHEN** a solid or surface of revolution appears and the meridian region is a recognisable shape whose centroid you know **DO** use Pappus rather than integrating **BECAUSE** the integral $\int 2\pi y \, dA$ is by definition 2π times the first moment, which is $2\pi \bar{y} A$; integrating recomputes a centroid you already know and costs more ink
- **WHEN** a region in $\mathbb{R}^3$ is defined by a relation in x, y only through $\sqrt{x^2 + y^2}$ **DO** substitute r and work in the meridian half plane **BECAUSE** the region is then rotationally invariant by construction, so a three variable problem is a two variable problem plus a known revolution
- **WHEN** an inequality relates a perimeter to an area, or a surface area to a volume **DO** test it against the isoperimetric inequality and check the circle or sphere case first **BECAUSE** the isoperimetric inequality is the sharp relation between those two quantities, so it is either the tool or the sanity check, and the equality case tells you the extremal shape before you start

9.13 Lattices and the geometry of numbers

Definition 9.13.1 (Lattice). A lattice $L \subset \mathbb{R}^n$ of full rank is $L = B\mathbb{Z}^n = \{B\mathbf{k} : \mathbf{k} \in \mathbb{Z}^n\}$ for an invertible real matrix B. Its determinant is $\det L = |\det B|$, the volume of a fundamental cell, and it does not depend on the basis because two bases differ by a matrix in $GL_n(\mathbb{Z})$, which has determinant ± 1.

Structurally, every discrete subgroup of $\mathbb{R}^n$ is a lattice of some rank, and every *closed* subgroup is $V + L$ with V a subspace and L a lattice in a complement; so a subgroup is discrete, dense in its span, or a mixture along complementary directions. That trichotomy is the tool behind Putnam 2023 A4. The rank one case, "every discrete subgroup of $\mathbb{R}$ is cyclic", is the one you use most often, and it is the reason the lattice points on a line through a lattice point form a coset $P_0 + \mathbb{Z}\mathbf{w}_0$; intersecting that with a convex set gives a set of *consecutive* integers k, whose coordinate projections are therefore arithmetic progressions. Putnam 2012 A4 converts a collinearity conclusion into the arithmetic progression the problem demands by exactly this route.

Index is the other essential: if $M \subseteq L$ are full rank lattices, the index $[L : M]$ equals $\det M / \det L$, and indices multiply in towers. Hence for a sublattice of $\mathbb{Z}^2$ of index q, any two independent vectors

span a parallelogram whose area is a positive multiple of q, so three noncollinear points of that sublattice span a triangle of area at least $q/2$. That quantisation is the whole obstruction in Putnam 2012 A4, and the case $q = 1$ is the primitive triangle fact used to prove Pick.

Theorem 9.13.2 (Minkowski's convex body theorem)**.** *Let $L \subset \mathbb{R}^n$ be a lattice and K a convex set, symmetric about the origin, with $\mathrm{vol}(K) > 2^n \det L$. Then K contains a nonzero point of L.*

Proof. First Blichfeldt: if S has $\mathrm{vol}(S) > \det L$ then S contains two distinct points differing by a lattice vector. Indeed, let F be a fundamental cell; the sets $(S \cap (P + F)) - P$ for $P \in L$ all lie in F and have total volume $\mathrm{vol}(S) > \mathrm{vol}(F)$, so two of them overlap, giving $s_1 - P_1 = s_2 - P_2$ with $P_1 \neq P_2$. Now apply this to $S = \frac{1}{2}K$, which has volume $2^{-n}\,\mathrm{vol}(K) > \det L$: there are distinct $\frac{1}{2}k_1, \frac{1}{2}k_2 \in S$ with $\frac{1}{2}(k_1 - k_2) \in L$ nonzero. By symmetry $-k_2 \in K$, and by convexity $\frac{1}{2}(k_1 - k_2) \in K$. □

Lattice symmetries license the "without loss of generality" steps that otherwise look unexamined: integer translations and quarter turns about a lattice point map $\mathbb{Z}^2$ to itself and preserve all distances, which is exactly what justifies placing one point at the origin and another on an axis in Putnam 2010 B2.

Which rotations can preserve a lattice at all is the crystallographic restriction, and it is the structural heart of Putnam 2023 A4.

Proposition 9.13.3 (Crystallographic restriction)**.** *If a rotation of order m preserves a lattice of rank n, its matrix in a lattice basis has integer entries, hence integer trace. In $\mathbb{R}^3$ the trace of a rotation by $2\pi/m$ is $1 + 2\cos(2\pi/m)$, so $2\cos(2\pi/m) \in \mathbb{Z}$ and $m \in \{1, 2, 3, 4, 6\}$. In particular no lattice is invariant under a rotation of order five, so no lattice has icosahedral symmetry, while the cube's rotations all have order $1, 2, 3$ or 4 and cause no obstruction.*

- **WHEN** a problem mixes a convex symmetric region with a lattice and asks for a nonzero lattice point **DO** compute the volume and compare it with $2^n \det L$ **BECAUSE** Minkowski's theorem is the only general existence statement of this kind, and its hypotheses (convex, symmetric, volume) are exactly the three things such a problem hands you
- **WHEN** lattice points and areas appear together **DO** use the index quantisation (det, hence area, is a multiple of the index) and Pick **BECAUSE** lattice areas are quantised in half integers, so a continuous bound plus quantisation forces a discrete conclusion, usually collinearity or an exact count
- **WHEN** a rotational symmetry of order five or greater than six is being imposed on a lattice or on an integer configuration **DO** apply the crystallographic restriction via the integer trace **BECAUSE** a lattice preserving map has an integer matrix in a lattice basis, so its trace is an algebraic integer constraint that only $m \in \{1, 2, 3, 4, 6\}$ can satisfy

9.14 Recursion, induction and self similarity in geometry

Geometry looks static, so solvers forget to look for recursion in it. A large share of the hard geometry in the archive is a recursion wearing a picture.

The general form. A geometric recursion is a map Φ on configurations together with a quantity q. Solve it in three moves. Identify the state: what finite data determines the configuration up to the symmetries you may use. Write $q(\Phi(X))$ in terms of $q(X)$ and check whether the relation is linear; if it is, the characteristic equation finishes it. Then find the invariant, the quantity unchanged or scaled by a fixed factor: in geometry it is almost always an area ratio, a dot product, a norm, an angle or a lattice index.

Linear recursions generated by a fixed transformation

Putnam 2010 A5 is the model case. With $\mathbf{x}$ fixed and $\mathbf{w}_0 \perp \mathbf{x}$, iterate $\mathbf{w}_{k+1} = \mathbf{x} \times \mathbf{w}_k$. The triple product gives $\mathbf{w}_{k+2} = -|\mathbf{x}|^2 \mathbf{w}_k$, characteristic roots $\pm i|\mathbf{x}|$, so the orbit is periodic of period four up to the scaling $|\mathbf{x}|^k$. Generalise the trigger: *any linear map A with $A^2 = -\lambda I$ on a subspace generates $\mathbf{w}_{k+2} = -\lambda \mathbf{w}_k$ there, hence a scaled quarter turn of period four*. Rotations by $2\pi/m$ give period m; Householder reflections give period two.

Putnam 2022 B2 is the same circuit with the scaling doing the work: once the vectors are orthogonal $|\mathbf{u} \times \mathbf{v}| = |\mathbf{u}||\mathbf{v}|$, so iterating produces norms in geometric progression, which a finite set cannot hold. The rule: *if $N(a * b) = N(a)N(b)$ and the ambient set is finite, then $N \equiv 1$ on the nonzero part*, since otherwise $N(a), N(a)^2, \ldots$ realise infinitely many values.

Self similar subdivision

Putnam 2009 A1 uses the midpoint square: joining the midpoints of consecutive sides of a square gives a smaller square with the same centre, and in general the midpoint polygon of a regular n gon is a regular n gon with the same centre. Iterating gives an infinite nested family sharing that centre, and each level supplies a fresh instance of the hypothesis about it. The method: *when a hypothesis is stated for every figure of a given type, generate a nested self similar family and gather as many independent linear equations as you need, then eliminate*. There the four small squares formed by a vertex, the two adjacent edge midpoints and the centre are what contain $f(P)$, and the midpoint square supplies the cancelling equation.

The same shape appears in self similar sets: if $X = F \cup \alpha X$ for a similarity α with $|\alpha| > 1$ and finite base F, any finite configuration in X may be divided by α until it meets F, so all configurations are α^m images of base ones. Termination is descent on a valuation, not luck.

Induction by peeling a boundary

Putnam 2007 A6 is the archetype. A diagonal of a simple polygon lying in its interior splits it into two simple polygons with $n_1 + n_2 = n + 2$, both at least 3 and hence both at most $n - 1$: that is the splitting case. When no such chord exists the triangulation near the boundary is forced and you peel a run of low degree boundary vertices. The bookkeeping tool is the angle sum: counting all angles of all T triangles two ways gives $T\pi = 2\pi I + (n-2)\pi$, hence $T = n + 2I - 2$, faster than Euler's formula and with no graph theory. The local statement is that in an edge to edge triangulation of a disc the link of an interior vertex is a cycle and of a boundary vertex a path, so degree a means a triangles inside and $a - 1$ on the boundary.

The rule: *for a triangulated region, induct by removing a boundary feature and use a two way count of angles, edges or area to relate before and after*. The sharp constant may not be what the induction gives: the extremal object there is a hexagonal patch of the triangular lattice with about $n^2/6$ triangles, so $n^2 - 2n - 2$ has the right order and the wrong constant.

Recursion in the dimension

Two items in the archive recurse on the ambient dimension rather than on a count.

Putnam 2013 A5: the statement for $\mathbb{R}^2$ implies it for every $\mathbb{R}^n$. Project along a uniformly random unit vector onto a hyperplane; a triangle spanning a plane W has its area multiplied by $\sqrt{1 - |P_W \mathbf{u}|^2}$, whose average over S^{n-1} is independent of W because $O(n)$ acts transitively on two dimensional subspaces and the sphere measure is invariant. One averaging step performs the induction $A_{n-1} \Rightarrow A_n$. Extract the method: *to remove a dimension, identify a group acting transitively on the data you want to forget, check that an invariant measure exists, and average the distortion factor into a constant*.

Putnam 2018 A6: the Cayley and Menger determinant is itself a dimensional recursion, since the $(k+2) \times (k+2)$ determinant computes the k volume and by vanishing detects that the points lie in a $(k-1)$ flat, where the determinant one size down takes over.

Putnam 2016 B3 recurses on dimension in the construction: the extremal simplex doubling that gives 4 in the plane gives k^k in $\mathbb{R}^k$, so 27 in three dimensions and not 8. Check it on the standard simplex: the hyperplanes through each vertex parallel to the opposite facet cut out a simplex with edge vectors k times as long. What generalises is the extremal principle, the parallel hyperplane doubling and the half space intersection; the clean medial triangle picture does not.

- **WHEN** a geometric construction can be applied to its own output **DO** iterate it, write the recursion for the one quantity you care about, and solve it as a linear recursion **BECAUSE** a construction that consumes a figure and returns a figure of the same type is a map on a state space, and the state space is usually two or three numbers, so the characteristic equation finishes it
- **WHEN** a hypothesis holds for every figure of a given type **DO** build a nested self similar family and collect one equation per level **BECAUSE** one instance rarely determines the unknown; a self similar family gives you unlimited independent instances about the same centre at zero construction cost
- **WHEN** a problem is about a triangulated region or a polygon with a subdivision **DO** induct by cutting along a chord or peeling a low degree boundary vertex, and control the counts with a two way angle or edge count **BECAUSE** the chord splitting identity $n_1 + n_2 = n + 2$ guarantees both pieces are strictly smaller, which is exactly what an induction needs
- **WHEN** a statement is posed in $\mathbb{R}^3$ but nothing in it is special to three dimensions **DO** prove the two dimensional case and transfer by averaging over random projections **BECAUSE** the invariant measure forces the distortion constant to be independent of the subspace, so one averaging step is the entire induction

9.15 Installation order

Install in this order, because this is the order in which the archive uses these objects.

First, without hesitation: polarisation, projection, the determinant as area and volume, the cross product from its determinant definition with the norm law and the triple product, the perpendicular bisector as the locus of a circle's centre, the incircle block, the barycentric weights of the centroid and incentre with the parallelism criterion, the shoelace formula, the triangle as an intersection of half planes, convex hull containment, and the reflection and homothety circuits. Each carries at least one priority five item: Putnam 2006 A1, 2007 A1, 2007 A2, 2008 B1, 2008 B3, 2010 B2, 2013 A5, 2015 A1, 2016 B3, 2017 B1, 2017 B5, 2019 A2, 2021 A3 and 2024 B2.

Second: Gram determinants, the isometry classification and orthogonal matrices, Householder reflections, affine invariance and the "normalise the triangle" reroute, power of a point and cyclic quadrilaterals, Ceva and Menelaus, Euler's polyhedron formula with the coordinates of the regular solids, Pappus, and the lattice index quantisation. These carry Putnam 2011 A4, 2011 B4, 2012 A4, 2013 A1, 2016 A3, 2018 A6, 2019 B3, 2022 B2 and 2023 A4.

Third, the ones that appear once or twice and when they appear *are* the problem: the Cauchy projection formula and zonoid linearisation (2013 A5), Crofton and Buffon, Efron and Sylvester (2006 A6), Cayley and Menger (2018 A6), the crystallographic restriction (2023 A4), Ravi substitution (2015 B4), Bretschneider (2024 B2), the arc and chord strip lemma (2024 A5), the hat box theorem (2019 A4) and the anharmonic group (2016 A3).

Last, and honestly: Helly, Carathéodory, Krein and Milman, Minkowski's convex body theorem, Ptolemy, Pick and the isoperimetric inequality appear nowhere by name in the 2006 to 2025 window. They are here because they complete structures the archive uses constantly (hull containment,

extreme points, lattice area quantisation, cyclic quadrilaterals) and a paper that wanted one would give no warning. Learn them last, as statements you can quote and apply, and do not spend arena time reconstructing their proofs.

One closing instruction, which is the whole chapter in a sentence. Geometry problems are lost by people who start drawing and start computing before they have named the object. Name the object first, go back to its definition, and the point of interest falls out: the centre falls out of the definition of a circle, the normal falls out of the definition of a hyperplane, the volume falls out of the definition of the determinant, and the extremal configuration falls out of the equality case of the inequality you were about to use.

Chapter 10

Probability, games, invariants and logic

10.1 What this chapter is for

A-level and Further Maths hand you three probabilistic objects and stop: the binomial distribution, the normal distribution, and conditional probability as the formula $\mathrm{P}(A \mid B) = \mathrm{P}(A \cap B)/\mathrm{P}(B)$. That is the whole toolkit. It contains no linearity of expectation used as a weapon, no indicator method, no tail sum formula, no conditioning used as a recursion engine, no order statistics, no probability generating functions, no exponential tilting, no concentration inequality beyond a normal table. It contains nothing at all about invariants, monovariants, colouring arguments, combinatorial games, or the discipline of quantifiers, which are not even topics in the school syllabus: they are a way of thinking, and the Putnam assumes you have it.

This chapter installs all of it. At the end you will write down the state and the sample space of a random process before computing anything; convert any count into a sum of indicators and take its mean term by term; set up a first step recursion and solve the affine ones in closed form; invert a distributional recursion into one on quantile functions when it refuses to iterate; quote the standard laws cold; turn a counting bound into a Chernoff bound by choosing the tilt that makes the target the tilted mean; compute geometric probabilities as volumes and prove existence by averaging; search for an invariant systematically rather than by staring; decide a two player game by finding the $\mathcal{P}$-positions or by pairing, knowing in advance which of the two the game's state structure permits; and negate a nested quantifier correctly, first time, because more Putnam marks are lost there than to any single piece of missing content.

The order below is the order of frequency, and Section 15 gives the full installation order. Exotica are flagged as exotica when they arrive.

10.2 Probability as measure: fix the sample space first

Definition 10.2.1 (Probability space)**.** A probability space is a triple $(\Omega, \mathcal{F}, \mathrm{P})$: the outcomes Ω; a family $\mathcal{F}$ of subsets called events, containing Ω and closed under complement and countable union; and $\mathrm{P} : \mathcal{F} \to [0,1]$ with $\mathrm{P}(\Omega) = 1$ and $\mathrm{P}\left(\bigcup_i A_i\right) = \sum_i \mathrm{P}(A_i)$ for pairwise disjoint A_i.

Go back to the definition and the point of interest falls out. A probability is a normalised measure and nothing else. On a finite Ω with equally likely outcomes, $\mathrm{P}(A) = |A|/|\Omega|$ and probability is counting; on a region of $\mathbb{R}^n$ with the uniform law, $\mathrm{P}(A) = \mathrm{vol}(A)/\mathrm{vol}(\Omega)$ and probability is volume. Every geometric probability problem on the Putnam is a volume computation wearing a costume, and the costume comes off the moment you write the sample space down.

That is the first circuit of the chapter, and it saves whole problems. "Chosen at random" is meaningless until you say with respect to which measure: choosing a chord by two uniform endpoints, by a uniform midpoint, or by a uniform offset along a uniform direction gives three

different laws. That is Bertrand's paradox, and it stops being a paradox once you accept that "a random chord" names no measure. Putnam 2024 A5 and 2021 B1 both require the fixing to be stated.

- **WHEN** a problem says "chosen at random" or "dropped at random" **DO** write Ω and P explicitly, in coordinates, before touching the question **BECAUSE** "at random" is not a distribution, and Bertrand's paradox is the standing proof that different fixings give different correct answers (Putnam 2021 B1, 2024 A5)
- **WHEN** the sample space is a region and the law is uniform **DO** replace every probability by a ratio of volumes and forget probability exists **BECAUSE** $\mathrm{P}(A) = \mathrm{vol}(A)/\mathrm{vol}(\Omega)$ is the definition, not a technique, so the problem was a volume problem all along (Putnam 2006 A6)

Definition 10.2.2 (Conditional probability, independence)**.** For $\mathrm{P}(B) > 0$, $\mathrm{P}(A \mid B) = \mathrm{P}(A \cap B)/\mathrm{P}(B)$. A family $(A_i)_{i\in I}$ is independent when $\mathrm{P}\left(\bigcap_{i\in J} A_i\right) = \prod_{i\in J} \mathrm{P}(A_i)$ for every finite $J \subseteq I$; random variables are independent when the events they generate are.

Pairwise independence is strictly weaker than independence of the family, so keep the quantifier "for every finite subset" in your statement.

Proposition 10.2.3 (Law of total probability and Bayes)**.** *If (B_i) is a countable partition into events of positive probability then $\mathrm{P}(A) = \sum_i \mathrm{P}(A \mid B_i)\,\mathrm{P}(B_i)$, and hence $\mathrm{P}(B_j \mid A) = \mathrm{P}(A \mid B_j)\,\mathrm{P}(B_j)/\sum_i \mathrm{P}(A \mid B_i)\,\mathrm{P}(B_i)$.*

Proof. A is the disjoint union of the $A \cap B_i$, so countable additivity and the definition give the first identity; substituting it into the definition of $\mathrm{P}(B_j \mid A)$ gives the second. □

Bayes is rarely the point on the Putnam. The law of total probability is the point constantly, because it is the engine of first step analysis: partition according to what happens first and you have a recursion.

Lemma 10.2.4 (Functions of disjoint blocks are independent)**.** *If $(X_i)_{i\in I}$ is independent and $S_1, S_2, \ldots$ are pairwise disjoint subsets of I, then the variables $f_j\left((X_i)_{i\in S_j}\right)$ are independent.*

This looks like bookkeeping and it is load bearing. It licenses treating the $\binom{2n}{2}$ variables $M_{ij} = A_{ij} - A_{ji}$ as independent in Putnam 2016 B4, and it makes the medians of disjoint triples independent in Putnam 2021 B6, where overlapping triples would leave no scalar recursion at all.

- **WHEN** you want to treat derived quantities as independent **DO** check that each is a function of a block of the underlying independent family and that the blocks are pairwise disjoint **BECAUSE** independence of the derived variables follows from disjointness of the blocks, and with overlapping blocks the claim is usually false (Putnam 2016 B4, 2021 B6)

10.3 Expectation, linearity and the indicator method

Definition 10.3.1 (Expectation)**.** For countable Ω, $\mathrm{E}[X] = \sum_\omega X(\omega)\,\mathrm{P}(\{\omega\})$ when the sum converges absolutely; for X with density f, $\mathrm{E}[X] = \int_{\mathbb{R}} x f(x)\,dx$. In both cases $\mathrm{E}[X]$ is the integral of X against P.

Theorem 10.3.2 (Linearity of expectation)**.** *If $X_1, \ldots, X_m$ have finite expectations then $\mathrm{E}[X_1 + \cdots + X_m] = \mathrm{E}[X_1] + \cdots + \mathrm{E}[X_m]$. No independence, no identical distribution, no positivity is required.*

Proof. On a countable Ω, absolute convergence permits the interchange $\sum_\omega \sum_j X_j(\omega)\,\mathrm{P}(\omega) = \sum_j \sum_\omega X_j(\omega)\,\mathrm{P}(\omega)$. In general it is linearity of the integral. □

Read the hypothesis list again: the absent hypotheses are the point, and the summands may be as dependent as you like. That is why expectation beats probability for counting: a union needs inclusion and exclusion, a mean does not.

Lemma 10.3.3 (Indicators)**.** *For an event A,* $\mathrm{E}[\mathbf{1}[A]] = \mathrm{P}(A)$.

- **WHEN** a problem asks for the average, mean or expected value of a count **DO** write the count as a sum of indicators, one per position or per object, and add the individual probabilities **BECAUSE** E is linear with no independence hypothesis, so heavy dependence between the indicators costs you nothing (Putnam 2006 A4, 2016 B4, 2020 A4, 2021 B1, 2022 A4, 2023 B3)

Example 10.3.4 (Putnam 2006 A4)**.** For a uniformly random permutation of $\{1, \dots, n\}$, let I_k indicate a local maximum at k. At $k = 1$ the condition is $\pi(1) > \pi(2)$, of probability $\frac{1}{2}$ by symmetry, and likewise at $k = n$; at interior k it says $\pi(k)$ is the largest of three values, of probability $\frac{1}{3}$ by exchangeability. So the answer is $\frac{1}{2} + \frac{1}{2} + (n-2)/3 = (n+1)/3$. The indicators are violently dependent, a local maximum at k forbidding one at $k+1$, and linearity does not care.

Example 10.3.5 (Putnam 2023 B3)**.** Let a be the length of the longest zigzag subsequence of $X_1, \dots, X_n$ drawn independently and uniformly from $[0, 1]$. The structural step is $a = 2 + \#\{\text{interior local extrema}\}$ for $n \geq 2$: you may always take the first and last terms, and each change of direction inside forces one more. With $\mathrm{P}(\text{three consecutive terms not monotone}) = \frac{2}{3}$ by exchangeability, $\mathrm{E}[a] = 2 + 2(n-2)/3 = (2n+2)/3$.

Notice what I did not do in either example: I did not compute the distribution of the count. That is the nominal solution and it costs enormous ink for a question that asked only for a mean. The bifurcation point is the phrase "expected number of".

Example 10.3.6 (Putnam 2016 B4, mean zero factors annihilate monomials)**.** A is $2n \times 2n$ with independent entries uniform on $\{0, 1\}$ and $M = A - A^t$, so $M_{ii} = 0$ and for $i < j$ the M_{ij} are independent, taking $-1, 0, 1$ with probabilities $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$. Expand $\det M = \sum_\sigma \operatorname{sgn}(\sigma) \prod_i M_{i\sigma(i)}$. A fixed point contributes $M_{ii} = 0$; a cycle of length at least three contains an entry to the first power, independent of the rest, so that term has mean zero. Only perfect matchings survive, and for these $\operatorname{sgn}(\sigma) = (-1)^n$ while $\prod_i M_{i\sigma(i)} = (-1)^n \prod_{\text{pairs}} M_{ij}^2$, so the signs cancel and the mean is $\prod \mathrm{E}[M_{ij}^2] = 2^{-n}$. With $(2n)!/(2^n n!)$ matchings,

$$\mathrm{E}[\det(A - A^t)] = \frac{(2n)!}{2^n n!} \cdot \frac{1}{2^n} = \frac{(2n)!}{4^n\, n!}.$$

Extract the rule: a monomial in independent mean zero variables has expectation zero as soon as one variable occurs to an odd power. That kills all but a vanishing fraction of the $(2n)!$ terms.

Theorem 10.3.7 (Tail sum)**.** *If Z takes values in $\{0, 1, 2, \dots\} \cup \{\infty\}$ then $\mathrm{E}[Z] = \sum_{n \geq 0} \mathrm{P}(Z > n)$, both sides finite or infinite together; if $Y \geq 0$ is real valued then $\mathrm{E}[Y] = \int_0^\infty \mathrm{P}(Y > t)dt$.*

Proof. $Z = \sum_{n \geq 0} \mathbf{1}[Z > n]$ pointwise, and Tonelli permits interchanging expectation with a sum of nonnegative terms; the continuous version uses $\int_0^\infty \mathbf{1}[Y > t]dt = Y$. □

This is the escape hatch whenever the law of Z is horrible but the events $\{Z > n\}$ are transparent, because those events usually depend on the first n steps only.

- **WHEN** you need the mean of a nonnegative integer valued variable with an unpleasant distribution **DO** write $\mathrm{E}[Z] = \sum_{n \geq 0} \mathrm{P}(Z > n)$ and look at the tail events **BECAUSE** the identity $Z = \sum_n \mathbf{1}[Z > n]$ replaces the law of Z by events determined by the first n steps (Putnam 2020 B3, 2022 A4)
- **WHEN** the count has a random upper limit, $\sum_{i=1}^{k} Y_i$ with k random **DO** rewrite it as $\sum_{i \geq 1} Y_i \mathbf{1}[i \leq k]$ and take expectations term by term **BECAUSE** the rewritten sum has a fixed index set, so linearity applies with no independence between Y_i and $\mathbf{1}[i \leq k]$ (Putnam 2022 A4)

Example 10.3.8 (Putnam 2022 A4, worked)**.** $X_1, X_2, \dots$ independent uniform on $[0,1]$, k least with $X_k < X_{k+1}$, and $S = \sum_{i=1}^{k} X_i/2^i$. Fire both circuits: $\mathrm{E}[S] = \sum_{i\geq 1} 2^{-i}\,\mathrm{E}[X_i \mathbf{1}[k \geq i]]$. Now $\{k \geq i\}$ says no comparison succeeded before i, that is $X_1 > \cdots > X_i$, an event depending on $X_1, \dots, X_i$ only, which is what removes X_{i+1} and makes the computation one dimensional. Condition on $X_i = t$:

$$\mathrm{E}[X_i \mathbf{1}[k \geq i]] = \int_0^1 t\,\frac{(1-t)^{i-1}}{(i-1)!}dt = \frac{1}{(i-1)!\,i(i+1)} = \frac{1}{(i+1)!},$$

using that $\{t < x_{i-1} < \cdots < x_1 < 1\}$ is a simplex of volume $(1-t)^{i-1}/(i-1)!$. Hence

$$\mathrm{E}[S] = \sum_{i\geq 1} \frac{(1/2)^i}{(i+1)!} = 2\left(e^{1/2} - 1 - \tfrac{1}{2}\right) = 2\sqrt{e} - 3.$$

Install the simplex volume on its own: for independent uniforms, $\mathrm{P}(X_1 > \cdots > X_m > t) = (1-t)^m/m!$, because the $m!$ orderings of m points in $(t,1)$ are equally likely and the magnitude event has probability $(1-t)^m$. That is exchangeability doing the integration for you.

10.4 Variance, covariance and second moment facts

Definition 10.4.1. $\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X]^2$ and $\mathrm{Cov}(X,Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]$.

Proposition 10.4.2. $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + 2\,\mathrm{Cov}(X,Y) + \mathrm{Var}(Y)$; $\mathrm{Cov}(X,Y)^2 \leq \mathrm{Var}(X)\,\mathrm{Var}(Y)$; *and independence gives* Cov $= 0$, *so variances of independent summands add.*

Proof. The first is expansion, the second is Cauchy and Schwarz for the centred variables under $\langle U, V\rangle = \mathrm{E}[UV]$, the third is $\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$. □

Putnam 2013 B4 asks for $\mathrm{Var}(fg) \leq 2\,\mathrm{Var}(f)M(g)^2 + 2\,\mathrm{Var}(g)M(f)^2$ on $[0,1]$. The cue is the word Var: the problem has told you the ambient space is $L^2[0,1]$ under the uniform measure, so read f, g as random variables, split $f = \mu(f) + f_0$ with $\mathrm{E}[f_0] = 0$, and absorb the cross terms with the covariance bound. Putnam 2023 A1 reverses it: for independent fair signs $\mathrm{E}[\exp(ix\sum \varepsilon_m m)] = \prod_m \cos(mx)$, and since $\phi''(0)$ is minus the second moment, additivity of the quadratic coefficients is additivity of variance.

Theorem 10.4.3 (Chebyshev's correlation inequality)**.** *If g, h are increasing and $g(X), h(X), g(X)h(X)$ are integrable then* $\mathrm{Cov}(g(X), h(X)) \geq 0$, *strictly if both are strictly increasing and X is not a point mass.*

Proof. Let X' be an independent copy. Then $(g(X) - g(X'))(h(X) - h(X')) \geq 0$ pointwise, and its expectation is $2\,\mathrm{Cov}(g(X), h(X))$. Strictness holds because the product is strictly positive whenever $X \neq X'$. □

That symmetrisation is the whole content and it makes the inequality visible. Putnam 2025 B2 is exactly it: with $f : [0,1] \to [0,\infty)$ strictly increasing and X of density proportional to f, the centroid of the region has $x_1 = \mathrm{E}[X]$ while the solid of revolution has $x_2 = \mathrm{E}[Xf(X)]/\,\mathrm{E}[f(X)]$, so

$$x_2 - x_1 = \frac{\mathrm{Cov}(X, f(X))}{\mathrm{E}[f(X)]} > 0.$$

One line, and the reason is visible: rotating weights the region by an extra increasing factor, so mass moves right.

- **WHEN** you must compare two weighted averages of x whose weights differ by a monotone factor **DO** write both as expectations under one law and read the difference as a covariance **BECAUSE** $\mathrm{Cov}(X, h(X)) \geq 0$ for increasing h by the symmetrisation $(X - X')(h(X) - h(X')) \geq 0$ (Putnam 2025 B2)

- **WHEN** a problem hands you μ, Var or a sup norm on $[0,1]$ **DO** treat the functions as random variables on $[0,1]$ and use variance identities instead of expanding integrals **BECAUSE** the language is already probabilistic, so $\operatorname{Var}(f+g)$ and $\operatorname{Cov}^2 \le \operatorname{Var}\operatorname{Var}$ apply and cost far less ink (Putnam 2013 B4)

Theorem 10.4.4 (Markov and Chebyshev)**.** *For $Y \ge 0$ and $t > 0$, $\mathrm{P}(Y \ge t) \le \mathrm{E}[Y]/t$; hence $\mathrm{P}(|X - \mathrm{E}\,X| \ge t) \le \operatorname{Var}(X)/t^2$.*

Proof. $t\mathbf{1}[Y \ge t] \le Y$ pointwise; for Chebyshev apply this to $(X - \mathrm{E}\,X)^2$ at threshold t^2. □

Markov looks too weak to be useful and is the base of every concentration bound here, because you choose which nonnegative variable to apply it to. Applied to $e^{\theta X}$ it becomes Chernoff.

10.5 Conditioning as machinery: tower law, first step analysis, stopping

Definition 10.5.1. If Y takes countably many values, $\mathrm{E}[X \mid Y]$ is the random variable equal to $\mathrm{E}[X \mid Y = y]$ on the event $\{Y = y\}$.

Theorem 10.5.2 (Tower law)**.** $\mathrm{E}[X] = \mathrm{E}\left[\mathrm{E}[X \mid Y]\right]$.

Proof. $\sum_y \mathrm{E}[X \mid Y = y]\,\mathrm{P}(Y = y) = \sum_y \sum_x x\,\mathrm{P}(X = x, Y = y) = \mathrm{E}[X]$. □

Take Y to be "what happened at the first step" and the tower law converts a global average into a one step average of the same quantity at the new state. That is a recursion, and Section 11 lives on it.

Theorem 10.5.3 (Affine drift)**.** *If $\mathrm{E}[X_{k+1} \mid X_k] = \lambda X_k + c$ then $e_k = \mathrm{E}[X_k]$ satisfies $e_{k+1} = \lambda e_k + c$, so for $\lambda \ne 1$, with $\mu = c/(1-\lambda)$, $e_k = \mu + \lambda^k(e_0 - \mu)$.*

Proof. Take expectations in the hypothesis and apply the tower law; the closed form is checked by substitution. □

The conditional law of X_{k+1} may be complicated, but only its mean enters, and if that mean is affine the sequence of means closes on itself with no knowledge of the law of X_k.

Example 10.5.4 (Putnam 2024 B4, worked)**.** With $a_{n,0} = 1$, a uniform $m \in \{1, \dots, n\}$ sends the state up with probability $(n-a)/n$, leaves it with probability $1/n$ and sends it down with probability $(a-1)/n$, so

$$\mathrm{E}[a_{n,k+1} \mid a_{n,k} = a] = a + \frac{n-a}{n} - \frac{a-1}{n} = \left(1 - \frac{2}{n}\right) a + \frac{n+1}{n}.$$

Affine drift with $\mu = (n+1)/2$ gives $e_k = \frac{n+1}{2} - \frac{n-1}{2}\left(1 - \frac{2}{n}\right)^k$, so $E(n)/n \to \frac{1}{2} - \frac{1}{2}e^{-2} = (1 - e^{-2})/2$.

Carry the second reading too, because it generalises differently. Interpreting the state as the number of lit bulbs among n with one stuck on, each free bulb is toggled a $\mathrm{Bin}(n, 1/n)$ number of times, so $E(n) = 1 + (n-1)\,\mathrm{P}(\mathrm{Bin}(n,1/n) \text{ odd})$, and since $\mathrm{P}(\mathrm{Bin}(m,p) \text{ odd}) = \left(1 - (1-2p)^m\right)/2$, obtained by subtracting $(q-p)^m$ from $(q+p)^m$, the limit is the same. Now the e^{-2} is explained rather than computed: it is the Poisson limit of $\mathrm{Bin}(n, 1/n)$. That chain is the Ehrenfest urn with one ball glued down, and recognising it supplies the check.

- **WHEN** a process has a one step conditional mean affine in the current state **DO** write $e_{k+1} = \lambda e_k + c$ at once and solve; compute no distribution **BECAUSE** the tower law needs only the conditional mean, so affine drift closes the recursion for the means alone (Putnam 2024 B4)

- **WHEN** a random process is described by a rule for its first step **DO** condition on that step, name the resulting quantity for every state, and write the system **BECAUSE** the law of total probability turns "what happens overall" into "one step plus the same problem from the new state", which is the only source of a recursion (Putnam 2017 A5, 2020 A4, 2020 B3)

Example 10.5.5 (Putnam 2017 A5, rotating roles)**.** Three players draw in the order A, B, C; the drawn card is uniform on $1, \dots, n$, the deck shrinks to that value minus one, and the roles rotate. With a_n, b_n, c_n the winning probabilities of the player to move, the next player and the third,

$$a_n = \frac{1}{n}\sum_{k=1}^{n} c_{k-1}, \qquad b_n = \frac{1}{n}\sum_{k=1}^{n} a_{k-1}, \qquad c_n = \frac{1}{n}\sum_{k=1}^{n} b_{k-1},$$

with $a_0 = b_0 = 0$ and $c_0 = 1$, the last encoding that whoever empties the deck has won. Section 11 gives the device that uncouples such a system.

Proposition 10.5.6 (Gambler's ruin)**.** *For the simple symmetric walk on $\{0, \dots, N\}$ started at j and absorbed at the ends, $\mathrm{P}(\text{absorb at } N) = j/N$ and the expected absorption time is $j(N-j)$.*

Proof. First step analysis gives $h(j) = \frac{1}{2}h(j-1) + \frac{1}{2}h(j+1)$, so h has constant increments and the boundary values 0 and 1 force $h(j) = j/N$. For the mean time, $T(j) = 1 + \frac{1}{2}T(j-1) + \frac{1}{2}T(j+1)$ makes the second difference -2, so T is a quadratic with leading coefficient -1, and $T(0) = T(N) = 0$ give $T(j) = j(N-j)$. □

The asymmetric case is the same computation with $h(j) = \big(1 - (q/p)^j\big) \,/\, \big(1 - (q/p)^N\big)$: derive it, do not memorise it.

Theorem 10.5.7 (Wald's identity)**.** *For independent identically distributed integrable Y_i and a stopping time T with $\mathrm{E}[T] < \infty$, $\mathrm{E}[Y_1 + \cdots + Y_T] = \mathrm{E}[T]\,\mathrm{E}[Y_1]$.*

Proof. $S_T = \sum_{i \geq 1} Y_i \mathbf{1}[T \geq i]$, and $\{T \geq i\}$ is determined by $Y_1, \dots, Y_{i-1}$, hence independent of Y_i. So each term has mean $\mathrm{E}[Y_1]\,\mathrm{P}(T \geq i)$, and summing with the tail sum formula for $\mathrm{E}[T]$ gives the result. □

That proof is the indicator decomposition again, with the extra observation that $\{T \geq i\}$ is decided before Y_i is seen. That observation is the definition of a stopping time and the only place the phrase earns its keep.

Definition 10.5.8 (Martingale)**.** (M_k) is a martingale for the information $(\mathcal{F}_k)$ if $\mathrm{E}\,|M_k| < \infty$ and $\mathrm{E}[M_{k+1} \mid \mathcal{F}_k] = M_k$.

Theorem 10.5.9 (Optional stopping)**.** *If (M_k) is a martingale and T is a bounded stopping time, or one for which $(M_{k \wedge T})$ is uniformly integrable, then $\mathrm{E}[M_T] = \mathrm{E}[M_0]$.*

The bounded case is induction on the bound plus the tower law; the uniformly integrable case needs a convergence theorem, which is the honest limit of what you can reconstruct in the arena, so quote it by name. At Putnam level martingales appear in disguise: in gambler's ruin S_k and $S_k^2 - k$ are martingales and optional stopping reproduces both answers instantly, and in Putnam 2020 B3 memorylessness makes the overshoot exponential, so Wald gives $\mathrm{E}[Z] = L + 1$ with the additive 1 as that overshoot.

- **WHEN** a process is stopped when it first crosses a level and you want the mean stopping index **DO** look for a quantity of zero drift, or apply Wald to the increments **BECAUSE** both convert "mean at a random time" into "mean time times mean increment", avoiding the law of the stopping time (Putnam 2020 B3)

A Markov chain on a finite state space is a matrix M of transition probabilities; if it is irreducible and aperiodic there is a unique stationary law and the distance to it decays like b^m, with b the largest modulus of a nontrivial eigenvalue. Putnam 2011 A6 is the cleanest instance: a step law supported on a generating set of a finite abelian group gives $M_{x,y} = P(x^{-1}y)$, doubly stochastic, with the characters of G as eigenvectors and eigenvalue $\sum_g P(g)\chi(g)$ at χ. That is the whole structural content: the support generates, so the chain is irreducible; it contains the identity, so the chain is aperiodic and $b < 1$; it is proper, so some nontrivial character survives and $b > 0$. Parseval finishes.

- **WHEN** a random walk multiplies by random generators of a finite group **DO** diagonalise with characters: the eigenvalue at χ is $\sum_g P(g)\chi(g)$, and uniform is stationary because the matrix is doubly stochastic **BECAUSE** the transition matrix is a convolution operator whose eigenvectors are exactly the characters, so mixing becomes a question about the largest nontrivial Fourier coefficient (Putnam 2011 A6)

10.6 The distributions you must be able to quote

Each entry gives the law, the moments and the one structural fact that makes it appear. Not to be looked up.

Uniform. U on $[0,1]$ has mean $\frac{1}{2}$, variance $\frac{1}{12}$, $\mathrm{P}(U \le t) = t$. Two structural facts: tU is uniform on $[0,t]$, so "x_{n+1} is uniform on $[0,x_n]$" is literally $x_{n+1} = x_n U_{n+1}$, hence $x_n = U_1 \cdots U_n$; and conditioning a uniform law on a subregion returns the uniform law there, which permits rescaling (Putnam 2006 A6, 2020 B3).

Binomial. $\mathrm{Bin}(m,p)$ has mean mp, variance $mp(1-p)$, generating function $(1-p+ps)^m$. Quote the parity formula $\mathrm{P}(\mathrm{Bin}(m,p) \text{ odd}) = \left(1-(1-2p)^m\right)/2$ and the symmetry $\mathrm{P}(\mathrm{Bin}(2k+1,\frac{1}{2}) \ge k+1) = \frac{1}{2}$, since in an odd number of fair flips exactly one of "majority heads" and "majority tails" occurs.

Geometric and negative binomial. The first success has $\mathrm{P}(X=n) = (1-p)^{n-1}p$, mean $1/p$, and is the unique memoryless law on $\{1,2,\dots\}$; the $(k+1)$st has $\mathrm{P}(X=n) = \binom{n-1}{k}p^{k+1}(1-p)^{n-k-1}$.

Example 10.6.1 (Putnam 2020 A2, worked)**.** Evaluate $\sum_{j=0}^{k} 2^{k-j}\binom{k+j}{j}$. Put $n = k+j+1$, so $\binom{k+j}{j} = \binom{n-1}{k}$ and $2^{k-j} = 2^{2k+1-n}$, giving

$$\sum_{j=0}^{k} 2^{k-j}\binom{k+j}{j} = 2^{2k+1}\sum_{n=k+1}^{2k+1}\binom{n-1}{k}2^{-n} = 2^{2k+1}\,\mathrm{P}(X \le 2k+1),$$

where X is the index of the $(k+1)$st head in fair flips. The event $\{X \le 2k+1\}$ is "at least $k+1$ heads in $2k+1$ flips", of probability $\frac{1}{2}$, so the sum is 4^k. The cue was $\binom{n-1}{k}2^{-n}$, which is a waiting time law and nothing else; without that fluency you induct, and induction is the nominal solution here.

Poisson. $\mathrm{P}(N=j) = e^{-\lambda}\lambda^j/j!$, mean and variance λ, generating function $e^{\lambda(s-1)}$; it is the limit of $\mathrm{Bin}(m,\lambda/m)$ and the count of points of a Poisson process.

Exponential and gamma. $\mathrm{P}(E > t) = e^{-\lambda t}$, mean $1/\lambda$, variance $1/\lambda^2$.

Lemma 10.6.2 (Memorylessness and the probability integral transform)**.** *The exponential laws are the only ones on $[0,\infty)$ with $\mathrm{P}(E > s+t \mid E > s) = \mathrm{P}(E > t)$, since $G(s+t) = G(s)G(t)$ for the monotone $G(t) = \mathrm{P}(E > t)$ forces $G(t) = e^{-\lambda t}$. If U is uniform on $[0,1]$ then $-\log U$ is exponential of rate one, since $\mathrm{P}(-\log U > t) = \mathrm{P}(U < e^{-t}) = e^{-t}$; more generally $F^{-1}(U)$ has continuous strictly increasing distribution function F, and $F(X)$ is uniform when X has distribution function F.*

Sums of independent rate one exponentials are gamma: T_m has density $t^{m-1}e^{-t}/(m-1)!$, proved by the convolution induction $g_m(t) = \int_0^t g_{m-1}(s)e^{-(t-s)}ds$. The partial sums are the arrival times of a rate one Poisson process, so $N(t) = \#\{m \ge 1 : T_m \le t\}$ is $\mathrm{Poisson}(t)$.

Example 10.6.3 (Putnam 2020 B3, worked end to end). $x_0 = 1$, x_{n+1} uniform on $[0, x_n]$, and Z least with $x_n < \delta$. Chain the facts: $x_n = U_1 \cdots U_n$ by scaling; $-\log x_n = T_n$ a sum of rate one exponentials; with $L = \log(1/\delta)$, the event $\{Z > n\}$ is $\{T_n \le L\}$; and by the tail sum formula

$$\mathrm{E}[Z] = 1 + \sum_{n \ge 1} \mathrm{P}(T_n \le L) = 1 + \mathrm{E}[N(L)] = 1 + \log \frac{1}{\delta}.$$

Every step was a named circuit. Nothing there had to be invented in the arena.

Normal. Density $e^{-x^2/2}/\sqrt{2\pi}$, moment generating function $e^{\theta^2/2}$, even moments $(2m-1)!!$; quote the central limit theorem by name. On the Putnam the normal is almost never the object: it tells you the right scaling before you look for an exact argument, so treat it as a compass, not a tool.

- **WHEN** a recursion says the next term is uniform on [0, current term] **DO** write $x_{n+1} = x_n U_{n+1}$, take logarithms, and read off a random walk with exponential increments **BECAUSE** scaling makes the recursion multiplicative and $-\log U$ is exponential, so the object becomes a Poisson process and every quantity is quotable (Putnam 2020 B3)
- **WHEN** a binomial appears with p of order $1/m$ **DO** pass to the Poisson limit for the shape of the answer, then confirm with the exact binomial identity **BECAUSE** $\mathrm{Bin}(m, \lambda/m) \to \mathrm{Poisson}(\lambda)$ explains constants such as e^{-2} that otherwise look unmotivated (Putnam 2024 B4)

10.7 Order statistics, exchangeability and the quantile transform

Proposition 10.7.1. *For independent Y_i with common distribution function F, $\max_i Y_i$ has distribution function F^m and density $mF^{m-1}f$, and the jth smallest has distribution function $\sum_{i=j}^{m} \binom{m}{i} F^i (1-F)^{m-i}$.*

Proof. $\{\max_i Y_i \le x\} = \bigcap_i \{Y_i \le x\}$; and $\{Y_{(j)} \le x\}$ says at least j of the values are at most x, a $\mathrm{Bin}(m, F(x))$ count. □

Lemma 10.7.2 (Exchangeability). *For independent identically distributed variables with a continuous law, almost surely no two are equal, each of the $m!$ relative orderings has probability $1/m!$, and the rank vector is independent of the vector of order statistics.*

This gives $\mathrm{P}(X_1 > \cdots > X_m) = 1/m!$ with no integration, the $\frac{1}{3}$ in Putnam 2006 A4, the $\frac{2}{3}$ in Putnam 2023 B3, and the factorisation of the descending run event in Putnam 2022 A4.

Theorem 10.7.3 (Quantile transform and monotone coupling). *A single uniform U realises a whole family of laws simultaneously through $F_\alpha^{-1}(U)$, monotonically in α. Since $(\varphi \circ F)^{-1} = F^{-1} \circ \varphi^{-1}$, a recursion $F_k = \varphi \circ F_{k-1}$ on distribution functions becomes the pointwise recursion $Y_k = \varphi^{-1}(Y_{k-1})$ driven by that one seed.*

Example 10.7.4 (Putnam 2021 B6, worked). Trim a list by replacing each block of three consecutive entries by its median. If the entries have distribution function F, the median of three independent copies has distribution function $\varphi(F) = 3F^2 - 2F^3$, since the median is at most x exactly when at least two of the three are, a $\mathrm{Bin}(3, F(x))$ count. Medians of disjoint triples are independent, so after k rounds the survivors have $F_k = \varphi^k$ applied to $F_0(x) = x$.

The nominal route now tries to prove $\mu_k \ge \frac{2}{3}\mu_{k-1}$ for $\mu_k = \mathrm{E}\,|X_k - \frac{1}{2}|$ by bounding φ below in the value slot. That route is not merely long, it is impossible: $\varphi(u) \ge \frac{2}{3}u$ fails near $u = 0$ where $\varphi(u) \approx 3u^2$. So the bifurcation point is visible, and the inequality must leave the value slot.

Invert. Let $\psi = \varphi^{-1}$ and drive everything by one uniform, so $X_k = \psi^k(U)$. Since $\varphi(\frac{1}{2}) = \frac{1}{2}$ and $\varphi'(y) = 6y(1-y) \le \frac{3}{2}$,

$$\left|\varphi(y) - \tfrac{1}{2}\right| \le \tfrac{3}{2}\left|y - \tfrac{1}{2}\right|, \qquad \text{hence} \qquad \left|\psi(u) - \tfrac{1}{2}\right| \ge \tfrac{2}{3}\left|u - \tfrac{1}{2}\right|$$

for every u, the second being the first with $u = \varphi(y)$. Apply it with $u = \psi^{k-1}(U)$ and take expectations: $\mu_k \ge \frac{2}{3}\mu_{k-1}$. With $\mu_0 = \mathrm{E}\,|U - \frac{1}{2}| = \frac{1}{4}$, after 2021 rounds $\mu \ge \frac{1}{4}\left(\frac{2}{3}\right)^{2021}$.

Extract the rule. Inequalities between distributions do not iterate, since a bound on a distribution function says nothing about the next round's shape, while inequalities between random variables on a common space do, because they hold pointwise. The quantile transform is the bridge and it costs one line.

- **WHEN** a process is described by a recursion on distributions or distribution functions **DO** couple everything to a single uniform seed by quantile functions and rewrite as a pointwise recursion on random variables **BECAUSE** pointwise inequalities compose under iteration and distributional ones do not, which is exactly why the value side bound fails in Putnam 2021 B6
- **WHEN** a symmetry is available and a scaling parameter is free **DO** condition on the extreme order statistic and rescale so the extreme becomes fixed **BECAUSE** max of m independent copies has distribution function F^m, so conditioning costs one density factor and spends the scaling symmetry once (Putnam 2006 A6)
- **WHEN** the problem involves the relative order of independent continuous variables **DO** use exchangeability: all $m!$ orders have probability $1/m!$ and the ranks are independent of the values **BECAUSE** ties have probability zero and permuting identically distributed independent variables preserves the law, so ordering probabilities are pure counting (Putnam 2006 A4, 2022 A4, 2023 B3)

10.8 Transforms: generating functions, moments and Chernoff

Definition 10.8.1. $G_X(s) = \mathrm{E}[s^X]$ for X in $\{0,1,2,\dots\}$; $M_X(\theta) = \mathrm{E}[e^{\theta X}]$; $\phi_X(\theta) = \mathrm{E}[e^{i\theta X}]$.

Proposition 10.8.2. *For independent X, Y each transform of the sum is the product of the transforms; $\phi_{aX}(\theta) = \phi_X(a\theta)$; $G_X^{(m)}(1)$ is the mth factorial moment $\mathrm{E}[X(X-1)\cdots(X-m+1)]$; $M_X^{(m)}(0) = \mathrm{E}[X^m]$ and $\phi_X''(0) = -\mathrm{E}[X^2]$.*

Independence becomes multiplication. That is the entire reason transforms exist, and it is the cue: when a problem sums independent contributions and asks about the sum, a transform turns the sum into a product.

Theorem 10.8.3 (Fourier inversion for integer valued variables)**.** $\mathrm{P}(X = b) = \frac{1}{2\pi}\int_0^{2\pi} e^{-ib\theta}\phi_X(\theta)d\theta$.

Proof. Expand $\phi_X(\theta) = \sum_j \mathrm{P}(X = j)e^{ij\theta}$ and integrate term by term against $e^{-ib\theta}$, using $\int_0^{2\pi} e^{i(j-b)\theta}d\theta = 2\pi\delta_{jb}$; absolute convergence licenses the interchange. □

Example 10.8.4 (Putnam 2022 B5, worked)**.** Independent X_i take $1, -1, 0$ with probabilities $p, p, 1-2p$, so $\phi_{X_i}(\theta) = 1 - 2p(1-\cos\theta)$, whose minimum $1-4p$ is nonnegative exactly when $p \le \frac{1}{4}$. For such p and $S = \sum_i a_i X_i$ we get $\phi_S(\theta) = \prod_i \phi_{X_i}(a_i\theta) \ge 0$, so by inversion

$$\mathrm{P}(S = b) \le \frac{1}{2\pi}\int_0^{2\pi} \phi_S(\theta)d\theta = \mathrm{P}(S = 0).$$

For $p > \frac{1}{4}$ the property fails: with $(a_1, a_2) = (1, 2)$, $\mathrm{P}(S = 0) = (1-2p)^2$ while $\mathrm{P}(S = 1) = p(1-2p) + p^2$, and the second wins once $p > \frac{1}{3}$; a longer vector such as $(1, 2, 2, 4)$ pushes the failure down to every $p > \frac{1}{4}$. The answer is $0 \le p \le \frac{1}{4}$.

The strengthening worth carrying: a sequence arising from a nonnegative characteristic function is positive definite in the sense of Herglotz, and such sequences satisfy $|c_b| \le c_0$. That is why the class is closed under convolution, and why nonnegativity of a single step is the right thing to demand.

Prescribing $\mathrm{E}[X^k]$ for $k \le d$ determines $\mathrm{E}[f(X)]$ for every polynomial of degree at most d and for nothing else; the set of laws with given moments up to degree d is convex, so every sharp bound on a linear functional comes from sandwiching it between polynomials of degree at most d. That converts a moment problem into a search for one polynomial.

Example 10.8.5 (Putnam 2014 A4, worked)**.** $X \geq 0$ integer valued with $\mathrm{E}[X] = 1$, $\mathrm{E}[X^2] = 2$, $\mathrm{E}[X^3] = 5$; minimise $\mathrm{P}(X = 0)$. Lower bound: any cubic f with $f(0) \leq 1$ and $f(j) \leq 0$ for integers $j \geq 1$ satisfies $\mathbf{1}[X = 0] \geq f(X)$ pointwise, so $\mathrm{P}(X = 0) \geq \mathrm{E}[f(X)]$, a number computed from the data. Upper bound: on support $\{0, 1, 3\}$ the moment equations $p_1 + 3p_3 = 1$, $p_1 + 9p_3 = 2$, $p_1 + 27p_3 = 5$ give $p_3 = \frac{1}{6}$, $p_1 = \frac{1}{2}$, $p_0 = \frac{1}{3}$. The answer is $\frac{1}{3}$. The signal that a generating function route exists is that in the factorial moment basis the data read $1, 1, 1$, since $\mathrm{E}[X^2] - \mathrm{E}[X] = 1$ and $\mathrm{E}[X^3] - 3\,\mathrm{E}[X^2] + 2\,\mathrm{E}[X] = 1$; three ones is a statement about G_X near $s = 1$ and says to build the certificate from falling factorials.

Theorem 10.8.6 (Chernoff)**.** *For independent X_i and $\theta > 0$, $\mathrm{P}\left(\sum_i X_i \geq a\right) \leq e^{-\theta a} \prod_i \mathrm{E}[e^{\theta X_i}]$.*

Proof. Apply Markov to $e^{\theta \sum X_i}$ at threshold $e^{\theta a}$ and factorise by independence. □

For counting: if $P(x) = \sum_m c_m x^m$ has nonnegative coefficients and $a(k, n) = [x^n]P(x)^k$, then $a(k, n) \leq x^{-n} P(x)^k$ for every $x > 0$, because all the other terms are nonnegative. That is Chernoff written without probability, and the optimal x solves $xP'(x)/P(x) = n/k$: tilt the one step law until its mean becomes the target average. The minimum is $e^{-kI(n/k)}$ with I the Legendre transform of $\log P(e^\theta)$, Cramer's rate function.

Example 10.8.7 (Putnam 2018 B6, worked)**.** With terms in $\{1, \ldots, 6, 10\}$ put $P(x) = x + x^2 + x^3 + x^4 + x^5 + x^6 + x^{10}$, so $|S| = [x^{3860}]P(x)^{2018} \leq x^{-3860} P(x)^{2018}$. At $x = \frac{1}{2}$, $P(\frac{1}{2}) = (512 + 256 + 128 + 64 + 32 + 16 + 1)/1024 = 1009/1024$, giving $|S| \leq 2^{3860}(2018/2048)^{2018}$ as required.

Do not file $x = \frac{1}{2}$ as a gift. Check stationarity: $xP'(x)$ at $x = \frac{1}{2}$ is $\sum_m m2^{-m} = 1930/1024$ over the allowed m, so $xP'(x)/P(x) = 1930/1009 = 3860/2018 = n/k$ exactly. The problem's n was chosen to make $x = \frac{1}{2}$ the optimal tilt, which also certifies that the bound is essentially sharp. Knowing the criterion means you generate $x = \frac{1}{2}$ instead of recognising it, and that is the difference between solving this in the arena and admiring it afterwards.

- **WHEN** you must bound the number of sequences with a prescribed sum **DO** write the count as a coefficient of a power of a generating polynomial and bound it by $x^{-n} P(x)^k$ **BECAUSE** all coefficients are nonnegative, so dropping the rest is legitimate, and this is Markov applied to $e^{\theta X}$ (Putnam 2018 B6)
- **WHEN** a Chernoff style bound has a free parameter **DO** choose it by the tilting equation $xP'(x)/P(x) = n/k$, making the tilted mean equal the target **BECAUSE** the bound is minimised at that stationary point and the exponent is Cramer's rate function, which also tells you whether the bound can be sharp
- **WHEN** independent contributions are summed and the question is about the sum **DO** move to G, M or ϕ so that independence becomes multiplication **BECAUSE** that factorisation is the only reason transforms exist (Putnam 2022 B5, 2023 A1)

10.9 Geometric probability

Lemma 10.9.1. *For U uniform on K and measurable $A \subseteq K$, $\mathrm{P}(U \in A) = |A|/|K|$; if A is random and independent of U then $\mathrm{P}(U \in A) = \mathrm{E}\,|A|/|K|$, by conditioning and Fubini.*

That lemma removes probability from a problem entirely. Putnam 2006 A6 asks for the probability that four uniform points in a disc are in convex position; the complement is that one lies inside the triangle of the other three, four disjoint events, so

$$\mathrm{P}(\text{convex}) = 1 - 4\,\mathrm{P}(S \in \triangle PQR) = 1 - \frac{4\,\mathrm{E}[T]}{\pi},$$

with T the area of the triangle on three uniform points. What remains is a mean area, computed by conditioning on the largest radius, whose density is $6c^5$, and rescaling; the value is $1 - 35/(12\pi^2)$. The named family is Sylvester's four point problem, worth $\frac{2}{3}$ for a triangle and $\frac{25}{36}$ for a square; Efron's identity $\mathrm{E}[V_m] = m\,(1 - \mathrm{E}[A_{m-1}]/|K|)$ cross checks it. Learn the family for the sanity checks.

Example 10.9.2 (Putnam 2021 B1, worked)**.** A unit square is dropped on the unit lattice with position and orientation independent of it. Fix the sample space first: the square has a rotation group of order four, so the angle θ is uniform on $[0, \pi/2)$ with density $2/\pi$ and the centre is uniform in a fundamental cell, independently.

Let N be the number of lattice points covered. The square has diameter $\sqrt{2}$ and side 1, so two covered points lie at distance 1 or $\sqrt{2}$, and the latter needs exact alignment, of probability zero; hence almost surely $N \in \{0, 1, 2\}$.

Averaging: translating a measurable set by a vector uniform modulo the lattice makes the expected number of lattice points in it equal to its area, since summing $|C \cap (P + K)|$ over lattice points reassembles K from the pieces $(C - P) \cap K$. So $\mathrm{E}[N] = 1$ for every θ, and combining $\mathrm{E}[N] = \mathrm{P}(N = 1) + 2\,\mathrm{P}(N = 2)$ with $\sum_j \mathrm{P}(N = j) = 1$ gives $\mathrm{P}(N = 0) = \mathrm{P}(N = 2)$: the diffuse event has been traded for a rigid one.

Finally, two covered points are lattice neighbours, so $\mathrm{P}(N = 2 \mid \theta)$ is the expected number of neighbouring pairs both covered, which by the same lemma is $|K_\theta \cap (K_\theta - v)|$ summed over $v = (1, 0)$ and $v = (0, 1)$. In the rotated frame $(1, 0)$ has coordinates $(\cos\theta, -\sin\theta)$, and the overlap of a unit square with its translate is the product of the one dimensional overlaps, namely $(1 - \cos\theta)(1 - \sin\theta)$; the vertical pair gives the same. Hence

$$\mathrm{P}(N = 0) = \frac{2}{\pi} \int_0^{\pi/2} 2(1 - \cos\theta)(1 - \sin\theta)\, d\theta = \frac{4}{\pi}\left(\frac{\pi}{2} - \frac{3}{2}\right) = 2 - \frac{6}{\pi}.$$

Putnam 2024 A5 needs decoupling: for independent uniform angles α, β, the chord joining $Re^{i\alpha}$ and $Re^{i\beta}$ is a uniformly rotated copy of the vertical chord at offset $R\cos\psi$, with rotation and ψ independent and uniform, because $(\alpha + \beta, \alpha - \beta)$ is again a pair of independent uniform angles, an integer matrix of nonzero determinant inducing a $|\det|$-to-one area multiplying map of the torus. Once decoupled, the hitting event is a one dimensional inequality per direction.

- **WHEN** a random point must land in a random region **DO** condition on the region and use $\mathrm{P}(U \in A) = \mathrm{E}\,|A|/|K|$ **BECAUSE** Fubini turns a probability about two random objects into a mean area (Putnam 2006 A6)
- **WHEN** a shape is dropped at random on a lattice **DO** compute the expected number of covered lattice points, which equals the area, and use it to trade a diffuse event for a rigid one **BECAUSE** translating by a uniform vector modulo the lattice reassembles the shape inside one cell, so the expected count is the area with no computation (Putnam 2021 B1)
- **WHEN** two independent uniform angles appear **DO** change variables to their sum and difference, again independent and uniform **BECAUSE** an integer matrix of nonzero determinant is measure preserving on the torus up to constant multiplicity (Putnam 2024 A5)

10.10 The probabilistic method and concentration

Theorem 10.10.1 (First moment method)**.** *If $\mathrm{E}[X] \le t$ then $X \le t$ somewhere; if $\mathrm{E}[X] = t$ and X is not almost surely constant then $X < t$ somewhere and $X > t$ somewhere.*

Proof. If $X > t$ everywhere then $\mathrm{E}[X] > t$. □

Trivial to prove, and it is an existence proof: to show an object with a property exists, construct a random object, compute the mean of a statistic, and conclude that some realisation beats the

mean. You never exhibit the object. Putnam 2021 B1 above is exactly this, as is Putnam 2024 A3, answered yes by the cells $(2,1)$ and $(1,2)$. Putnam 2013 B5 is the method run as an exact identity: a uniformly random $f : \{1,\dots,n\} \to \{1,\dots,n\}$ funnels every point into a fixed k element subset with probability exactly k/n, so the count is $n^n \cdot k/n = kn^{n-1}$, and the reformulation predicts before any algebra that the answer is linear in k.

For concentration the Putnam needs three tools and no more: Markov when you have only a mean, Chebyshev when you have a variance and want a two sided statement, Chernoff when the variable is a sum of independent bounded pieces and you need exponentially small tails. The second moment method mirrors the first: $\mathrm{E}[X] > 0$ with $\mathrm{Var}(X)$ small against $\mathrm{E}[X]^2$ gives $X > 0$ with positive probability. Putnam 2011 B5 has the counting version, occupancies q_z summing to n with $\sum_z q_z^2 \geq n^2/k$ by Jensen.

- **WHEN** a problem asks you to prove an object with some property exists and constructing one looks hopeless **DO** build a random object, compute the mean of the relevant statistic, and conclude that some realisation is at least as good as the mean **BECAUSE** $\mathrm{E}[X] \leq t$ forces $X \leq t$ somewhere, an existence proof that never exhibits the object (Putnam 2021 B1, 2024 A3)
- **WHEN** a count over a large family looks exactly computable **DO** divide by the family size and read it as a probability for a uniformly random member **BECAUSE** an exact probability often has a symmetry proof where the count has none, which is how kn^{n-1} becomes "probability k/n" (Putnam 2013 B5)

10.11 The recursion layer

Every recursion in this chapter, in the probability, games and logic material alike, belongs to one of the following families. Learn the family, not the instance.

First step recursions. $u(x) = \sum_y p_{xy} u(y)$ for a harmonic quantity, or $u(x) = 1 + \sum_y p_{xy} u(y)$ for an expected time, with u prescribed on the absorbing set. On a line this is a constant coefficient recurrence, so the general solution is a combination of geometric terms plus a particular solution, and the constants are pinned by the fact that the value vector is the unique bounded solution. Recognise it at "at each step, one of the following happens". Gambler's ruin; Putnam 2017 A5, 2020 A4, 2024 B4.

Affine drift and its matrix version. $x_{k+1} = \lambda x_k + c_k$ solves by variation of constants as $x_k = \lambda^k x_0 + \sum_{j<k} \lambda^{k-1-j} c_j$. For vector states λ becomes a transition matrix and the solution is spectral. Putnam 2024 B4, and the mixing rates of Putnam 2011 A6.

Rotating roles. $u_n^{(i)} = \sum_k p_{nk} u_k^{(i+1 \bmod r)}$. Pair the r sequences with a character of $\mathbb{Z}/r$, forming $v_n = \sum_i \omega^i u_n^{(i)}$; the system becomes a single scalar recursion and the originals are recovered by inverting the finite Fourier transform. Putnam 2017 A5 with $r = 3$.

Multiplicative recursions. $x_{n+1} = x_n U_{n+1}$ has solution $x_0 \prod_i U_i$, which logarithms turn into the additive random walk $\log x_n = \log x_0 + \sum_i \log U_i$, the only form in which crossing times are available. Putnam 2020 B3.

Convolution and renewal. $g_m = g_{m-1} * g_1$ solves by transforms, $\widehat{g_m} = \widehat{g}^m$, with renewal density $\widehat{g}/(1-\widehat{g})$; the gamma law of Putnam 2020 B3. The discrete twin is Cauchy's formula for repeated integration: $F_{i+1}(t) = \int_t^1 F_i$ with $F_1 \equiv 1$ gives $F_i(t) = (1-t)^{i-1}/(i-1)!$, the simplex volume of Putnam 2022 A4.

Recursions on distribution functions. $F_k = \varphi \circ F_{k-1}$ inverts to $Y_k = \varphi^{-1}(Y_{k-1})$ on one uniform seed, so the kth quantile function is the kth iterate of φ^{-1}. Putnam 2021 B6 with $\varphi = 3u^2 - 2u^3$; with medians of $2m+1$ the same argument runs with $\varphi_m = I_u(m+1, m+1)$ and constant $(2m+1)!/(m!\,m!\,4^m)$ in place of $\frac{3}{2}$.

Fragmentation. A random cut splits a structure of size N into parts of sizes $j-1$ and $N-j$ and a statistic is additive across the split, so $w(N) = \sum_j \mathrm{P}(\text{cut at } j)\,(w(j-1) + w(N-j))$. The licence is

linearity of expectation across the split: each part need only be marginally a copy of the original, never independent of the other. Putnam 2020 A4, where conditioning on the first square to change colour gives $2(N-1)$ equally likely effective cuts.

Backward recursions on games. $\mathcal{P}$ and $\mathcal{N}$ computed upward from the terminal positions, or $G(P) = \operatorname{mex}\{G(P') : P \to P'\}$. Guess the answer set and verify the fixed point conditions instead of descending through every level. Putnam 2006 A2, 2013 B6, 2020 B2. The scoring version is the mutual recursion $A(P) = \max_\mu B(P^\mu)$, $B(P) = \min_\mu A(P^\mu)$, well founded whenever every move strictly decreases a nonnegative integer statistic: Putnam 2022 A5.

Closure recursions. $S_0 = A$, $S_{k+1} = S_k \cup \bigcup_i f_i(S_k)$, with S the least fixed point. Two methods, one per direction: to prove S is everything, find a complexity measure strictly decreased by a composite and induct; to prove it is not, exhibit one closed proper superset. Putnam 2009 A4, 2017 A1, 2025 B3, all worked in Section 14.

Reduction to a smaller copy. After m moves the position is the initial position of the same game on a smaller parameter, so $V(n) = b(m) + V(n-m)$ with $b(m)$ banked in the first m moves. Putnam 2023 A6, where $V(n) \equiv n \pmod 2$.

Backward induction on a potential. To prove every state is reachable, show every non-initial state has an undo move to one of strictly smaller potential, valued in a well ordered set. Putnam 2023 B1, with the inversion number.

- **WHEN** a process problem is put in front of you and you do not know where to start **DO** ask which of the eleven families above it belongs to before writing anything **BECAUSE** each family has a general solution method and a general invariant, so identifying the family is worth more than any computation on the instance
- **WHEN** a recursion involves several interlocking sequences with cyclically permuted roles **DO** form the character combinations $\sum_i \omega^i u^{(i)}$ for each rth root of unity **BECAUSE** the cyclic symmetry is diagonalised by the characters of $\mathbb{Z}/r$, which uncouples the system into scalar recursions (Putnam 2017 A5)

10.12 Invariants, monovariants and the extremal principle

This is not a topic, it is a way of thinking, and it decides every process problem on the paper. A process problem is any problem saying "repeat the following move", and there are exactly two questions to ask about one: what cannot change, and what must change.

Definition 10.12.1. An invariant of a process is a function of the state that no legal move alters. A monovariant, or semi-invariant, is a function of the state that every legal move changes in one direction only. A complete invariant is one whose value determines the state up to the equivalence you care about.

Each does a specific job. An invariant proves *impossibility*: if I is constant and $I(\text{start}) \neq I(\text{target})$, the target is unreachable. A monovariant proves *termination*: an integer valued Φ, bounded above and strictly increasing at every move, forces the process to stop. A complete invariant proves *exact reachability*, the strongest and rarest.

Theorem 10.12.2 (Well ordering, in its three working forms)**.** *Every nonempty set of positive integers has a least element. Hence a nonincreasing sequence of positive integers is eventually constant; no infinite strictly decreasing such sequence exists, which is infinite descent; and any nonempty class of counterexamples has a minimal member, which is the extremal principle.*

Example 10.12.3 (Putnam 2025 A1, worked)**.** Distinct positive integers evolve by $m_k/n_k = (2m_{k-1}+1)/(2n_{k-1}+1)$ in lowest terms; prove $\gcd(2m_k+1, 2n_k+1) = 1$ for all but finitely many k. Let d_k be that gcd at stage $k-1$, so $\Delta_k = m_k - n_k$ satisfies $\Delta_k = 2\Delta_{k-1}/d_k$. Now d_k is odd, being a divisor

of odd numbers, and it divides $2\Delta_{k-1}$, hence divides Δ_{k-1} and so its odd part. Writing u_k for the odd part of $|\Delta_k|$, multiplication by 2 leaves it alone and division by d_k shrinks it, so $u_k = u_{k-1}/d_k$ is a nonincreasing sequence of positive integers, strictly decreasing exactly when $d_k > 1$. By well ordering it is eventually constant, so $d_k = 1$ eventually; each strict decrease divides u by at least 3, so there are at most $\log_3 u_0$ of them.

The command term "for all but finitely many" has exactly one proof shape: monovariant plus well ordering. There is no other.

Example 10.12.4 (Putnam 2008 A3, worked)**.** Repeatedly choose $j < k$ with $a_j \nmid a_k$ and replace the pair by gcd and lcm. Since $\gcd \cdot \operatorname{lcm} = a_j a_k$ and $a_j \nmid a_k$, the gcd is strictly below a_j and the lcm strictly above a_k. So $\Phi = \sum_i i \log a_i$ changes by $(k-j)\delta > 0$ with $\delta = \log(a_j/\gcd)$, and it is bounded above because for each prime the multiset $\{v_p(a_i)\}$ is invariant, the move replacing a pair by its min and max: the state space is finite, hence termination. At the end $a_j \mid a_k$ for all $j < k$, so $v_p(a_1) \le v_p(a_2) \le \cdots$, and with the multiset invariant each $v_p(a_i)$ is forced to be its ith smallest element. The final sequence is unique.

That shape has a name: a rewriting system that is *terminating* (no infinite rewrite sequence) and *confluent* (any two maximal sequences from one start reach the same normal form) has a unique normal form reachable by any strategy. Recognise the pair and you split such a problem correctly in the first minute.

Parity, colouring and weighting

Parity is the cheapest invariant in existence and you check it first, every time. In Putnam 2021 A1 each hop of length 5 is a signed permutation of $(3,4)$ or $(0,5)$, so it changes $x+y$ by an odd amount and after n hops $x + y \equiv x_0 + y_0 + n \pmod 2$; the target forces an even hop count, consistent with the answer 578, and decisive had the divisional bound come out odd. In Putnam 2022 A5 each domino covers two squares, so the uncovered count is invariant modulo 2. In Putnam 2018 B1 the totals $\sum a = 303$ and $\sum b = 15150$ must become even after deleting $\mathbf{v} = (a, b)$, forcing $a = 1$ and b even, and a pairing construction supplies sufficiency. Never present only one of those two halves.

Weighting generalises colouring: instead of two colours, attach a number to each cell and demand that moves conserve the total, usually modulo something. Putnam 2013 B6 is the finest instance on the paper, and the modulus is derived rather than guessed.

Example 10.12.5 (Putnam 2013 B6, the weight and the ghost cells)**.** Spaces $1, \dots, n$ with n odd. A move either places a stone in an empty s, or removes a stone from an occupied s and places stones in the nearest empty spaces on each side, where they exist. Extend the board by ghost cells labelled 0 and $n+1$ and work modulo $n+1$, so a missing flank contributes 0 and both rules obey one formula. Put $w(x) = \sum_{s \text{ occupied}} s \bmod (n+1)$. A placement at s changes w by $+s$; a type two move at s with flanking empty cells $i < s < j$ changes it by $i + j - s$.

Count: with m genuine empty cells there are m placements and $n - m$ type two moves, so n moves against n nonzero residues, and the correspondence is a bijection. List the empty cells with ghosts as $0 = e_0 < e_1 < \cdots < e_m < e_{m+1} = n+1$. Placements realise the increments $e_1, \dots, e_m$, the empty positions. For a type two move at s strictly between e_t and e_{t+1}, the increment $e_t + e_{t+1} - s$ reflects s in the midpoint of that gap, so as s runs over the occupied cells of the gap the increment runs over the same set: the type two moves realise exactly the occupied positions. So every increment in $\{1, \dots, n\}$ occurs exactly once.

Hence no move fixes w, so from a position with $w \equiv (n+1)/2$ every move leaves that class, and from any other position exactly one move enters it. The empty board has $w = 0$, so Alice's only candidate is a placement at the central space $s = (n+1)/2$, which is the answer; brute force for $n = 1, 3, 5$ confirms uniqueness.

The honest caveat: the no repetition clause makes this a game on histories, so "exactly one move enters the class" still needs that move to be legal, and the complete argument quarantines the repetition rule with a potential that never decreases, confining any residual cycles to a small end zone. The weight removes everything else.

How to search for an invariant

1. Write the state as a mathematical object, not a picture: a multiset, a vector, a subset, a function. Putnam 2022 A5 is unattackable as a board of 2022 squares and immediate as a multiset of run lengths.
2. Write the moves as maps on that object, in coordinates.
3. Compute the change under a general move of every simple statistic: the sum, the sum of squares, residues modulo small numbers, an alternating sum, the number of parts, the maximum, the number of inversions. You want a change that is always zero or always of one sign.
4. If nothing is invariant, weight the coordinates by s, or 2^s, or ζ^s for a root of unity, and see which weighting makes the move rule collapse. If a boundary defect spoils conservation, add ghost cells so it vanishes; that fixes the modulus.
5. If both players can change the target quantity, stop looking, as in Putnam 2023 A6, and construct a strategy instead.
6. If the question is termination or eventual behaviour rather than impossibility, switch to a monovariant and well ordering.

Putnam 2024 B2 is step six in pure form: the reflection defining partner quadrilaterals fixes the circumcentre and permutes the side lengths, so centre, radii and side multiset are invariant and, with finitely many angular sign choices, fix the shape up to rotation. Finitely many classes and infinitely many terms force a repeat, so the answer is no. Putnam 2023 B1 is the complete invariant case: the hole set is always a staircase, and every staircase is reached by descending on the inversion number backwards, giving the count $\binom{m+n-2}{m-1}$.

- **WHEN** a problem says "repeat the following move" and asks whether a target is reachable **DO** hunt for an invariant by weighting the coordinates and testing sums, sums of squares and residues under a general move **BECAUSE** an invariant separating start from target settles impossibility outright, and no amount of case work substitutes for it (Putnam 2018 B1, 2023 B1, 2024 B2)
- **WHEN** a problem says the process "must eventually stop", or says "for all but finitely many" **DO** find an integer valued monovariant, bounded in the direction it moves, and cite well ordering **BECAUSE** a strictly monotone integer quantity on a bounded range cannot run forever, and this is the only proof shape for that command term (Putnam 2008 A3, 2020 B2, 2025 A1)
- **WHEN** a conservation law almost holds but fails at the boundary **DO** add ghost cells carrying the neutral value and change the modulus so the defect becomes zero **BECAUSE** one formula then covers interior and boundary and the modulus is derived, not guessed (Putnam 2013 B6)
- **WHEN** both players can move the quantity you hoped to freeze **DO** abandon the invariant search at once and construct an explicit strategy **BECAUSE** an invariant argument needs that no legal move changes the target quantity, so if any does the method cannot decide the game (Putnam 2023 A6)

10.13 Combinatorial games

Definition 10.13.1. A finite impartial game under the normal play convention is one in which both players have the same moves available from a position and the player unable to move loses. A position is a $\mathcal{P}$-position if the player about to move loses under optimal play, and an $\mathcal{N}$-position otherwise.

Theorem 10.13.2 (Characterisation of $\mathcal{P}$-positions)**.** *Let $\mathcal{S}$ be a set of positions such that every terminal position lies in $\mathcal{S}$, every move from a position of $\mathcal{S}$ leaves $\mathcal{S}$, and from every position outside $\mathcal{S}$ some move enters $\mathcal{S}$. Then $\mathcal{S}$ is exactly the set of $\mathcal{P}$-positions.*

Proof. Induct on the length of the longest play, which is finite. If $x \in \mathcal{S}$ every move leads outside $\mathcal{S}$, from which by induction the mover wins, so whoever moved from x loses; a terminal x is in $\mathcal{S}$ and its mover has already lost. If $x \notin \mathcal{S}$ some move enters $\mathcal{S}$, handing the opponent a losing position. □

Those three conditions are the fixed point equations of the backward recursion, so guessing the set and verifying them solves every level at once. That is why you guess first.

Example 10.13.3 (Putnam 2006 A2, worked)**.** From a heap of n stones a player removes one less than a prime, and whoever takes the last stone wins; show Bob wins for infinitely many n. Let B be the set of $\mathcal{P}$-positions, which contains 0, and suppose $B = \{b_1, \dots, b_r\}$ were finite. Then every $n \notin B$ admits b_i with $n - b_i + 1$ prime. Falsify that: choose distinct primes $q_1, \dots, q_r$ and pick, by the Chinese remainder theorem, $n \equiv b_i - 1 \pmod{q_i}$ for every i with n larger than every b_i and every $q_i + b_i$. Each $n - b_i + 1$ is then a multiple of q_i exceeding q_i, hence composite, so n has no move into B and $n \notin B$: contradiction. The proof never learns one element of B.

Theorem 10.13.4 (Sprague and Grundy)**.** *Every position of a finite impartial game under normal play has a Grundy value $G(P) = \operatorname{mex}\{G(P') : P \to P'\}$, where* mex *of a set of nonnegative integers is the least nonnegative integer not in it. Then P is a $\mathcal{P}$-position if and only if $G(P) = 0$, and for a disjunctive sum of independent components $G(P \oplus Q) = G(P) \oplus G(Q)$, where $\oplus$ is binary addition without carrying. In particular Nim is a loss for the mover exactly when the binary digit sum of the pile sizes is zero.*

The proof of the sum rule is the two part argument you should be able to produce cold: from a nonzero nim sum you can move to zero by adjusting the pile whose entry carries the leading bit of the sum, and from a zero nim sum every move makes it nonzero.

Now the negative knowledge, worth more on this paper than the theorem. Grundy theory is a statement about disjunctive sums of independent impartial components under normal play and has no content outside that hypothesis. Putnam 2013 B6 is not such a sum. Putnam 2022 A5 is a scoring game and also resembles the octal game Dawson's Kayles, whose Grundy sequence is eventually periodic with period 34: recognise the resemblance so that you reject it deliberately rather than wonder. Putnam 2020 B2 resembles coin sliding on a strip, which is staircase Nim on the gaps when jumping is forbidden, but jumping is allowed there and the gap structure is destroyed. In all three cases, look for a pairing.

- **WHEN** an impartial normal play game splits into independent components **DO** compute Grundy values by mex and take their nim sum **BECAUSE** the value of a disjunctive sum is the exclusive or of the component values, and zero means the mover loses
- **WHEN** the game does not split into independent components, or the payoff is a score **DO** do not compute Grundy values; look for a pairing, an invariant, or a pair of value functions **BECAUSE** Sprague and Grundy is a theorem about disjunctive sums under normal play (Putnam 2013 B6, 2020 B2, 2022 A5)

Pairing, mirroring and symmetry

This is the highest frequency game technique on the Putnam, for a structural reason: a pairing strategy is a purely local rule, verified by induction on rounds with no search at all.

Theorem 10.13.5 (Pairing)**.** *Suppose the arena carries a fixed point free involution σ preserving legality, that the initial position is σ-stable, and that whenever the first player makes a move g the image move $\sigma(g)$ is legal. Then the second player wins by always replying $\sigma(g)$.*

Proof. Induct on rounds with the invariant that after each reply the used set is a union of complete σ-pairs. It holds initially, and if it holds after round k then the first player's move disturbs exactly one pair, whose partner is still available, and the reply restores it. So the second player always has a legal reply and the first player runs out first. □

The load bearing hypothesis is legality of the image move, not existence of the involution. That is where every incomplete write up fails and it is what a strict grader probes.

Putnam 2008 A2: Alan and Barbara alternately fill an empty 2008×2008 array, Alan wanting a nonzero determinant. Pair the cells $(i, 2j-1)$ with $(i, 2j)$ and let Barbara copy Alan's number into the partner cell, which the invariant keeps vacant. At the end columns 1 and 2 are identical, so Barbara wins.

Putnam 2014 B5: players alternately choose distinct elements of $GL_n(\mathbb{F}_p)$, each commuting with all previous choices. For odd p take $\sigma(g) = -g$: since $-I$ is central, $-g$ commutes with everything g does and $-g \neq g$, so the second player wins. Second reading of the same fact: the game ends when the chosen set is a maximal abelian subgroup A, so the length is $|A|$, and with a forced length the first player wins exactly when it is odd; for odd p every such A contains $-I$ and has even order. The case $p = 2$ is where the parity is actually contested, and it needs separate work.

Example 10.13.6 (Putnam 2020 B2, worked)**.** k pegs occupy the leftmost of n holes, a move slides one peg rightwards into a vacant hole, and whoever cannot move loses. Bob wins exactly when n and k are both even.

Suppose both are even and pair the holes $(1,2), (3,4), \ldots, (n-1, n)$, calling a position balanced if each pair is jointly occupied or jointly empty; the start and the terminal position are balanced since k and $n-k$ are even. Alice moves a peg from a to $b > a$, and these lie in different pairs, because within a pair balance means both holes occupied or both empty while b was vacant and a was not. Every hole of b's pair exceeds every hole of a's pair, so Bob's reply, moving the partner of a to the partner of b, is rightwards, legal, and restores balance. Termination must be stated: $\Phi = \sum_p p$ strictly increases at each move and is bounded above, so Alice runs out first.

Otherwise Alice converts to an even-even position with Bob to move, using the fact that in a one directional game a vacant leftmost hole can never be filled and an occupied rightmost hole can never be vacated or entered, so both may be deleted to give an isomorphic game on a shorter board.

Example 10.13.7 (Putnam 2025 A3 and vertex geography)**.** A token moves along the edges of a finite graph from a start vertex s and may never revisit a vertex. If $G - s$ has a perfect matching M, the second player wins by always moving to the M-partner of the vertex just occupied, since before each of the first player's moves the visited set is $\{s\}$ together with complete M-pairs and any newly visited vertex has an unvisited partner; if instead G has a perfect matching covering s, the first player wins by the same rule started a move earlier. The theorem of Fraenkel, Scheinerman and Ullman is that the first player wins undirected vertex geography if and only if every maximum matching covers s.

In Putnam 2025 A3 the arena is $\{0,1,2\}^n$ with edges joining strings differing by one in a single digit. Deleting the all zero string leaves a perfect matching: pair each nonzero string with the one got by swapping 1 and 2 in its first nonzero digit, an adjacent vertex and a fixed point free involution. So Bob wins for every n.

Putnam 2023 A6 combines pairing with blocking. Bob names a goal for the parity of $\#\{k : k \text{ was chosen on turn } k\}$, and everything follows from one parity alignment: a fixed point at k can be created only on turn k, and Alice moves on odd turns while Bob moves on even ones, so Alice can only fix odd numbers and Bob only even ones. For even n Bob names "even" and uses the matching $\{2k-1, 2k\}$, since a pair contributes 0 or 2 fixed points. For odd n he names "odd" and blocks: on his turn s he takes $s+1$, the unique number that could score for Alice next turn and which cannot score for him, after which nothing scores at all. Two exceptional branches are closed by reduction to a smaller copy. So Bob wins for every n, with goal the parity of n.

Strategy stealing needs its own sentence because it is so often misapplied. Where an extra move can never hurt you, if the second player had a winning strategy the first could make an arbitrary move and steal it, so the first player wins or draws. It proves existence and never exhibits a strategy, and its hypothesis fails in no revisit games and in scoring games.

- **WHEN** the arena decomposes into pairs and the total number of moves is even **DO** give the second player the rule "reply inside the pair the opponent just disturbed" and prove the invariant by induction on rounds **BECAUSE** the invariant guarantees a legal reply at every turn, so the opponent runs out first, and it forces any statistic that is constant on pairs (Putnam 2008 A2, 2014 B5, 2020 B2, 2023 A6, 2025 A3)

- **WHEN** the rules and the start are symmetric under a reflection with a unique fixed cell **DO** take the fixed cell, then mirror every opponent move **BECAUSE** a mirrored reply is legal whenever the position is symmetric, which is the ten second guess that the centre is the answer in Putnam 2013 B6

- **WHEN** the game forbids revisiting states, or consumes a resource, so the state is the history **DO** look for a matching strategy, not Grundy values and not strategy stealing **BECAUSE** only local rules can describe a history-state strategy, and strategy stealing needs "an extra move never hurts", which a no revisit rule destroys (Putnam 2025 A3)

- **WHEN** every completed play has the same length **DO** ignore strategy and compute the parity of that length **BECAUSE** under normal play the first player wins if and only if the number of moves is odd, so the game reduces to one arithmetic question (Putnam 2014 B5)

Scoring games

Definition 10.13.8. In a scoring game the payoff is a number, one player maximising and one minimising. With $A(P)$ and $B(P)$ the values with the maximiser and the minimiser to move, $A(P) = \max_\mu B(P^\mu)$ and $B(P) = \min_\mu A(P^\mu)$, with $A = B =$ score at terminal positions; the recursion is well founded whenever every move strictly decreases a nonnegative integer statistic. A component worth $m + \tau$ to whoever moves in it first and $m - \tau$ to the other is a switch of mean m and temperature τ, and a sum of switches of equal temperature has value $\sum_i m_i \pm \tau(t \bmod 2)$ with t the number of components of positive temperature.

Example 10.13.9 (Putnam 2022 A5)**.** Alice and Bob alternately place dominoes on a row of 2022 squares, Alice maximising and Bob minimising the uncovered count. The first move is abstraction: a position is the multiset of lengths of maximal uncovered runs, a move replaces an entry $n \geq 2$ by i and $n-2-i$, and the terminal score is the number of entries equal to 1. Without that, the board geometry hides that positions are multisets and moves are splits. The parity monovariant gives the free bound and the switch structure gives the shape, every position being a sum of switches of temperature 1. Running the recursion, a single run of length $N = 7q + 4 + j$ with $0 \leq j \leq 6$ is worth $q + 2$ for even j and $q + 1$ for odd j; since $2022 = 7 \cdot 288 + 4 + 2$, the answer is 290.

10.14 The logic layer

Nothing here is content. All of it is discipline, and it is where the cheapest marks on the paper are won and lost.

Quantifiers and their negations

The negation of "for all x there exists y with $Q(x, y)$" is "there exists x such that for all y, not $Q(x, y)$". Flip every quantifier in order and negate the kernel, on the page, in symbols, before arguing: the order of quantifiers changes the mathematics completely and fluency does not protect you at speed.

The asymmetry to internalise: membership of an existentially defined set supplies one adversarially chosen witness, while non membership supplies information at every point of the domain. Putnam 2022 B3 runs entirely on it. With D the set of distances realised between two points of the same colour, "$d \notin D$" translates into "$\chi(t) \neq \chi(t + d)$ for every t", a functional identity, whereas "$d \in D$" gives one pair of points and nothing more. That translation makes the problem computable, and the conclusion is that the recolouring map satisfies $\Phi^2 = \text{constant}$, so the all red colouring is the unique fixed point and is reached by time two.

Related: in an existence question where you choose the objects, choose those for which the event is decided by a single forced element, which in an ordered structure means near the minimum or the maximum. That is how Putnam 2024 A3 is answered yes, by the cells $(2, 1)$ and $(1, 2)$.

- **WHEN** a hypothesis or a goal has nested quantifiers **DO** write the statement and its negation in symbols, in order, before doing anything else **BECAUSE** negation flips every quantifier and negates the kernel, and the asymmetry between one witness and a universal condition decides which side carries the information (Putnam 2022 B3)
- **WHEN** a set is defined by "there exists t with $Q(t, d)$" and you need non membership **DO** convert it to "for every t, not $Q(t, d)$" and treat it as a functional identity **BECAUSE** non membership of an existentially defined set is universally quantified, which is strictly more usable than a single witness

Disjunctions, contradiction and contrapositive

To prove "P or Q" you prove that the failure of one implies the other, and the choice of which to negate is not free: negate the alternative whose negation produces a usable quantitative object. Putnam 2019 A6 asks you to prove that either $g'(x) \to 0$ or $\limsup x^r |g''(x)| = \infty$. Negating the second gives $x^r |g''(x)| \leq C$ near 0, an inequality with a constant and hence a working hypothesis; negating the first gives only a sequence along which g' stays away from zero. Negate the second, and the architecture is fixed in the first minute. Putnam 2010 A5 uses the same shape on the hypothesis: "either $\mathbf{a} \times \mathbf{b} = \mathbf{a} * \mathbf{b}$ or $\mathbf{a} \times \mathbf{b} = 0$" is used by negating the trivial branch, so a nonzero cross product is forced to equal the group product and to lie in G.

Contradiction and contrapositive are not the same move. The contrapositive of "P implies Q" is a direct proof of an equivalent statement, and it is honest when $\neg Q$ is more concrete than P. Genuine contradiction is honest when assuming falsity manufactures a new finite object. Putnam 2006 A2 is the second kind: finiteness of the losing set manufactures the list the Chinese remainder construction needs. Putnam 2025 B1 is the first: "three points of one colour force their circumcentre to be that colour" is unusable forwards, since you cannot manufacture monochromatic triples, but its contrapositive is a countable budget, every circle centred at a red point carrying at most two green points and vice versa.

- **WHEN** the goal is "P or Q" **DO** assume the negation of whichever alternative yields a quantitative hypothesis and prove the other **BECAUSE** negating an unboundedness statement

gives an inequality with a constant, while negating a limit gives only a sequence (Putnam 2019 A6, 2010 A5)

- **WHEN** a closure hypothesis cannot be used forwards **DO** take its contrapositive and read it as a cap on how many objects of the other kind sit on each fibre **BECAUSE** "several of colour c force output c" is equivalent to "an output of colour c admits at most threshold minus one inputs of the other colour", which is a counting budget (Putnam 2025 B1)

The architecture of "find all" and "for which"

An image determination question always splits into an obstruction and a construction: prove a necessary property of every value, then exhibit a preimage for every value with that property. Half the marks sit in each direction, so half an architecture earns half the marks. Putnam 2019 A1 and 2018 B1 are exactly this shape, and a claim "$P(N)$ if and only if $Q(N)$" is split into its two directions before any computation because they want structurally different tools, as in Putnam 2014 B1, whose answer is the integers having no digit zero.

Two asymmetries. Disproving a universally quantified statement needs one witness, proving it needs an argument for an arbitrary parameter, which is why in Putnam 2017 B1 the backward direction may fix $\lambda = 1$ while the forward direction must handle arbitrary λ and P. And "for which n does there exist" is neutral: the answer may be every n, and the interrogative form is not evidence of an obstruction. Putnam 2019 B6 is the trap, its answer being every $n \geq 1$, so test a construction first unless an obstruction is immediately visible. Finally, when a hypothesis guarantees an object without a formula, name it and use only its defining property: in Putnam 2012 A2 you name e and g and use only $d * e = a$ and $g * a = b$.

Closure and least fixed points

Definition 10.14.1. The set generated by a seed A under partial maps f_i is the least fixed point of $X \mapsto A \cup \bigcup_i f_i(X)$, equivalently the union of the iterates $S_0 = A$, $S_{m+1} = S_m \cup \bigcup_i f_i(S_m)$, equivalently the intersection of all sets containing A and closed under every f_i.

Those three descriptions are three proof techniques: the iterates prove membership, by a finite chain of rule applications; the intersection proves non membership, by one closed set omitting your element; the fixed point is what you verify.

Example 10.14.2 (Putnam 2017 A1)**.** S is smallest with $2 \in S$, with $n \in S$ whenever $n^2 \in S$, and with $(n+5)^2 \in S$ whenever $n \in S$. Upper bound: the integers that are neither 1 nor multiples of 5 contain 2 and form a closed set, since $5 \nmid n$ gives $5 \nmid (n+5)^2$ and $5 \nmid n^2$ gives $5 \nmid n$, while 1 is excluded because its only square root is itself and $(n+5)^2 \geq 36$. Lower bound: every other positive integer is reached by an explicit chain. The answer is 1 together with the multiples of 5. Note the shape: $n \in S$ gives $(n+5)^2 \in S$ and hence $n+5 \in S$, so S is an up-set within each residue class modulo 5 it meets, and the description reduces to one least element per class.

Example 10.14.3 (Putnam 2009 A4, worked)**.** S contains 0 and is closed under $x \mapsto x \pm 1$ and under $x \mapsto 1/(x(x-1))$ for $x \notin \{0,1\}$. Must $S = \mathbb{Q}$? No, and the certificate is one closed proper superset, which you design rather than guess. Write $x = p/q$ in lowest terms with $q > 0$. The shifts act as $(p,q) \mapsto (p \pm q, q)$, and since $\gcd(q^2, p(p-q)) = 1$,

$$\frac{1}{x(x-1)} = \frac{q^2}{p(p-q)} \quad \text{is already in lowest terms.}$$

So the third rule always produces a numerator $\pm q^2$: a square up to sign. That is the design input. Choose a prime ℓ for which -1 is a quadratic residue, the smallest being $\ell = 5$, so that the squares modulo 5 are $\{1,4\} = \{\pm 1\}$ and the condition "is a square" is blind to sign. Put

$$T = \{p/q \text{ in lowest terms} : \ 5 \nmid q, \text{ or } p \equiv \pm 1 \pmod 5\}.$$

Then $0 \in T$. The shifts change p by a multiple of q, so when $5 \mid q$ they fix p modulo 5 and T is closed under them. For the third rule, suppose 5 divides the new denominator $p(p-q)$; then $5 \mid p$ or $5 \mid p-q$, and in either case $5 \nmid q$, since $5 \mid q$ with either alternative would force $5 \mid p$ against coprimality. Hence $q^2 \equiv \pm 1 \pmod 5$ and the new numerator $\pm q^2$ is again ± 1, so T is closed with no hypothesis on x at all. Since $2/5 \notin T$ we get $S \subseteq T \subsetneq \mathbb{Q}$, and the answer is no.

Example 10.14.4 (Putnam 2025 B3)**.** If $n \in S$ then every positive divisor of $2025^n - 15^n$ is in S, and S is nonempty; must $S = \mathbb{N}$? Yes. First $1 \in S$. Now strong induction: given $n \geq 2$, produce $k \leq n-1$ with $n \mid 2025^k - 15^k = 15^k(135^k - 1)$. Factor $n = 3^a 5^b d$ with $\gcd(d, 15) = 1$, put $c = \max(a, b)$ and $k = \max(c, 1)\varphi(d)$. Then $3^a 5^b \mid 15^k$ since $k \geq c$, and $d \mid 135^k - 1$ by Euler since $\varphi(d) \mid k$ and $\gcd(135, d) = 1$; the factors are coprime, so $n \mid 2025^k - 15^k$, and $k \leq n-1$ follows from $\varphi(d) \leq d$ with $\max(c, 1) \leq 3^c - 1$ when $c \geq 1$, and $\varphi(d) \leq d - 1$ when $c = 0$.

- **WHEN** a set is defined as the smallest one containing a seed and closed under given rules **DO** prove an explicit candidate closed for the upper bound, and produce a rule chain to each element for the lower bound **BECAUSE** the generated set is at once the least fixed point, the union of the iterates and the intersection of all closed sets, and each description powers one direction (Putnam 2017 A1)
- **WHEN** you suspect a closure is not everything **DO** read off what shape the rules force on their output and impose exactly that shape as a congruence **BECAUSE** in Putnam 2009 A4 the third rule produces numerators $\pm q^2$, so a quadratic residue condition modulo a prime with -1 a residue is automatically preserved
- **WHEN** a set is divisor-closed under some F and you must show it is everything **DO** use strong induction, designing a witness $k < n$ with $n \mid F(k)$ from the prime factorisation of n **BECAUSE** closure hands you every divisor of $F(k)$ at once, so one designed witness per n finishes the induction (Putnam 2025 B3)

Cardinality, choice and transfinite arguments

The Putnam uses countability lightly and precisely: a countable union of countable sets is countable, $\mathbb{Q}$ is countable and $\mathbb{R}$ is not, and a countable family of constraints cannot exhaust an uncountable set, which is how you produce a point avoiding all of them. Putnam 2025 B1 finishes this way once the contrapositive budget has bounded how many points of one colour each point of the other can obstruct.

The axiom of choice appears only through Zorn's lemma: if every chain in a nonempty partially ordered set has an upper bound, the set has a maximal element. On the Putnam it surfaces at most as "take a maximal such object", and in a finite or countable situation the extremal principle or ordinary induction does the same job with no axiom. Do not invoke Zorn where well ordering of $\mathbb{N}$ suffices; it is not wrong, it is noise, and noise costs marks. Transfinite induction proper has not been needed by any problem behind this chapter.

One last tool, logic wearing an analysis coat: a real valued quantity varying continuously along a connected path and taking both signs vanishes somewhere on it. Putnam 2006 B3 uses exactly this to show that only the partitions through a given point can have two preimages under a restriction map, a statement otherwise asserted by inspection and then not believed.

- **WHEN** you must produce an object avoiding countably many bad conditions **DO** count: the bad set is a countable union of small sets and cannot exhaust an uncountable one **BECAUSE** cardinality gives existence with no construction, and it is the standard finish for colouring problems in the plane (Putnam 2025 B1)
- **WHEN** a solution wants to assert a sign change by inspection **DO** name the continuous parameter, name the connected path, and cite the intermediate value theorem **BECAUSE** a

sign change on a connected path is a theorem, not an observation, and writing it as one closes the gap a grader will probe (Putnam 2006 B3)

10.15 The order in which to install this

Priority five, drilled until reflexive: linearity of expectation with indicators, which carries Putnam 2006 A4, 2016 B4, 2020 A4, 2021 B1, 2022 A4 and 2023 B3; the tail sum formula; conditioning and first step analysis; exchangeability; $\mathcal{P}$-positions and $\mathcal{N}$-positions; pairing and mirroring, which carry Putnam 2008 A2, 2014 B5, 2020 B2, 2023 A6 and 2025 A3; parity as the first invariant checked; quantifier negation; obstruction plus construction as the architecture of every "find all".

Priority four: affine drift and its closed form; order statistics and the quantile transform; probability and moment generating functions; Chernoff with the tilting criterion; the uniform to exponential to Poisson chain; geometric probability by volume ratios; monovariants and well ordering; weighting with ghost cells; closure as a least fixed point; reduction to a smaller copy.

Priority three, an afternoon each: covariance inequalities; characteristic functions and Fourier inversion; the truncated moment problem; Wald and optional stopping; scoring games and switches; random walks on groups and the spectral gap.

Rare, learn last, and know why. Sylvester's four point problem and Efron's identity appeared once, in Putnam 2006 A6, as cross checks rather than as method. Octal games appeared once, as negative knowledge in Putnam 2022 A5. Undirected vertex geography appeared once, in Putnam 2025 A3, and the pairing lemma beneath it is the part you need. Zorn's lemma and transfinite induction have not been needed at all.

Two closing rules that belong to no section. Name the state before you compute: the classification of games in Putnam 2025 A3 and the abstraction to run multisets in Putnam 2022 A5 are nothing but that, and both problems are impenetrable until it is done. And whenever you meet an average, find the object whose mean is wanted and write it as a sum. Those two habits account for more marks here than all the named theorems together.