PAPER VI

# Transfer Semigroups and Memory

## Probabilistic refinement, fractional operators, and exact hidden-state elimination

Forty-two worked and reference pages follow complete transfer, probability aggregation, Bayesian information, spectral clocks, and the exact memory of hidden states.

Komorebi P.

Visual guide • 9 October 2026

**How to use this volume**

The guide follows small models whose matrices, probability laws, and transfer formulas can be calculated explicitly. The traditional volume supplies the general arguments. The Lean companion records a separate formal contract.

## Routes through the guide

The labels V1 through V42 identify worked pages independently of PDF pagination. The references on V–1 through V–28 use the theorem numbering of Traditional Part I; V–29 onward names the corresponding strengthened result in Traditional Part II. These references distinguish preserved manuscript branches rather than merging their result counts.

| Route | Mathematical question | Visual pages |
|---|---|---|
| Complete transfer | Which coordinates and modes reconstruct the same evolution? | 1–6, 29 |
| Probability and evidence | What survives averaging, aggregation, or an observation? | 7–14, 26, 32, 34–36 |
| Spectral evolution | How do rates, gaps, fractional clocks, and Gaussian modes interact? | 15–17, 25, 33, 41 |
| Green and path formulas | How do grid units and weighting conventions enter? | 18–20 |
| Eliminating hidden states | What memory returns, and how can it be compressed? | 21–24, 37–40 |
| Passing to limits | What additional conditions give convergence and closure? | 27, 30–31 |
| Evidence | Which calculations and formal records support which claims? | 28, 42 |

### Three distinctions to keep visible

A coordinate decomposition retains the whole vector if all components are kept. An aggregation map usually loses information. Elimination recovers exact retained dynamics only after adding a memory term and any hidden-state initial forcing.

An exact example checks its stated finite model. A theorem provides an argument under quantified assumptions. A Lean record certifies its declared formal statement under the recorded environment. None of these automatically certifies the other two.

### Preservation and revision

The original 26 TikZ plates and the mathematical text of the first 28 visual pages are retained. The monograph updates presentation, reading references, edition wording, and the distinction between current and inherited evidence. Fourteen editorial vector plates support the extended examples. All forty plates were regenerated from their supplied Python sources for this edition. All recovered source snapshots are included separately in the final source package.

The inherited formal baseline records 33 scoped results, 341 supporting theorem declarations, and 47 proof modules. The strengthened traditional branch is a separate written-proof collection. Its broader content is not added to the formal denominator.

Traditional Part III retains complementary results with an R prefix: reversibility and Dirichlet forms (R2), prior disintegrations and filtering (R3), potential perturbation and coarse Feynman–Kac transfer (R4), and finite realizations (R5). Related models appear on V–5–8, V–11, V–13, V–20, V–26, and V–39–40. The bounded-operator reciprocal result migrated from Paper IV is R7.2; V–25 and V–41 explain related finite matrix calculus. These models do not replace the supplementary hypotheses.

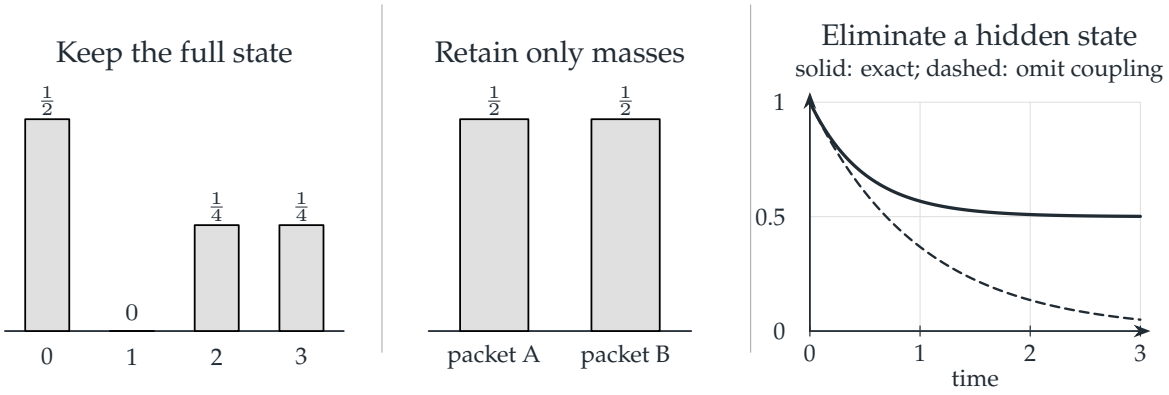
**Read this as:** Choose a question, work through its small model, then check the hypotheses in the corresponding traditional proof and formal contract.

V1 / PAPER VI / A WORKED VISUAL ATLAS

# Three ways to reduce a model

Traditional pp. 2, 3, 5, 13, 28 / Part I: Sections 1–3, 8, and 13. Read the three operations separately.

This guide follows small objects that can be computed completely. A four-state chain connects packet transfer, Markov aggregation, and Bayesian conditioning. A two-state coupled evolution exposes memory. A three-node Dirichlet problem makes the Green-operator normalization visible. These models are related by the paper's language, not by pretending that their state spaces are identical.




Left and center use one probability vector and its two block masses. The right panel is a separate retained/hidden model, developed on V22. Every curve is drawn from an explicit formula.

### Coordinates retain; aggregation forgets; elimination remembers

Keeping every component $P_a x$ is an exact change of representation when $\sum_a P_a = I$. Reconstruct by summing the components. In contrast, replacing $(p_0, p_1, p_2, p_3)$ by $(p_0 + p_1, p_2 + p_3)$ forgets how mass was distributed inside either pair. That information can matter for later observations.

Eliminating a hidden state is different again. Its coordinates disappear from the displayed state, but the exact retained equation acquires a history integral. Simply deleting the coupling produces the dashed curve above, not the original evolution.

| Question | Worked route | Pages |
|---|---|---|
| When is coarse evolution exact? | Aggregate columns, then test intertwining. | V4–V6 |
| When is a coarse posterior exact? | Compare two hidden shapes with the same masses. | V9–V13 |
| What changes under a fractional power? | Keep modes, change spectral rates. | V15–V17 |
| What returns after elimination? | Compute the self-energy and memory kernel. | V21–V25 |

> **Read this as:** The paper supplies exact identities and conditional bounds. Whether a particular coarse model satisfies their hypotheses is a separate calculation.

V2 / READ BEFORE MULTIPLYING / CONVENTIONS

# A dictionary with orientations

Traditional pp. 3, 7, 9, 11, 27 / Part I: Sections 2, 4, 6, 7, and 12. Column probabilities and backward observables.

An apparent disagreement between two transfer formulas often comes from acting on different objects. Throughout this guide, probabilities are columns and observables are functions written as columns. Their pairing is $f^{\mathsf{T}}p$. The transpose distinguishes the forward action on distributions from the backward action on observations.

| Object | Exact meaning | What to check |
|---|---|---|
| $p$ | A nonnegative column with $\mathbf{1}^{\mathsf{T}}p = 1$. | It is not a vector of function values. |
| $Q$ | Continuous Markov generator. | Off-diagonal entries are nonnegative and columns sum to zero. |
| $K$ | One discrete stochastic step. | $\mathbf{1}^{\mathsf{T}}K = \mathbf{1}^{\mathsf{T}}$ and $K_{ij} \geq 0$. |
| $K^{\mathsf{T}}f$ | Expected next observable given the current state. | The same dynamics, with the opposite action on the dual object. |
| $P_a$ | A coordinate packet projector. | All packets together reconstruct the full state. |
| $C$ | Sums fine probability masses into coarse blocks. | Usually rectangular and not invertible. |
| $E_\pi$ | Conditional expectation of observables. | A weighted average, self-adjoint for the $\pi$ inner product. |
| $\mathcal{B}_\lambda$ | $\operatorname{diag}(\lambda)p/(\lambda^{\mathsf{T}}p)$. | Nonnegative likelihood and positive total evidence. |
| $L$ | A nonnegative self-adjoint operator. | Its evolution is $e^{-tL}$, not $e^{tL}$. |
| $R(z)$ | $(zI - A)^{-1}$. | $z$ belongs to the resolvent set. |
| $\mathcal{K}(t)$ | The memory kernel after hidden-state elimination. | It is not the transition matrix $K$. |

### One convention test

For

$$K_2 = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}, \qquad p = e_0, \qquad f = (2,1)^{\mathsf{T}},$$

we obtain $K_2 p = (3/4, 1/4)^{\mathsf{T}}$ and

$$f^{\mathsf{T}}K_2 p = 7/4 = (K_2^{\mathsf{T}}f)_0.$$

With departure weights $w_i = e^{-V(i)}$, the backward Feynman–Kac operator is $\operatorname{diag}(w)K_2^{\mathsf{T}}$. Its forward dual is $K_2 \operatorname{diag}(w)$. Interchanging these factors changes when the weight is applied.

> **Read this as:** Always name the object being transported. The transpose and the position of the likelihood or potential are mathematical choices, not typographic details.

V3 / CONTINUOUS TRANSFER / THE ENTIRE SOLUTION

# A clock generated by one matrix

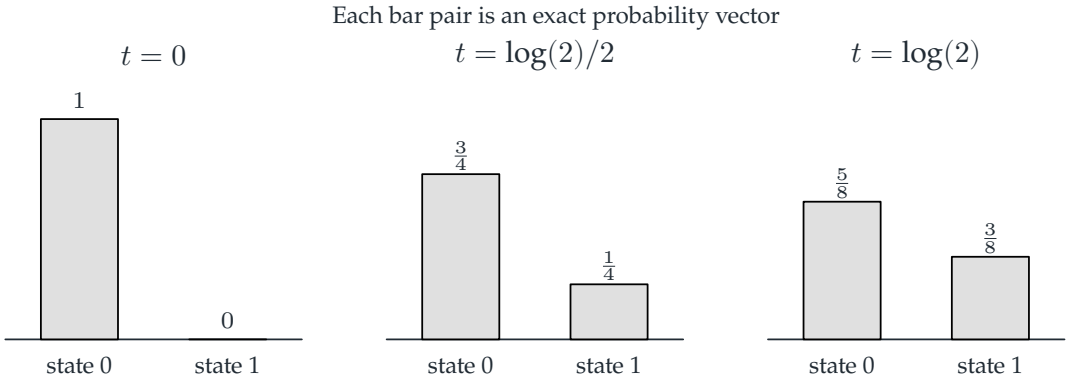
Traditional pp. 3, 7 / Part I: Theorems 1.2 and 4.1. Finite matrix exponentials and Markov positivity.

Take the generator

$$G = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

The total-mass direction $(1, 1)$ has eigenvalue zero. The contrast $(1, -1)$ has eigenvalue $-2$. Resolving those two directions gives the full transfer, not only its first-order approximation:

$$T(t) = e^{tG} = \frac{1}{2}\begin{pmatrix} 1 + e^{-2t} & 1 - e^{-2t} \\ 1 - e^{-2t} & 1 + e^{-2t} \end{pmatrix}.$$




A point seed at state zero approaches the uniform distribution. The displayed times were selected so that every bar label is rational. This is continuous-time evolution, not a simulated trajectory.

### Two checks explain the mechanism

The constant mode is unchanged, while each elapsed interval multiplies the contrast by $e^{-2t}$. Consequently two intervals multiply their contrast factors:

$$e^{-2s}e^{-2t} = e^{-2(s+t)}, \qquad T(s)T(t) = T(s+t).$$

For $t \geq 0$, the entries are nonnegative and every column sums to one. Differentiation gives $T'(t) = GT(t)$. These three facts separately check the initial condition, the evolution equation, and preservation of probability.

At $t = \log 2$, the retained first coordinate is $5/8$. It would instead be $1/2$ under the isolated scalar equation $\dot{x} = -x$. The discrepancy is the mass that can leave state zero and return. On V22, we remove the second coordinate correctly and recover exactly the same $5/8$ using a memory integral.

> **Read this as:** The semigroup lives on the complete state. One diagonal entry of a semigroup need not itself satisfy a scalar semigroup law.

V4 / FOUR-STATE MODEL / COORDINATES AND MODES

# One evolution, two packet choices

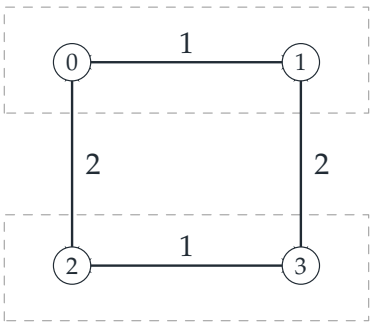
Traditional p. 4 / Part I: Theorem 2.1, Corollary 2.2, and the packet path expansion.

Our main finite model has generator

$$Q = \begin{pmatrix} -3 & 1 & 2 & 0 \\ 1 & -3 & 0 & 2 \\ 2 & 0 & -3 & 1 \\ 0 & 2 & 1 & -3 \end{pmatrix}.$$

The graph shows rates between actual states. The four normalized sign vectors on the right form an orthonormal basis $u_j$. Their projectors $P_j = u_j u_j^{\mathsf{T}}$ sum to the identity and satisfy $P_i P_j = 0$ for $i \neq j$.

Off-diagonal rates of $Q$




Top pair A; bottom pair B

Mode signs (each divided by 2)

| | 0 | 1 | 2 | 3 | rate |
|---|---|---|---|---|---|
| constant | 1 | 1 | 1 | 1 | 0 |
| within | 1 | −1 | 1 | −1 | −2 |
| between | 1 | 1 | −1 | −1 | −4 |
| mixed | 1 | −1 | −1 | 1 | −6 |

State coordinates have off-diagonal transfers. Spectral packets reduce the generator and evolve independently. The two dashed state groups are the coarse partition used next.

### The same evolution has a closed formula

Writing $q = e^{-2t}$ gives

$$e^{tQ} = P_0 + qP_1 + q^2 P_2 + q^3 P_3.$$

At $q = 1/2$, the point seed $e_0$ becomes $(15, 5, 9, 3)^{\mathsf{T}}/32$. The polynomial identity $T(q)T(r) = T(qr)$ checks composition exactly, without numerical matrix exponentiation.

Any other complete packet family still has exact block convolution. For the coordinate groups A and B,

$$P_b e^{(s+t)Q} P_a = \sum_{c \in \{\mathrm{A},\mathrm{B}\}} (P_b e^{sQ} P_c)(P_c e^{tQ} P_a).$$

The intermediate packet cannot generally be omitted. Only a reducing family eliminates all cross terms.

For later discrete calculations define $K = I + Q/8$. This matrix is stochastic, but *it is not* $e^{Q/8}$. It is the step kernel in the uniformization identity $e^{tQ} = e^{-8t} \sum_{n \geq 0} (8t)^n K^n / n!$.

> **Read this as:** Changing packet coordinates does not change the evolution. Replacing continuous time by one discrete step does.

V5 / EXACT LUMPABILITY / AGGREGATE COLUMNS

# When two states are truly equivalent

Traditional p. 8 / Part I: Theorem 5.1. The hypothesis is an intertwining identity.

Group states $0, 1$ into A and states $2, 3$ into B. The aggregation map and the candidate coarse kernel are

$$C = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \qquad \bar{K} = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}.$$

A fine state in A has probability $1/4$ of entering B in one step, regardless of which fine state in A was occupied. The analogous statement holds for B.

$8K$

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 5 | 1 | 2 | 0 |
| 1 | 1 | 5 | 0 | 2 |
| 2 | 2 | 0 | 5 | 1 |
| 3 | 0 | 2 | 1 | 5 |

$CK$: aggregate transition columns

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| A | $\frac{3}{4}$ | $\frac{3}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ |
| B | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{3}{4}$ | $\frac{3}{4}$ |

$$\bar{K} = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}$$

Fine trajectories differ. Coarse predictions agree.

Columns 0 and 1 agree; columns 2 and 3 agree.

The fine columns are visibly different. Their sums over each target packet agree within each source packet. That equality is the reason a two-state description works.

### A one-step calculation controls every discrete horizon

The displayed columns give $CK = \bar{K}C$. Multiplication then yields

$$CK^n = \bar{K}^n C \qquad (n \geq 0).$$

For example, $e_0$ and $e_1$ have the same coarse initial vector $(1, 0)^{\mathsf{T}}$. Their fine evolutions differ, but their coarse vectors agree at every step. No choice of the within-A conditional distribution is needed for the prediction.

The continuous generator also closes:

$$CQ = \bar{Q}C, \qquad \bar{Q} = \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}.$$

The Duhamel identity therefore gives $Ce^{tQ} = e^{t\bar{Q}}C$. This is a second consequence of the corresponding generator identity, not an inference that $K$ is a fixed-time sample of $Q$.

Aggregation is still not invertible. The theorem says the discarded details are irrelevant for *these predictions*. It does not say they are irrelevant to every later likelihood. The observation counterexample on V10 uses this exact same partition.

> **Read this as:** A partition is not automatically a valid Markov model. The aggregate transition columns must be constant within each source block.

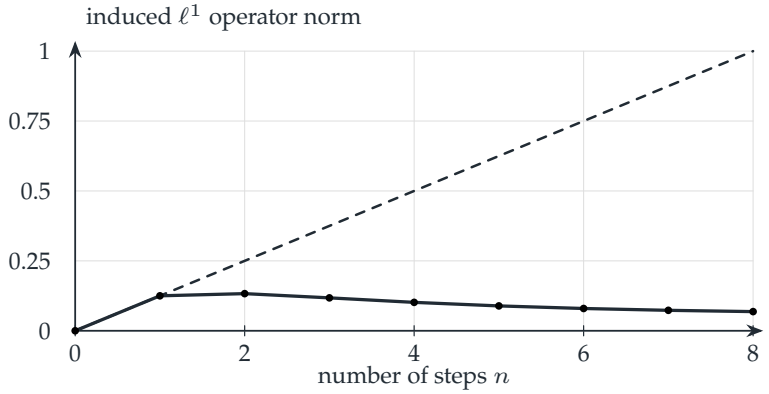
V6 / DEFECTS / A GUARANTEE WITHOUT A FIT

# A broken lump and its error budget

Traditional pp. 5, 6, 8 / Part I: Theorems 3.1, 3.3, and 5.2. Continuous and discrete defect identities.

Modify only column zero of $K$: move $1/16$ of its mass from target state zero to target state two. The perturbed matrix $K_\varepsilon$ remains stochastic, but its aggregate column now differs from column one. Relative to the unchanged coarse kernel,

$$\Delta = CK_\varepsilon - \bar{K}C = \begin{pmatrix} -1/16 & 0 & 0 & 0 \\ 1/16 & 0 & 0 & 0 \end{pmatrix}, \qquad \|\Delta\|_1 = 1/8.$$




$\|\Delta\|_1 = 1/8$

solid: exact defect
dashed: $n/8$

The bound is a guarantee, not a fitted trend.

The solid values are computed with rational matrix powers. The dashed certificate uses only stochasticity and the exact one-step defect. It need not be sharp.

### Telescoping tells us where every error originates

For $n \geq 0$,

$$CK_\varepsilon^n - \bar{K}^n C = \sum_{j=0}^{n-1} \bar{K}^{n-1-j} \Delta K_\varepsilon^j.$$

Each summand propagates one defect between an earlier fine evolution and a later coarse evolution. Column-stochastic matrices have induced $\ell^1$ norm one, so the norm is at most $n/8$. The empty sum at $n = 0$ is zero.

The continuous counterpart is equally explicit. With fine generator $-1$, coarse generator $-2$, and $C = 1$, the generator defect is $1$ and

$$e^{-t} - e^{-2t} = \int_0^t e^{-2(t-r)} e^{-r}\, dr.$$

Its resolvent version is $1/(z+1) - 1/(z+2) = 1/((z+1)(z+2))$ for $z > 0$.

> **Read this as:** A small defect does not disappear. The identity specifies how it accumulates, and the norm assumptions turn that identity into a finite-horizon certificate.

V7 / OBSERVABLES / WEIGHTED ORTHOGONALITY

# Averaging is not summing

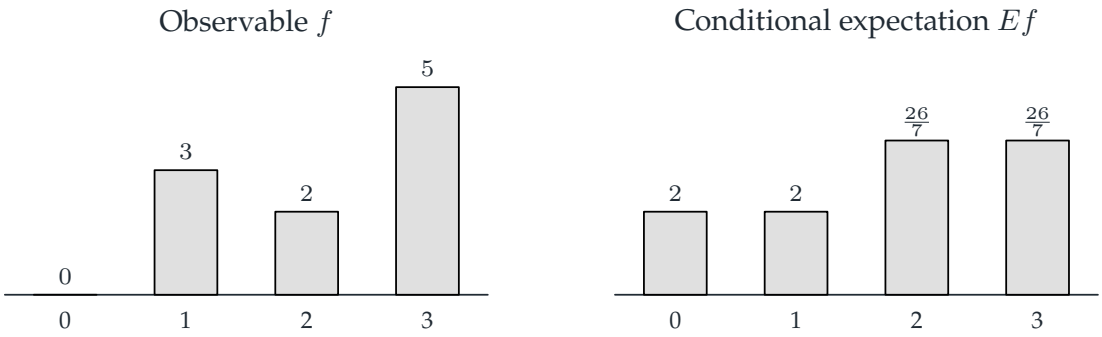
Traditional p. 9 / Part I: Theorem 6.1. Strictly positive weights are part of the setting.

For this observable example, choose the fixed probability weights $\pi = (1, 2, 3, 4)/10$. They are not being identified with the stationary law of the earlier symmetric chain. On the same two-block partition, conditional expectation is

$$E = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 3/7 & 4/7 \\ 0 & 0 & 3/7 & 4/7 \end{pmatrix}.$$

It replaces function values by their *weighted block average*. In particular, $f = (0, 3, 2, 5)^{\mathsf{T}}$ becomes $Ef = (2, 2, 26/7, 26/7)^{\mathsf{T}}$.

Weights $\pi = (1, 2, 3, 4)/10$; the block values are 2 and $26/7$.




The bars are function values, not state probabilities. Their total need not be one. The block masses $3/10$ and $7/10$ determine the denominators in the averaging formula.

### Which inner product makes this an orthogonal projection?

Put $W = \mathrm{diag}(\pi)$. Direct multiplication checks

$$E^2 = E, \qquad WE = E^{\mathsf{T}}W, \qquad E\mathbf{1} = \mathbf{1}.$$

The middle equation is weighted self-adjointness. Ordinary symmetry fails: $E_{01} = 2/3$ whereas $E_{10} = 1/3$. Testing $E = E^{\mathsf{T}}$ would therefore reject a correct conditional-expectation projector.

The range consists exactly of functions constant on each block. The error $f - Ef$ has weighted mean zero inside every block, so it is weighted-orthogonal to every block-constant function. This is the finite projection interpretation of conditional expectation.

In contrast, $C$ acts on probability columns by summing masses. The shapes of $C$ and $E$ differ, and their domains have different interpretations. The notation "coarsening" alone does not specify which operation is intended.

> **Read this as:** An exact weighted projector can be oblique in the unweighted Euclidean picture. Keep the measure and the inner product in the model.

V8 / MARTINGALE PACKETS / THREE EXACT PIECES

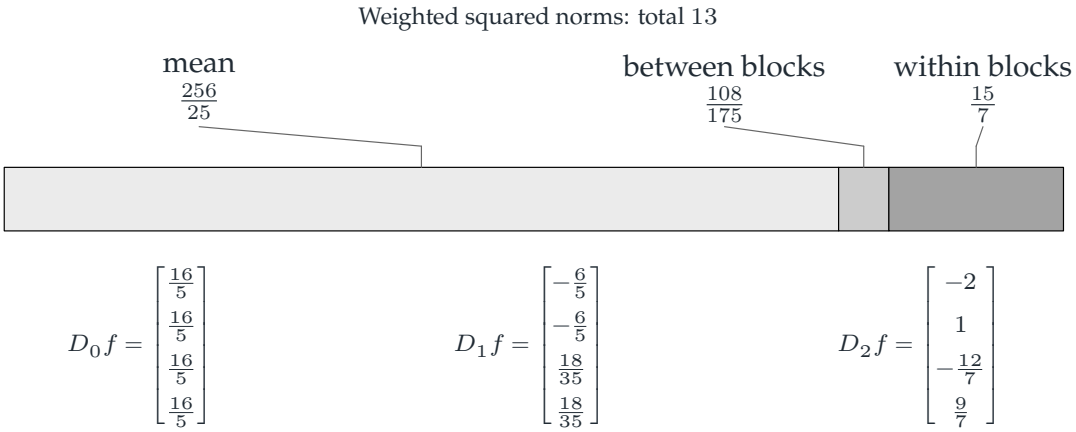
# Where the variance goes

Traditional p. 10 / Part I: Theorem 6.2. Nested weighted projections and their differences.

Keep the weights and observable from V7. Let $E_0 f$ be the constant global mean, let $E_1 = E$ average within A and B, and let $E_2 = I$ retain every value. Define

$$D_0 = E_0, \qquad D_1 = E_1 - E_0, \qquad D_2 = I - E_1.$$

Then $D_0 f$ is the mean, $D_1 f$ describes the contrast between block means, and $D_2 f$ is the fluctuation left inside each block.




The strip lengths are proportional to the three exact weighted energies. The between-block component is small but not zero. The vectors below specify the corresponding directions, not only their sizes.

### The energy identity follows from actual orthogonality

The tower identities give $E_0 E_1 = E_1 E_0 = E_0$. Together with weighted self-adjointness, these imply

$$D_i D_j = 0 \ (i \neq j), \qquad \sum_{j=0}^{2} D_j = I.$$

For the chosen $f$, the total weighted energy is $\|f\|_\pi^2 = 13$ and the mean is $16/5$. Consequently

$$\mathrm{Var}_\pi(f) = 13 - \left(\frac{16}{5}\right)^2 = \frac{69}{25} = \frac{108}{175} + \frac{15}{7}.$$

The first term is the energy of the between-block contrast, and the second is the weighted within-block variance. This is a decomposition of one observable in one fixed weighted space.

Changing the probability weights between levels would change the inner products and the averaging operators. The displayed orthogonal decomposition would no longer follow merely from nesting the sets. Likewise, choosing unrelated projections can preserve the telescope $\sum D_j = I$ while losing orthogonality and the energy sum.

> **Read this as:** The differences between nested conditional expectations are not just bookkeeping. They are orthogonal, separately measurable parts of the original signal.

V9 / BAYESIAN TRANSFER / AN EXACT POSTERIOR

# Evidence changes shape and mass

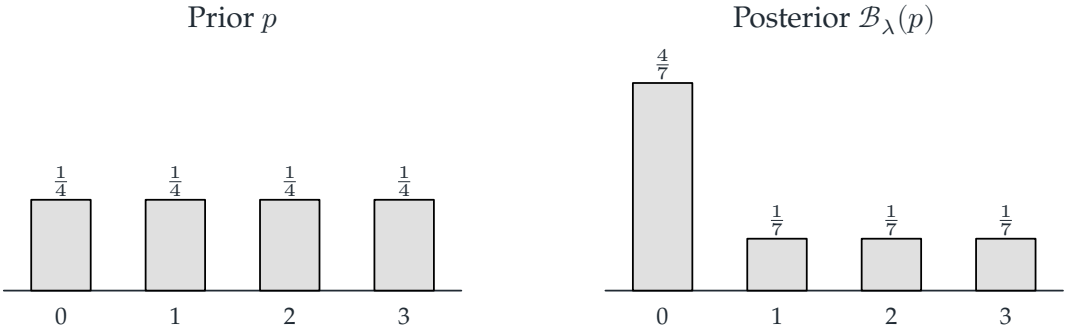
Traditional pp. 11, 12 / Part I: Theorems 7.1–7.3. Normalization, packet reconstruction, and positive evidence.

Start with $p = (1,1,1,1)^{\mathsf{T}}/4$ and likelihood weights $\lambda = (4,1,1,1)^{\mathsf{T}}$. The unnormalized vector is $(1,1/4,1/4,1/4)^{\mathsf{T}}$, its evidence is $Z = 7/4$, and

$$\mathcal{B}_\lambda(p) = \frac{(4,1,1,1)^{\mathsf{T}}}{7}.$$

Likelihood weights may be rescaled, so an evidence larger than one is not a probability claim. Multiplying all likelihoods by one common positive factor leaves this posterior unchanged.

$\lambda = (4,1,1,1)$; evidence $Z = 7/4$




The A packet becomes heavier and its internal shape changes. The B packet retains an equal internal split but loses total mass. These are different effects of the same likelihood.

### Reconstruct from packet mass times packet shape

The unnormalized packet evidences are $Z_{\mathrm{A}} = 5/4$ and $Z_{\mathrm{B}} = 1/2$. The resulting masses and conditional shapes are

$$m_{\mathrm{A}} = 5/7, \quad q_{\mathrm{A}} = (4/5, 1/5, 0, 0)^{\mathsf{T}}, \qquad m_{\mathrm{B}} = 2/7, \quad q_{\mathrm{B}} = (0, 0, 1/2, 1/2)^{\mathsf{T}}.$$

Thus $m_{\mathrm{A}} q_{\mathrm{A}} + m_{\mathrm{B}} q_{\mathrm{B}}$ is exactly the displayed posterior. Keeping only the masses would not reconstruct the fine state.

There is no need to divide by a zero packet evidence. For likelihood $(0,0,1,1)$, packet A has mass zero and contributes the zero vector. Only the occupied B packet has a conditional probability vector. If the *total* evidence is zero, however, the Bayesian update is not defined and the implementation rejects the input.

> **Read this as:** Bayesian reconstruction uses two kinds of data: how much posterior mass a packet carries and how that mass is distributed inside the packet.

V10 / SUFFICIENCY / A COMPUTABLE FAILURE WITNESS

# Two priors that fool the coarse model

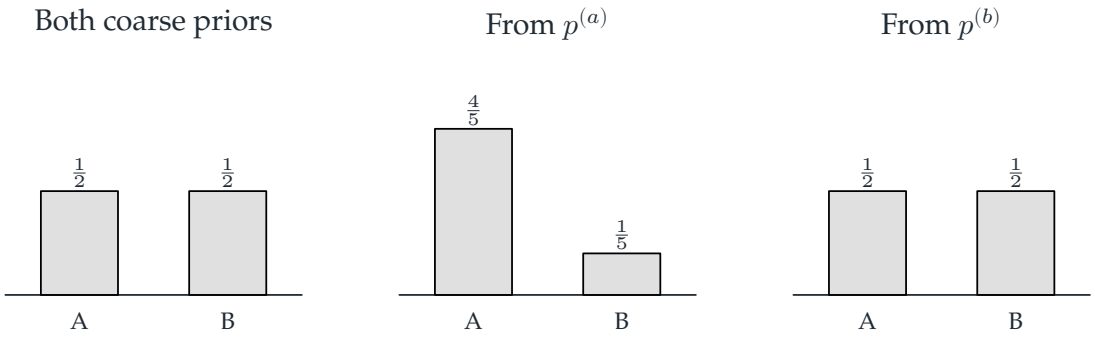
Traditional p. 14 / Part I: Theorem 8.2 and the one-packet exception.

Keep the same likelihood $\lambda = (4, 1, 1, 1)^{\mathsf{T}}$ and partition A$= \{0, 1\}$, B$= \{2, 3\}$. Compare

$$p^{(a)} = (1/2, 0, 1/2, 0)^{\mathsf{T}}, \qquad p^{(b)} = (0, 1/2, 1/2, 0)^{\mathsf{T}}.$$

Both have coarse prior $(1/2, 1/2)^{\mathsf{T}}$. A procedure given only that two-entry vector cannot distinguish them.

Hidden redistribution changes the packet-A posterior from $1/2$ to $4/5$.




After the identical fine likelihood is applied, the coarse posteriors are $(4/5, 1/5)$ and $(1/2, 1/2)$. A single deterministic update of the common coarse prior cannot produce both.

### Why this witnesses the necessity of blockwise constancy

The A evidence for $p^{(a)}$ is 2, whereas it is $1/2$ for $p^{(b)}$. Their B evidence is $1/2$ in both cases. The posterior odds therefore distinguish the hidden distribution inside A.

For strictly positive fine and coarse likelihoods, and at least two coarse blocks, the manuscript proves the all-prior criterion

$$C\mathcal{B}_{\lambda}(p) = \mathcal{B}_{\bar{\lambda}}(Cp) \quad \text{for every } p \quad \Longleftrightarrow \quad \lambda_i = c\bar{\lambda}_a \quad (i \in I_a)$$

for one global $c > 0$. The global scale is invisible after normalization. A different arbitrary scale for each packet is not invisible: it changes posterior odds.

If all states are merged into a single packet, every normalized posterior coarsens to $(1)$. That degenerate commutation holds for every likelihood and cannot identify any within-packet information. It is explicitly excluded from the nontrivial criterion.

> **Read this as:** A prior-dependent effective likelihood can always summarize the update. A fixed likelihood that works for every prior is a much stronger requirement.

V11 / BAYESIAN DEFECT / MORE THAN AN ERROR BAR

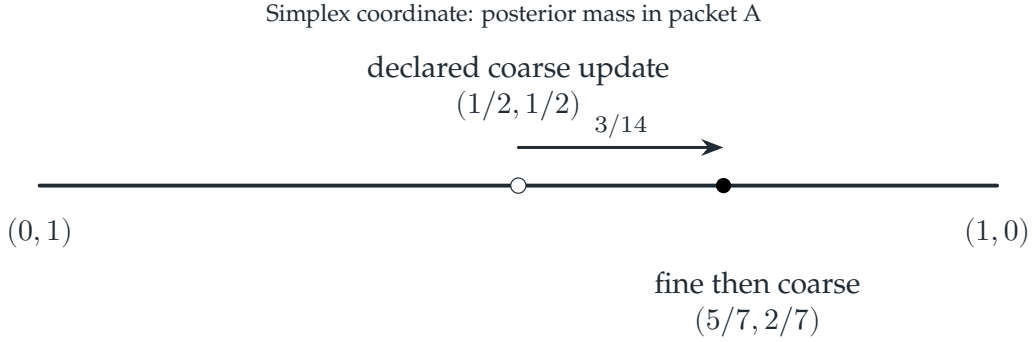
# The exact information-loss vector

Traditional pp. 13, 14, 15 / Part I: Theorems 8.1 and 8.4, and Corollary 8.6.

For a prior $p$, let $P = Cp$ and let $r_a(p)$ be the conditional average of the fine likelihood inside packet $a$. For the uniform prior and $\lambda = (4, 1, 1, 1)$,

$$P = (1/2, 1/2)^{\mathsf{T}}, \qquad r = (5/2, 1)^{\mathsf{T}}.$$

The exact coarse posterior is $\mathcal{B}_r(P) = (5/7, 2/7)^{\mathsf{T}}$. Suppose instead that the declared coarse likelihood is $\bar{\lambda} = (1, 1)$.




The posterior error lies along the mass-preserving direction of the two-state simplex. Its entries are opposite because both endpoints have total mass one.

### Compute the factorization, including the null direction

Here $\bar{b} = \mathcal{B}_{\bar{\lambda}}(P) = P$, $\delta = r - \bar{\lambda} = (3/2, 0)^{\mathsf{T}}$, and $Z_r = 7/4$. Hence

$$\frac{1}{Z_r}\left(\operatorname{diag}(P) - \bar{b}P^{\mathsf{T}}\right)\delta = \frac{4}{7}\begin{pmatrix} 1/4 & -1/4 \\ -1/4 & 1/4 \end{pmatrix}\begin{pmatrix} 3/2 \\ 0 \end{pmatrix} = \begin{pmatrix} 3/14 \\ -3/14 \end{pmatrix}.$$

This is the actual fine-to-coarse discrepancy, not a fitted approximation. The matrix kills $\bar{\lambda}$, recording the common-scale null direction. Its $\ell^1$ error is $3/7$.

For a family of likelihoods, states may be grouped by equality of their *whole likelihood profile*. The family $(3, 3, 1, 1)$ and $(1, 2, 1, 2)$ separates all four states even though either likelihood alone permits a coarser grouping. The common refinement is therefore the discrete partition. The theorem's one-packet exception still remains.

> **Read this as:** The defect factors through exactly the conditional evidence information missing from the declared coarse model. A new observation family can require a finer partition.

V12 / BAYESIAN FLOW / COMPOSITION WITHOUT LINEARITY

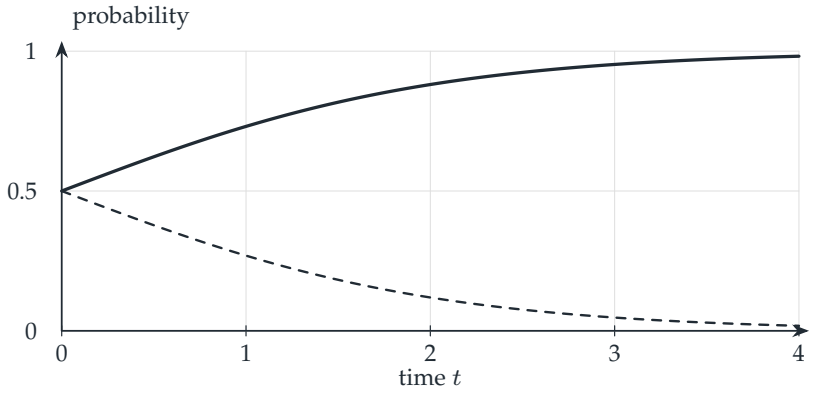
# A nonlinear semigroup of odds

Traditional pp. 11, 18 / Part I: Theorems 7.1 and 9.1. Likelihood composition and the replicator field.

Let the score vector be $g = (1, 0)$ and start at $(1/2, 1/2)$. The likelihood at time $t$ is $(e^t, 1)$, so the first posterior coordinate is

$$p_0(t) = \frac{e^t}{1 + e^t}, \qquad p_1(t) = \frac{1}{1 + e^t}.$$

The odds $p_0/p_1$ grow exponentially, while the probabilities remain inside the simplex. At $t = \log 2$ the posterior is exactly $(2/3, 1/3)$.




The solid and dashed curves sum to one. The curvature is caused by the shared evidence denominator, not by a nonlinear likelihood multiplication law.

### Two properties that must not be conflated

If $a > 0$ multiplies the odds, the first-coordinate map is

$$\phi_a(x) = \frac{ax}{1 + (a - 1)x}.$$

Direct algebra gives $\phi_a(\phi_b(x)) = \phi_{ab}(x)$. Thus $a = e^t$ gives a nonlinear semigroup. Nevertheless, $\phi_3(1/2) = 3/4$, while the mean of $\phi_3(0)$ and $\phi_3(1)$ is $1/2$. Superposition fails.

Differentiation of the normalized flow yields the centered score equation

$$\dot{p}_i = (g_i - \langle g, p \rangle) p_i.$$

For the displayed model, $\dot{p}_0 = p_0(1 - p_0)$. The negative average-score term is what keeps the total derivative zero. Omitting the derivative of the normalizer would not preserve mass.

Likelihood composition also preserves the evidence identity: the first evidence times the second conditional evidence equals the combined evidence. Normalization cancels in the posterior, but that factor still matters when comparing data models.

> **Read this as:** A semigroup is a composition law in time. It does not imply that the maps are linear or that they preserve mixtures.

V13 / HIDDEN MARKOV FILTERING / TIME ORDER

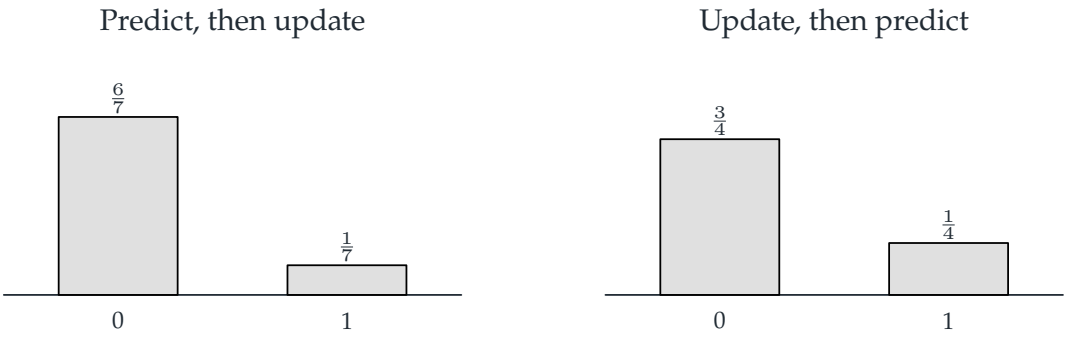
# Prediction and observation do not commute

Traditional p. 19 / Part I: Proposition 9.3. Linear prediction followed by a positive-evidence update.

Use the two-state kernel $K_2$ from V2, the initial distribution $e_0$, and likelihood $(1, 1/2)$. Predicting first gives $(3/4, 1/4)$. Its likelihood-weighted vector is $(3/4, 1/8)$, with evidence $7/8$, so the posterior is $(6/7, 1/7)$.

Updating first leaves $e_0$ unchanged. Predicting afterward gives $(3/4, 1/4)$ instead. The order of operations expresses which time the observation concerns.

Same prior $e_0$, transition $K_2$, and likelihood $(1, 1/2)$




Both routes produce valid probability vectors. Their difference is therefore not a normalization bug. It is a different conditioning problem.

### Keep an unnormalized path alongside the filter

For a sequence of observations, let $D_r = \operatorname{diag}(\lambda^{(r)})$. The forward recursion is

$$\hat{p}_r = K p_{r-1}, \qquad p_r = D_r \hat{p}_r / Z_r.$$

Its unnormalized companion is

$$v_r = D_r K \cdots D_1 K p_0, \qquad p_r = \frac{v_r}{\mathbf{1}^\mathsf{T} v_r}, \qquad \mathbf{1}^\mathsf{T} v_r = \prod_{j=1}^{r} Z_j.$$

The product of evidences records the normalization lost at each individual step. The four-state computation on V26 checks both paths independently, including their coarse versions.

Here $D_r K$ means an observation after prediction. In the Feynman–Kac example on V20, departure-state weighting has forward operator $K D_w$. Both conventions are legitimate once their meanings are fixed, but they cannot be substituted for one another.

> **Read this as:** Write the chronological sequence before multiplying matrices. A product can be normalized correctly and still answer the wrong time-ordered question.

V14 / BETA MODEL / POSTERIOR AND EVIDENCE

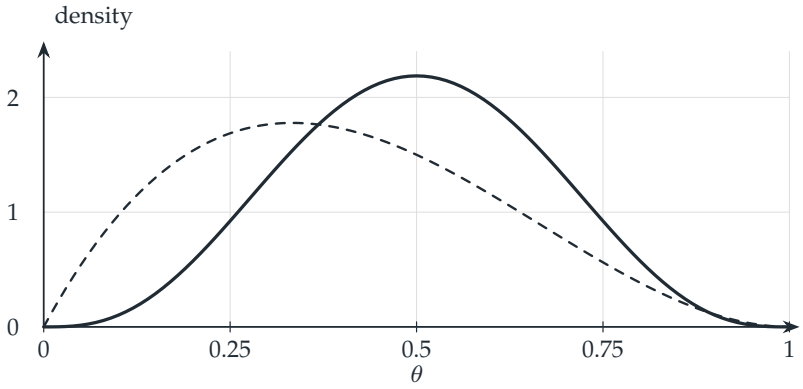
# Conjugacy tracks exponents

Traditional p. 19 / Part I: Example 9.2. A finite-parameter conjugate family, not a universal posterior model.

For a small exact example, take the Beta$(2, 3)$ density

$$\rho(\theta) = 12\theta(1-\theta)^2, \qquad 0 \le \theta \le 1.$$

Observe a specified ordered sequence with two successes and one failure. Its likelihood is $\theta^2(1-\theta)$. Multiplication adds exponents, so the posterior belongs to Beta$(4, 4)$ with density $140\theta^3(1-\theta)^3$.




Prior: Beta$(2, 3)$

Posterior: Beta$(4, 4)$

ordered data: $3/35$
count data: $9/35$

Same posterior, different evidence.

The dashed prior is asymmetric. The solid posterior is symmetric with mean $1/2$. Polynomial integration verifies the normalizers exactly; the curves are sampled drawings of those polynomials.

### The observation convention changes the evidence

The integral of the unnormalized posterior is

$$\int_0^1 12\theta^3(1-\theta)^3\,d\theta = \frac{3}{35}.$$

Dividing by $3/35$ gives the displayed normalized density. If the observation is instead the *count* of two successes in three trials, there are $\binom{3}{2} = 3$ compatible ordered sequences. The count likelihood is therefore $3\theta^2(1-\theta)$ and its evidence is $9/35$.

The extra factor does not depend on $\theta$, so it cancels from the posterior. The two experiments thus have the same posterior distribution but different data probabilities. Evidence is not determined by the final normalized density alone.

The exact check also integrates both densities to one and integrates $\theta$ against the posterior to obtain $1/2$. This demonstrates one conjugate family in elementary terms. It does not show that arbitrary likelihoods preserve that family or that all posterior information can be compressed into two exponents.

> **Read this as:** Conjugacy gives a closed parameter update when a specified likelihood respects a specified family. Preserve the evidence as well as the normalized posterior.

V15 / FRACTIONAL DIFFUSION / NEW COUPLINGS

# The modes stay; their clocks change

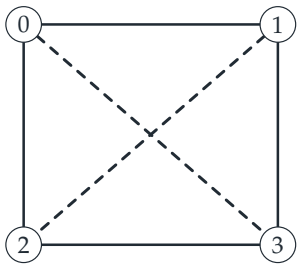
Traditional p. 20 / Part I: Theorem 10.1. A positive self-adjoint finite model and its spectral powers.

Return to the four-state generator on V4 and put $L = -Q$. It is real symmetric and nonnegative, with orthogonal projectors $P_0, P_1, P_2, P_3$ and eigenvalues $0, 2, 4, 6$. Its square root is

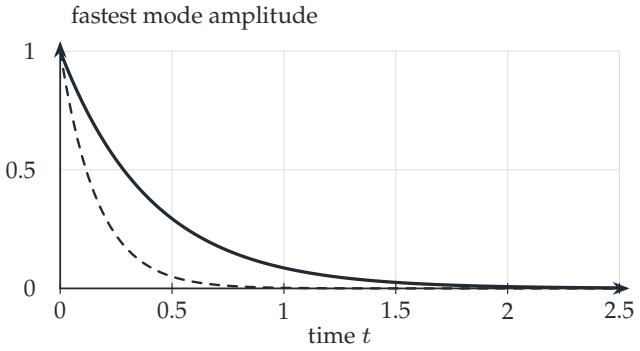
$$L^{1/2} = \sqrt{2}P_1 + 2P_2 + \sqrt{6}P_3, \qquad (L^{1/2})^2 = L.$$

Both heat evolutions retain these modes. Ordinary diffusion multiplies mode $j$ by $e^{-2jt}$; square-root diffusion multiplies it by $e^{-\sqrt{2j}\,t}$. The constant mode is unchanged in both.

Solid: $L$; dashed: new in $\sqrt{L}$




solid: $e^{-t\sqrt{6}}$; dashed: $e^{-6t}$




The highest mode decays at rates 6 and $\sqrt{6}$. The graph beside the curves shows the additional couplings in the fractional operator. Curves display the explicit formulas, not an empirical fit.

### Fractional does not mean entrywise

Although $L_{03} = 0$, its spectral square root satisfies

$$(L^{1/2})_{03} = \frac{\sqrt{6} - \sqrt{2} - 2}{4} < 0.$$

Thus $-L^{1/2}$ permits a direct transition between states 0 and 3, which were not neighbors for $Q$. Taking the square root of each matrix entry would neither produce this value nor square back to $L$.

The exact block aggregation remains compatible in this example:

$$CL^{1/2} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} C.$$

The coarse matrix is the positive square root of $-Q_c$. This is a concrete naturality identity, checked on the full matrices. It is not a claim that an arbitrary partition commutes with every fractional construction.

The computation uses a nonnegative self-adjoint operator and a specified spectral branch. General nonnormal matrices, unbounded domains, and positivity of other fractional evolutions need their own hypotheses.

> **Read this as:** Fractional powers alter modal decay rates. Reassembling those modes can introduce long-range spatial interactions even when the original generator was sparse.

V16 / FRACTIONAL DEFECTS / NEAR ZERO VERSUS AWAY FROM ZERO

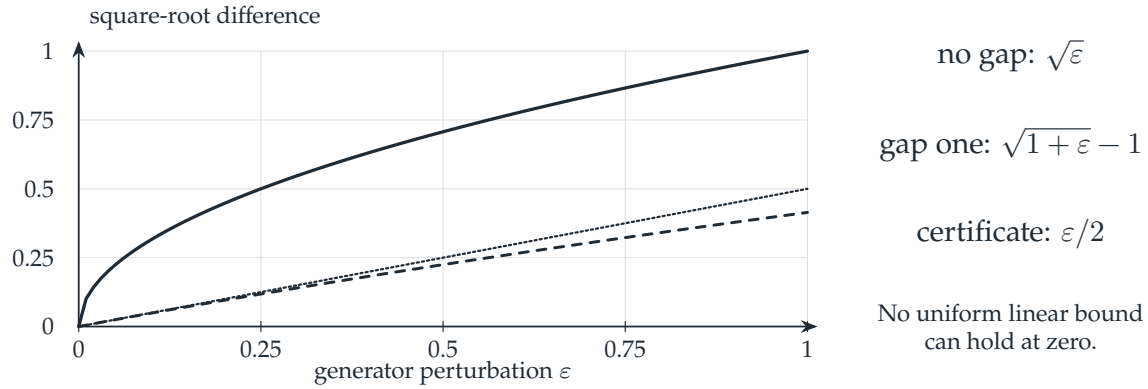
# A gap changes the error scale

Traditional pp. 22, 23 / Part I: Theorem 10.3 and Corollary 10.4. Separate generator error from fractional error.

The function $a \mapsto \sqrt{a}$ changes most rapidly near zero. A scalar model already shows why a spectral gap appears in fractional-defect estimates:

$$|\sqrt{\varepsilon} - \sqrt{0}| = \sqrt{\varepsilon}, \qquad \sqrt{1+\varepsilon} - 1 = \frac{\varepsilon}{\sqrt{1+\varepsilon}+1} \le \frac{\varepsilon}{2}.$$

The same input difference has a square-root scale at zero and a linear bound above a fixed positive lower bound.




The upper curve compares $0$ with $\varepsilon$. The lower curve compares $1$ with $1+\varepsilon$. The dotted line $\varepsilon/2$ bounds the latter. All plotted values come from the displayed scalar formulas.

### Read an operator bound with its hypotheses

For positive self-adjoint finite matrices $A, B \ge mI$ and $0 < \alpha < 1$, the manuscript's gapped estimate has the form

$$\|CA^{\alpha} - B^{\alpha}C\| \le \alpha m^{\alpha-1} \|CA - BC\|.$$

It concerns an intertwining defect, not just two eigenvalue lists. With $C = 1$, $A = 4$, $B = 9$, $m = 4$, and $\alpha = 1/2$, the generator defect has size $5$, the fractional defect has size $1$, and the bound is $5/4$.

The common zero mode of a Markov generator cannot simply be included in a positive-gap estimate. One must either work on an appropriate gapped subspace, handle the zero mode separately, or use the gapless estimate. In particular, a good estimate on each finite level need not have a constant uniform in the refinement level.

On V27, the sequences $1/n^2$ and $1/n$ turn this change in error scale into a concrete distinction between summable and nonsummable defects.

> **Read this as:** A fixed gap is quantitative information. It prevents the derivative of the fractional function from becoming singular along the comparison.

V17 / SUBORDINATION / A DISTRIBUTION OF ORDINARY TIMES

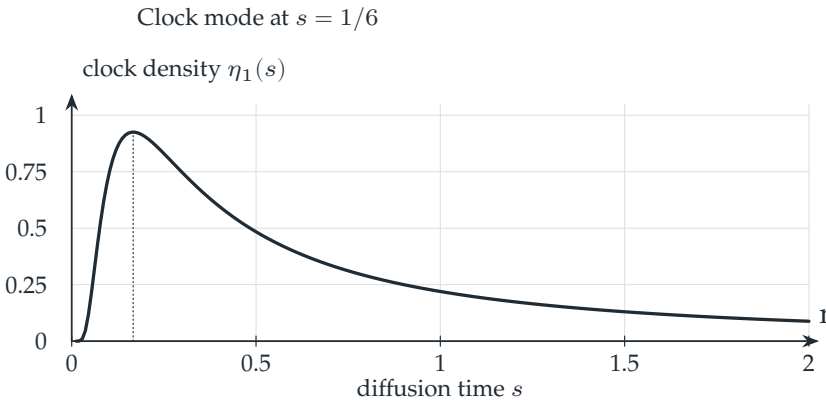
# A random diffusion clock

Traditional p. 21 / Part I: Theorem 10.2. The Poisson subordination formula for the finite positive model.

Square-root evolution can be written as an average of ordinary diffusion over a positive clock variable $s$:

$$e^{-t\sqrt{L}} = \int_0^\infty \eta_t(s) e^{-sL}\, ds, \qquad \eta_t(s) = \frac{t}{2\sqrt{\pi}} s^{-3/2} e^{-t^2/(4s)}, \quad t > 0.$$

Here $t$ is the external time; $s$ is the ordinary diffusion time being mixed. They are not interchangeable variables.




$$\int_0^\infty \eta_t(s)\, ds = 1$$

$$\int_0^\infty s\eta_t(s)\, ds = \infty$$

A probability clock need not have a finite mean.

The density $\eta_1$ peaks at $s = 1/6$ but has a long right tail. The drawing ends at $s = 2$; it does not claim that the missing tail has zero mass.

### Check the normalizer before interpreting a density

Under $s = t^2/(4u^2)$, reversing the integration limits gives

$$\eta_t(s)\,|ds| = \frac{2}{\sqrt{\pi}} e^{-u^2}\, du.$$

Its integral over $u > 0$ is one, so $\eta_t$ is a probability density. On each eigenmode of $L$, the operator formula reduces to the scalar Laplace identity

$$\int_0^\infty \eta_t(s) e^{-\lambda s}\, ds = e^{-t\sqrt{\lambda}}.$$

The manuscript supplies that scalar identity and its spectral assembly. The appendix separately checks the change of variables and normalization; sampled curves are not used to certify an improper integral.

A normalized density need not have a finite mean. For large $s$, the exponential factor tends to one and $\eta_t(s)$ is proportional to $s^{-3/2}$. Consequently $s\eta_t(s)$ has a nonintegrable $s^{-1/2}$ tail. The clock has infinite mean even though its total mass is one.

> **Read this as:** Subordination is a weighted mixture of diffusion times. Replacing the clock distribution by a single average time is not justified, and here that average is infinite.

V18 / DIRICHLET GRID / THE MISSING FACTOR OF h

# A Green kernel is an inverse with units

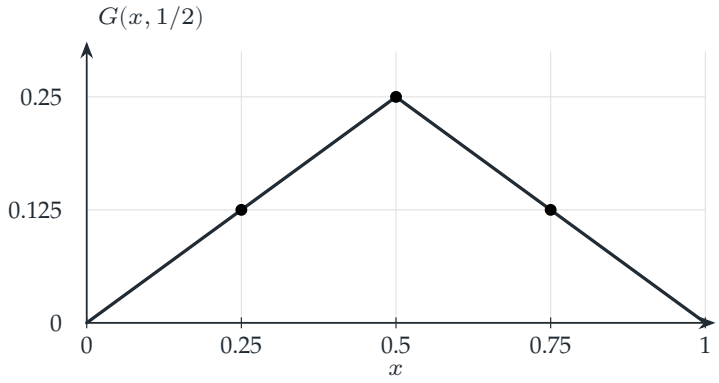
Traditional p. 25 / Part I: Theorem 11.1. Exact nodal Green reconstruction for this one-dimensional stencil.

Take three interior nodes $x_i = i/4$ on $[0, 1]$, with zero boundary values and $h = 1/4$. The discrete negative second derivative is

$$L_h = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}.$$

For the continuous Dirichlet problem, the Green function is $G(x, y) = \min(x, y) - xy$. At these nodes the exact comparison is

$$L_h^{-1} = h\,[G(x_i, x_j)]_{i,j=1}^{3} = \frac{1}{64}\begin{pmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{pmatrix}.$$




$64L_h^{-1}$, $h = 1/4$

| | | |
|---|---|---|
| 3 | 2 | 1 |
| 2 | 4 | 2 |
| 1 | 2 | 3 |

$L_h^{-1} = hG_{\text{nodes}}$

The tent is the central column of the continuous Green kernel. It has nodal values $(1, 2, 1)/8$; its discrete inverse column is $(1, 2, 1)/32$, smaller by the factor $h$.

### A forcing vector is not a quadrature-weighted measure

For sampled forcing $f_j$, the finite solution is

$$u_i = \sum_j (L_h^{-1})_{ij} f_j = h \sum_j G(x_i, x_j) f_j.$$

The factor $h$ converts samples into the quadrature weights appropriate to this formula. Dropping it multiplies the computed solution by four on the displayed grid.

A discrete unit coordinate vector and a discretized unit point mass use different normalizations. For $f = e_2$, the result is the second inverse column. For $f = e_2/h$, it is the unscaled tent sampled at the nodes. Either convention can be useful, but it must be specified before comparing answers.

Exact checks multiply $L_h$ by the proposed inverse at $1, 3, 7$, and $15$ interior nodes. The general identity is the manuscript theorem. These finite checks do not extend it to arbitrary meshes, dimensions, or boundary conditions.

> **Read this as:** An inverse kernel carries a measure convention. A correct-looking tent can give a wrong solution if the grid weight is omitted.

V19 / FINITE DIFFERENCES / A QUARTIC TEST PROBLEM

# Convergence with a nonzero exact error

Traditional p. 25 / Part I: Theorems 11.1–11.2. Exact nodal inversion and the scale of low-mode errors.

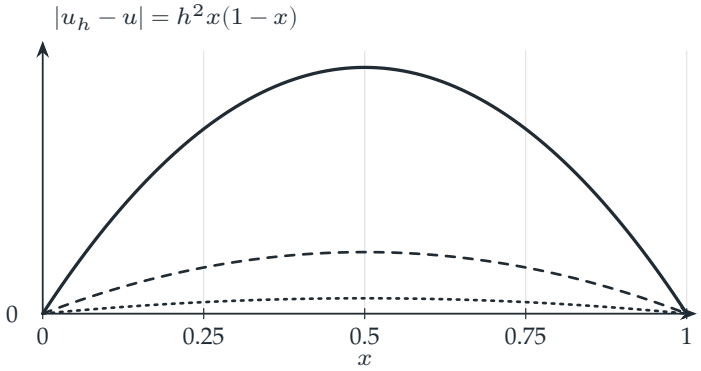
Use $u(x) = x - x^4$, which vanishes at both endpoints and solves $-u'' = 12x^2$. Its centered second difference is not exact:

$$L_h u|_{\text{nodes}} = 12x_i^2 + 2h^2.$$

Since the stencil differentiates the quadratic $x(1-x)$ exactly, the solution $u_h$ to the sampled forcing $12x_i^2$ satisfies

$$\boxed{u_h(x_i) - u(x_i) = -h^2 x_i(1-x_i).}$$

This gives both the sign and size of the error, rather than merely a log–log slope.




| grid spacing $h$ | maximum error |
|---|---|
| $\frac{1}{4}$ | $\frac{1}{64}$ |
| $\frac{1}{8}$ | $\frac{1}{256}$ |
| $\frac{1}{16}$ | $\frac{1}{1024}$ |

The three curves show magnitudes of the exact nodal errors for $h = 1/4, 1/8, 1/16$. Lines between grid points are visual guides, not an additional continuous numerical solution. The error is negative on every interior node.

### One refinement divides this error by four

All three grids include $x = 1/2$, where the maximum absolute nodal error is $h^2/4$. It equals $1/64$, $1/256$, and $1/1024$, respectively. Exact rational matrix inversion verifies these values independently of the closed error formula.

The manuscript also compares the discrete eigenvalues

$$\lambda_{k,h} = 4h^{-2}\sin^2(k\pi h/2)$$

with $(k\pi)^2$, with error at most $(k\pi)^4 h^2/12$. For a fixed mode $k$, the bound tends to zero quadratically. It does not claim a uniform small error for modes whose index grows like $1/h$.

The nodal Green identity is exact, while the discretization of the differential equation still has a truncation error. These statements concern different comparisons and do not contradict one another.

> **Read this as:** An exact inverse of the discrete operator does not make that operator equal to the differential operator. Keep algebraic inversion and continuum approximation separate.

V20 / FEYNMAN–KAC / DEPARTURE-STATE WEIGHTS

# Eight paths, one weighted expectation

Traditional p. 27 / Part I: Theorem 12.1. A finite path sum with its time convention displayed.

Let $K_2 = \frac{1}{4}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$, start at state $0$, use weights $w = (1, 1/2)$, and terminal observable $f = (2, 1)$. For three transitions, there are eight possible state sequences after the fixed initial state. We evaluate

$$\mathbb{E}_0 \left[ \prod_{r=0}^{2} w(X_r)\, f(X_3) \right].$$

Each path contributes its transition probability times its three departure weights times its terminal value.

| Path | probability | weight | terminal | contribution |
|---|---|---|---|---|
| 0000 | $27/64$ | $1$ | $2$ | $27/32$ |
| 0001 | $9/64$ | $1$ | $1$ | $9/64$ |
| 0010 | $3/64$ | $1/2$ | $2$ | $3/64$ |
| 0011 | $9/64$ | $1/2$ | $1$ | $9/128$ |
| 0100 | $3/64$ | $1/2$ | $2$ | $3/64$ |
| 0101 | $1/64$ | $1/2$ | $1$ | $1/128$ |
| 0110 | $3/64$ | $1/4$ | $2$ | $3/128$ |
| 0111 | $9/64$ | $1/4$ | $1$ | $9/256$ |

Total probability $= 1$; weighted terminal expectation $= 311/256$.

Every possible path is included. The probability column sums to one. The contribution column sums to $311/256$, agreeing exactly with the transfer calculation.

### Check which side of the transition gets weighted

The backward operator on observables is $D_w K_2^{\mathsf{T}}$. Therefore the expectation is the zeroth coordinate of

$$(D_w K_2^{\mathsf{T}})^3 f.$$

Its forward dual is $K_2 D_w$: weight the mass at the departing state, then move it. Weighting arrival states instead gives $D_w K_2$, a different expression. The exact tests deliberately check that the two conventions do not agree in this example.

The answer $311/256$ exceeds one, but it is not a probability. The terminal observable can equal two. With $0 \le w_i \le 1$, the weighted expectation is bounded by the maximum terminal value, not necessarily by one.

This calculation certifies a finite three-step identity for a stated model. The continuous-time theorem uses the corresponding infinitesimal generator and potential, with the direction of action kept consistent.

> **Read this as:** A Feynman–Kac weight changes the value assigned to each path. Its location relative to a transition is part of the model, not a typographical choice.

V21 / SCHUR–FESHBACH / RETAINED RESOLVENT

# Elimination makes a frequency-dependent term

Traditional p. 28 / Part I: Theorem 13.1. A two-state Schur complement with the inverse checked directly.

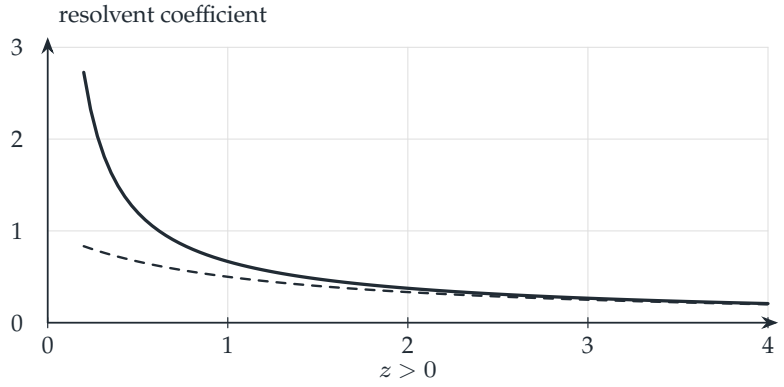
The two-state evolution from V3 has generator

$$A = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Retain the first coordinate and eliminate the second. For $z$ away from the relevant poles, the retained entry of $(zI - A)^{-1}$ is

$$R_x(z) = \left[ z + 1 - \frac{1}{z+1} \right]^{-1} = \frac{z+1}{z(z+2)}.$$

The term $\Sigma(z) = 1/(z+1)$ is the hidden coordinate's self-energy.




exact: $\dfrac{z+1}{z(z+2)}$

omit hidden state: $\dfrac{1}{z+1}$

The pole at zero survives through hidden feedback.

The full inverse gives a rational response with two poles, at $0$ and $-2$. Its partial fractions are $R_x(z) = \frac{1}{2}z^{-1} + \frac{1}{2}(z+2)^{-1}$. The hidden intermediate inverse has its own excluded point $z = -1$.

### An eliminated variable has not disappeared

If the coupling to the second coordinate were simply deleted, the retained resolvent would be $(z+1)^{-1}$. That has a different pole and a different long-time limit. The correction is frequency dependent, so it cannot be represented by one constant scalar generator on the retained coordinate.

The Schur expression first requires $z+1 \neq 0$, as well as invertibility of the full matrix. Its simplified rational form has a removable intermediate singularity at $z = -1$, where the retained entry of the full inverse is well-defined and equals zero. Simplification does not retroactively make the eliminated block invertible there.

For $z > 0$, the Laplace transform of $(1 + e^{-2t})/2$ is exactly $R_x(z)$. The next page derives the time-domain equation producing that response.

> **Read this as:** The Schur complement is exact elimination, not independent dynamics on a smaller state space. Hidden evolution returns as a $z$-dependent self-energy.

V22 / EXACT RETAINED DYNAMICS / NOT A SCALAR SEMIGROUP

# Hidden state returns as memory

Traditional p. 29 / Part I: Theorem 13.2. The retained-state memory identity, including hidden initial data.

Write the same system as

$$x' = -x + y, \qquad y' = x - y.$$

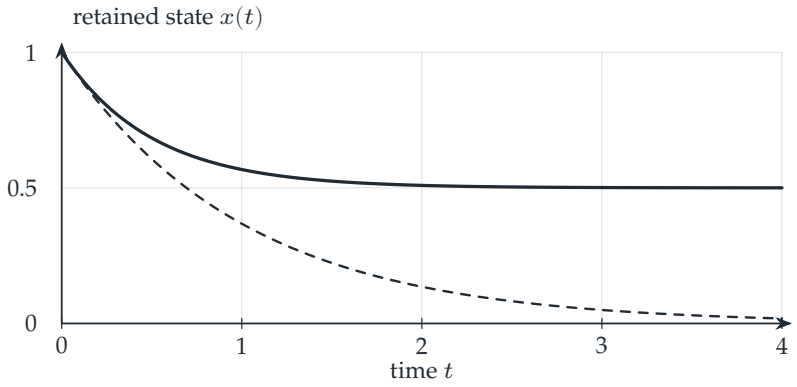
Solving the second equation first gives

$$y(t) = e^{-t}y(0) + \int_0^t e^{-(t-r)}x(r)\,dr.$$

Substitution is exact:

$$x'(t) = -x(t) + e^{-t}y(0) + \int_0^t e^{-(t-r)}x(r)\,dr.$$

The hidden initial state produces a separate forcing term. It must be retained even when the memory kernel is known.




exact: $\dfrac{1+e^{-2t}}{2}$

naive deletion: $e^{-t}$

$K(t) = e^{-t}$
returns hidden excursions.

For $(x(0), y(0)) = (1, 0)$, the exact retained coordinate is $(1 + e^{-2t})/2$. The dashed curve $e^{-t}$ solves the equation obtained by deleting the hidden coordinate rather than eliminating it.

### Projection can destroy the semigroup law

Let $R(t) = (1 + e^{-2t})/2$, the response to this fixed initial preparation. At $t = \log 2$,

$$R(t) = \frac{5}{8}, \qquad R(2t) = \frac{17}{32}, \qquad R(2t) - R(t)^2 = \frac{9}{64}.$$

The full two-state propagator is a semigroup, but this retained scalar response is not. Restarting it as a scalar model silently resets the hidden coordinate, which generally carries information from the first interval.

The appendix verifies the full differential equation, initial condition, integral memory equation, and Laplace resolvent in this model. Its kernel is $e^{-t}$, whose transform is the self-energy on V21.

> **Read this as:** Exact reduction retains a record of the eliminated state through convolution and initial forcing. A memoryless approximation must justify the information it removes.

V23 / NONDIAGONAL HIDDEN DYNAMICS / POLYNOMIAL EXPONENTIALS

# A repeated pole remembers a Jordan chain

Traditional pp. 29, 31 / Part I: Theorem 13.2 and Section 15. Functional calculus must retain nilpotent data.

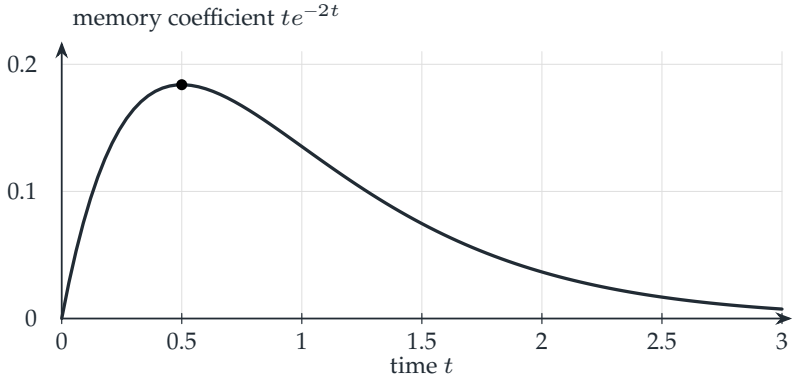
Now use a two-dimensional hidden block

$$J = -2I + N, \qquad N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad N^2 = 0.$$

Choose the entrance column $V = (0, 1)^{\mathsf{T}}$ and exit row $U = (1, 0)$. The hidden exponential is not $e^{-2t}I$:

$$e^{tJ} = e^{-2t}(I + tN), \qquad K_{\text{mem}}(t) = Ue^{tJ}V = te^{-2t}.$$

The nilpotent link carries an input from the second hidden coordinate to the first.




$$J = -2I + N, \quad N^2 = 0$$

$$e^{tJ} = e^{-2t}(I + tN)$$

$$\widehat{K}(z) = \frac{1}{(z+2)^2}$$

A repeated pole carries a jet.

The memory kernel starts at zero, becomes positive, and peaks at $t = 1/2$. A rule that retained the eigenvalue $-2$ but discarded the Jordan chain would predict the zero kernel for these entrance and exit maps.

### The transfer function sees pole order

The exact transform is

$$U(zI - J)^{-1}V = \frac{1}{(z+2)^2} = \int_0^\infty te^{-(z+2)t}\,dt, \qquad \operatorname{Re} z > -2.$$

The repeated pole and the polynomial factor $t$ are the same information in two domains. For a longer nilpotent chain, higher powers of $t$ occur with factorial denominators.

A zero value $K_{\text{mem}}(0)$ does not make the entire kernel zero. Nor do stable eigenvalues alone specify every transient contribution: entrance maps, exit maps, and nilpotent structure matter. Here the full kernel is nevertheless elementary, so its peak and transform can be checked without numerical fitting.

This page is an explicit finite realization. It does not show that an arbitrary continuum memory kernel admits a finite-dimensional representation of this form.

> **Read this as:** Repeated eigenvalues may carry directional information through Jordan chains. Keep the finite jet of the functional calculus, not just its values on the spectrum.

V24 / PHASE-TYPE SURVIVAL / TWO SUCCESSIVE STAGES

# A zero initial hazard can still lead to certain exit

Traditional p. 30 / Part I: Theorem 14.1. Survival, density, hazard, and moments have distinct meanings.

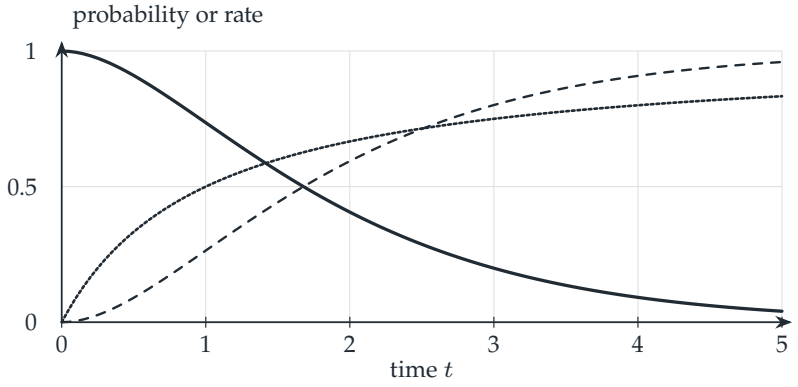
Let a transient process move through two successive rate-one stages, then leave the system. With column probabilities, the transient generator and starting vector are

$$T = \begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix}, \qquad p(0) = e_0.$$

Thus $p(t) = e^{-t}(1, t)^{\mathsf{T}}$. Its total remaining mass, exit density, and hazard are

$$S(t) = (1+t)e^{-t}, \qquad f(t) = te^{-t}, \qquad h(t) = \frac{t}{1+t}.$$

The initial hazard is zero because an exit needs two stages, not one.




solid: survival $S(t)$

dashed: absorbed mass

dotted: hazard $h(t)$

$h(0) = 0$, but eventual exit has probability one.

Survival falls from one to zero, while absorbed mass $1 - S$ rises to one. The dotted hazard is an instantaneous rate, not a probability, even though it shares the displayed numerical scale. Time is measured in inverse stage-rate units.

### Missing transient mass is absorbed mass

The column sums of $T$ are $0$ and $-1$. The exit-rate row is therefore $(0, 1)$, and $f(t) = (0, 1)p(t)$. Adding an absorbing state restores a three-state probability vector with total mass one.

Exact integration gives

$$\int_0^\infty f(t)\,dt = 1, \qquad \mathbb{E}[\tau] = 2, \qquad \mathbb{E}[\tau^2] = 6, \qquad \mathrm{Var}(\tau) = 2.$$

Equivalently, the mean is $\mathbf{1}^{\mathsf{T}}(-T)^{-1}e_0$. A single exponential waiting time with mean two would have constant hazard $1/2$ and variance four, so matching the mean alone would not reproduce this distribution.

The hazard formula is used where $S(t) > 0$. The model has that property for every finite $t \geq 0$. No division by an already vanished survival mass is required.

> **Read this as:** Survival tracks mass, density tracks exits per unit time, and hazard conditions on survival. Equal means or equal initial rates do not identify a waiting-time law.

V25 / HOLOMORPHIC CALCULUS / A RECTANGULAR COMPARISON

# Naturality also transports derivatives

Traditional p. 32 / Part I: Theorem 15.1. Exact intertwining passes to the specified functional calculus.

Consider a fine operator, a coarse operator, and a rectangular map

$$A_f = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 9 \end{pmatrix}, \qquad A_c = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}, \qquad C_r = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Direct multiplication gives $C_r A_f = A_c C_r$. For $F(z) = 1/(1+z)$, which is defined at both spectra, compute the matrices by inversion:

$$F(A_f) = \begin{pmatrix} 1/3 & -1/9 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/10 \end{pmatrix}.$$

$A_f$

| | | |
|---|---|---|
| 2 | 1 | 0 |
| 0 | 2 | 0 |
| 0 | 0 | 9 |

$F(A_f)$ for $F(z) = (1+z)^{-1}$

| | | |
|---|---|---|
| $\frac{1}{3}$ | $-\frac{1}{9}$ | 0 |
| 0 | $\frac{1}{3}$ | 0 |
| 0 | 0 | $\frac{1}{10}$ |

$$F(2) = 1/3$$

$$F'(2) = -1/9$$

The dashed block is $F(A_c)$, not just $F(2)I$.

The discarded eigenvalue $9$ contributes $1/10$ only on the fine side. On the retained Jordan block, the off-diagonal entry is $F'(2) = -1/9$, not zero. The rectangular comparison preserves that entry.

### Why spectral values are not enough

On the retained block $A_c = 2I + N$, with $N^2 = 0$, one has

$$F(A_c) = F(2)I + F'(2)N = \tfrac{1}{3}I - \tfrac{1}{9}N.$$

Hence $C_r F(A_f) = F(A_c) C_r$. Replacing the block by $F(2)I$ would lose the nonzero derivative term and no longer give the actual inverse of $I + A_c$.

For polynomials, naturality follows by transporting every operator power. The manuscript's finite holomorphic statement uses finite Taylor data at each primary block and the corresponding nilpotent orders. This example verifies a rational instance with an actual inverse rather than asserting that arbitrary scalar substitutions are meaningful.

The same mechanism explains why the Jordan memory kernel on V23 contains $te^{-2t}$. Exponentiation also has nonzero derivatives at repeated spectral values.

> **Read this as:** A functional calculus acts on operators, including their nilpotent parts. Exact intertwining transports that full action, not merely a list of scalar eigenvalue images.

V26 / END-TO-END MODEL / TWO OBSERVATIONS

# One complete coarse filter

Traditional pp. 8, 14, 19 / Part I: Proposition 9.3 with Theorems 5.1 and 8.2. Markov closure and likelihood sufficiency are separate.

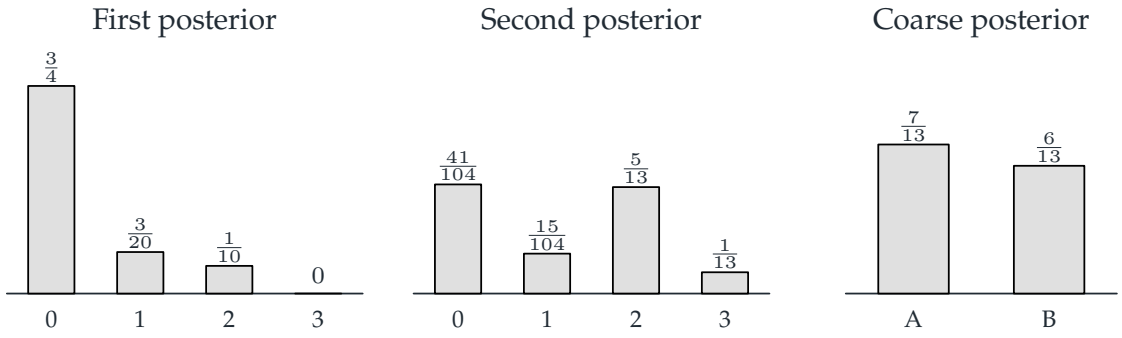
Use the four-state kernel $K$ and aggregation $C$ from V5, starting at $p_0 = e_0$. Observe two likelihood vectors constant inside both packets:

$$\lambda^{(1)} = (3/4, 3/4, 1/4, 1/4), \qquad \lambda^{(2)} = (1/2, 1/2, 1, 1).$$

Predict, then update at each observation. The fine calculation gives

$$p_1 = \tfrac{1}{20}(15, 3, 2, 0)^{\mathsf{T}}, \qquad p_2 = \tfrac{1}{104}(41, 15, 40, 8)^{\mathsf{T}}.$$

Evidence $Z_1 = 5/8$, $Z_2 = 13/20$, and $Z_1 Z_2 = 13/32$




Within-packet probabilities change during the two observations, but the packet totals agree exactly with an independent two-state filter. The second coarse posterior is $(7/13, 6/13)$.

### Check the normalizers, not only the final vector

The first and second predictive evidences are

$$Z_1 = \frac{5}{8}, \qquad Z_2 = \frac{13}{20}, \qquad Z_1 Z_2 = \frac{13}{32}.$$

The unnormalized product $D_2 K D_1 K p_0$ has mass $13/32$ and normalizes to $p_2$. Because the likelihood entries here lie in $[0, 1]$, they can serve as probabilities of the specified observations. This makes the product their joint evidence in the stated model.

Two distinct equalities justify the coarse computation. First, $CK = K_c C$ transports prediction. Second, the likelihood is constant on each packet, so $CD_r = \overline{D}_r C$ transports observation weighting. Their composition preserves both the posterior packet totals and the evidence factors.

The coarse filter does not recover the within-packet shape. If a later observation distinguishes states inside a packet, those missing probabilities can become necessary, as the counterexample on V10 shows.

> **Read this as:** Exact filtering at a coarse level needs both dynamical lumpability and observation sufficiency. Agreement under one step does not excuse testing the other.

V27 / REFINEMENT LIMITS / WHAT THE FINITE MODELS DO NOT PROVE

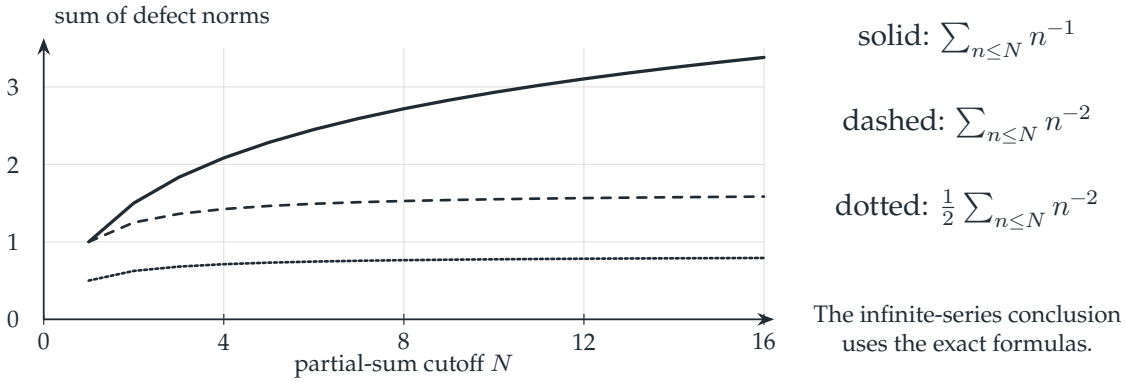
# Small at each level is not summable across levels

Traditional pp. 23, 34 / Part I: Corollary 10.5 and Section 17. Summability and uniform control are additional data.

For each $n \geq 1$, compare the two scalar operators $A_n = 0$ and $B_n = 1/n^2$ through the identity. Their ordinary defects have sizes $1/n^2$, while their square-root defects have sizes $1/n$:

$$\sum_n |A_n - B_n| < \infty, \qquad \sum_n |A_n^{1/2} - B_n^{1/2}| = \infty.$$

These are independent scalar comparisons, not a constructed inverse system. They isolate an error-scale obstruction without claiming a coherent refinement tower.




Partial sums through $n = 16$ illustrate the contrast. The convergence and divergence statements follow from the displayed series, not from the finite plot. A common positive gap gives the third, summable upper bound.

### A uniform hypothesis can change the conclusion

Replace the pair by $A_n = 1$ and $B_n = 1 + 1/n^2$. Then

$$\left|B_n^{1/2} - A_n^{1/2}\right| = \frac{1/n^2}{\sqrt{1 + 1/n^2} + 1} \leq \frac{1}{2n^2}.$$

This is the elementary form of the uniform-gap mechanism on V16. Without such information, transferring a summable generator-defect budget through a fractional function can fail.

The formal release keeps three original research questions separate: realization of broader transfer semigroups, uniform stochastic-refinement control, and latent packet closure. The exact finite examples here do not settle their existence, domain, topology, or uniformity requirements.

Likewise, exact hidden-state elimination only yields the finite memory models that were supplied. It does not establish a finite latent realization for every memory law. A complete list of matrix identities can still leave an infinite-dimensional approximation problem untouched.

> **Read this as:** A finite example can expose a missing hypothesis. It cannot supply a projective-limit construction, a uniform constant, or a continuum realization merely by passing at several levels.

V28 / READER'S GUIDE / CALCULATION, DRAWING, AND FORMAL CONTRACT

# How to use the evidence

Source release: Paper VI v0.5.0, Checked Fifth Block. Inherited evidence: September 2026.

The first 28 pages retain the worked mathematical atlas from the September release. Traditional Part I carries the corresponding core manuscript, including its figures and formal-status appendix; its theorem numbers are the references beneath these titles. The standalone guide uses internal visual-page links. The added pages beginning at V–29 address the strengthened branch in Traditional Part II.

The chosen release records all 33 scoped results complete, with 341 supporting theorem declarations in 47 proof modules and a saved audit of 782 namespace declarations. Those are inherited verification records, not a new Lean run and not counts of historically novel results. The three original conjectures retain their stated status.

### Read each kind of evidence at its actual strength

| Evidence | What is checked here | What it does not establish |
|---|---|---|
| Exact matrices | Stochasticity, intertwining, projector identities, inverse formulas, and specific defects. | The same properties for arbitrary partitions or unbounded operators. |
| Rational algebra | Posterior/evidence identities, polynomial integrals, nodal errors, and complete finite path sums. | An unstated observation model or a continuum convergence theorem. |
| Symbolic formulas | The displayed exponentials, Laplace transforms, derivatives, and finite spectral powers under stated conditions. | A general-purpose formal proof of every analytic identity. |
| Vector drawings | Reproducible geometry and sampled curves from the explicit models. | Numerical quadrature certificates or an exhaustive infinite-time experiment. |
| Saved Lean records | The original contract, source files, dependency pins, and earlier verification logs. | Fresh kernel verification of this editorial addition. |

### Reproduce a calculation before adapting it

The visual-appendix tree contains the mathematical models, drawing generator, and exact-example verifier. Its README records their paths and execution commands. The inherited September verifier records 204 exact or symbolic assertions and 11 deliberately false controls rejected. The monograph reproduces that symbolic suite with the available SymPy and mpmath environment. Its current execution record is identified on V–42. Its JSON record names every check and its finite scope.

The finite families are explicit: three-step Feynman–Kac paths are fully enumerated, selected rational priors are checked exhaustively within their stated denominator bound, and Green inverses are checked on four named grids. No finite count is used as a substitute for a general theorem.

To change a model, specify the state convention, partition, likelihood meaning, topology, and domain first. Then rerun the calculations and inspect the regenerated drawings. A preserved formula with changed assumptions is not automatically a preserved result.

> **Read this as:** Use the guide to see why a hypothesis matters and to compute a small case. Use the manuscript and its formal map for the scope of the general claim.

V29 / DIRECTED TRANSFER / A FINITE DYSON EXAMPLE

# Acyclic paths stop after finitely many transfers

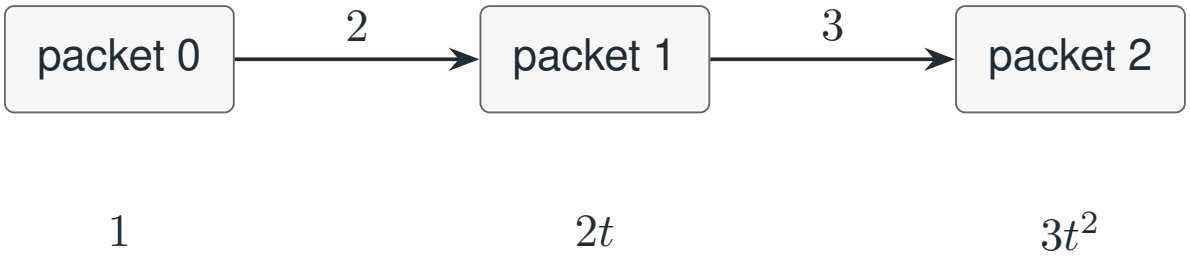
Traditional pp. 45, 46 / Part II: Theorem E2.8 and Corollary E2.9. Acyclicity concerns the transfer graph.

Let three scalar packets form the directed chain $0 \to 1 \to 2$, with transfer coefficients $2$ and $3$. The generator and its powers are

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix}, \qquad A^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 6 & 0 & 0 \end{pmatrix}, \qquad A^3 = 0.$$

Therefore the exponential is the exact polynomial

$$e^{tA} = I + tA + \frac{t^2}{2}A^2.$$




At $t = 1$: $(1, 2, 3)$ At $t = 2$: $(1, 4, 12)$

Only the empty path, one-edge paths, and the two-edge path exist. The coefficient $3t^2$ includes both the edge product $2 \cdot 3$ and the ordered-time simplex area $t^2/2$.

### Path length counts transfers, not elapsed time

Starting at $e_0$, the complete state is $(1, 2t, 3t^2)^{\mathsf{T}}$. No fourth packet or repeated cycle can contribute. Direct differentiation gives $x'(t) = Ax(t)$, including the relation $x_2'(t) = 3x_1(t)$.

If nonzero diagonal generators are added, a path acquires an integral of the within-packet evolutions between its transfer times. Acyclicity still bounds the number of transfers. It does not make the diagonal exponential factors polynomial in time.

This is a linear transfer example. The columns of $A$ do not sum to zero; the total of the displayed state is $1 + 2t + 3t^2$. It is consequently not a Markov probability model. Positivity of off-diagonal coefficients alone does not imply conservation of probability.

> **Read this as:** An acyclic packet graph terminates the transfer expansion. Nilpotence makes this particular full exponential polynomial; the two statements have different hypotheses.

V30 / REFINEMENT / A PERSISTENT RANGE EXCURSION

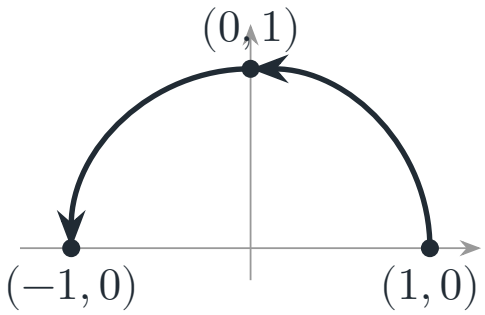
# Convergence does not close the dynamics

Traditional p. 51 / Part II: Theorem E4.8 and Example E4.10. Summability and closure are different tests.

Let the coarse space be $\mathbb{R}$ and every finer space be $\mathbb{R}^2$. The first connecting maps are $C(x,y) = x$ and $Rx = (x,0)$; all later maps are identities. Let the coarse evolution be the identity and all fine evolutions be rotations:

$$T(t) = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}, \qquad U_n(t) = CT(t)R = \cos t \quad (n \geq 1).$$

The reconstructed sequence is already constant above its first level. Only one adjacent defect is nonzero, so its defect sum is finite on every bounded time interval.




full two-step return: $-1$

project after each step: $0$

The full state returns to the coarse line with negative sign after two quarter turns. Resetting the hidden coordinate after the first quarter turn loses the returning contribution.

### A limit can fail the semigroup identity

At $s = t = \pi/2$,

$$U(s)U(t) = 0, \qquad U(s+t) = -1.$$

Write $P = RC = \operatorname{diag}(1,0)$. The exact algebraic difference is

$$U(s)U(t) - U(s+t) = CT(s)(P - I)T(t)R.$$

Its value is $1$ at these times, matching the missing hidden excursion. This is a failure of closure, not a failure of convergence.

The strengthened closure theorem asks for the corresponding excursion norm to tend to zero with refinement, uniformly on bounded time intervals. Generator control of range leakage supplies a sufficient condition in its stated contractive setting. The present rotation example never meets that vanishing condition.

> **Read this as:** A convergent sequence of reconstructed evolutions need not converge to a semigroup. The limit must also forget no excursion that can later return.

V31 / SUMMABLE DEFECTS / A SCALAR COMPLETION

# A tail bound states what refinement buys

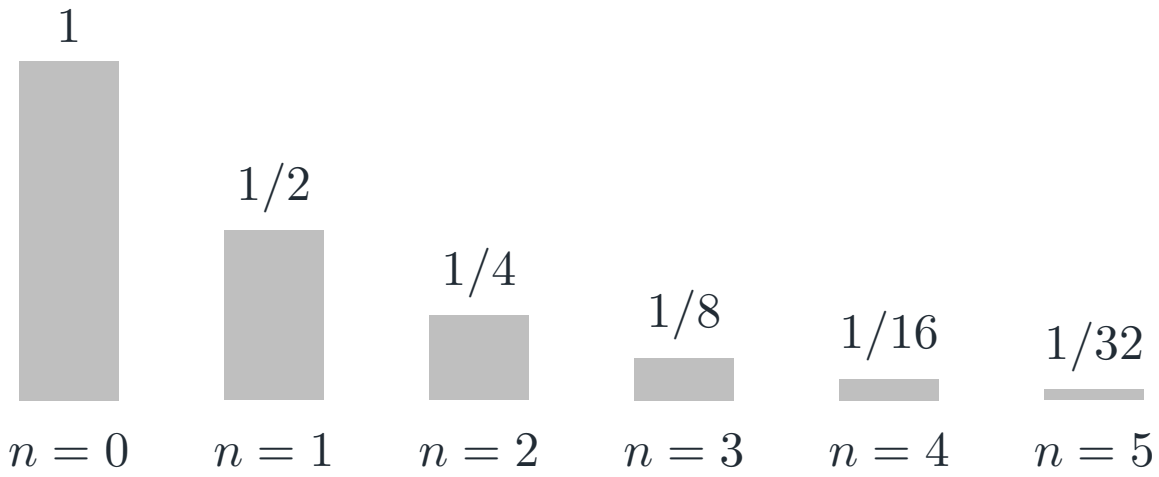
Traditional p. 50 / Part II: Theorem E4.5 and Corollary E4.7. This example has identity connecting maps.

Take $X_n = \mathbb{R}$, $C_n = R_n = 1$, and

$$a_n = 1 - 2^{-n}, \qquad A_n = -a_n, \qquad T_n(t) = e^{-a_n t} \quad (n \geq 0).$$

Each semigroup is contractive for $t \geq 0$. The exact adjacent generator-defect magnitude is

$$|A_{n+1} - A_n| = 2^{-(n+1)}, \qquad \sum_{j=n}^{\infty} |A_{j+1} - A_j| = 2^{-n}.$$




The bars show the certified tail $2^{-n}$ at horizon $T = 1$. They are bounds on a whole time interval, not errors sampled at the right endpoint.

### The estimate follows before any graph is drawn

For $0 \leq t \leq T$, the derivative of $a \mapsto e^{-at}$ has absolute value at most $T$ when $a \geq 0$. Thus

$$\sup_{0 \leq t \leq T} |T_n(t) - e^{-t}| \leq T 2^{-n}.$$

The limit is $e^{-t}$, which does have the semigroup law. Here every reconstruction is the identity, so the hidden-range excursion is identically zero. That closure fact is additional to the tail estimate.

In a general tower, the weighted estimate includes norms of composed connecting and reconstruction maps. Dropping those weights without contraction or another uniform bound can change a convergent error budget into an invalid one. Even a small defect at each individual level does not control the sum over all levels.

> **Read this as:** A summable tail gives an explicit stopping rule for a fixed horizon. It does not remove weights, supply uniformity on all time, or prove closure without its own hypotheses.

V32 / MARKOV DESCENT / TEST THE AGGREGATION KERNEL

# The obstruction lives in the forgotten directions

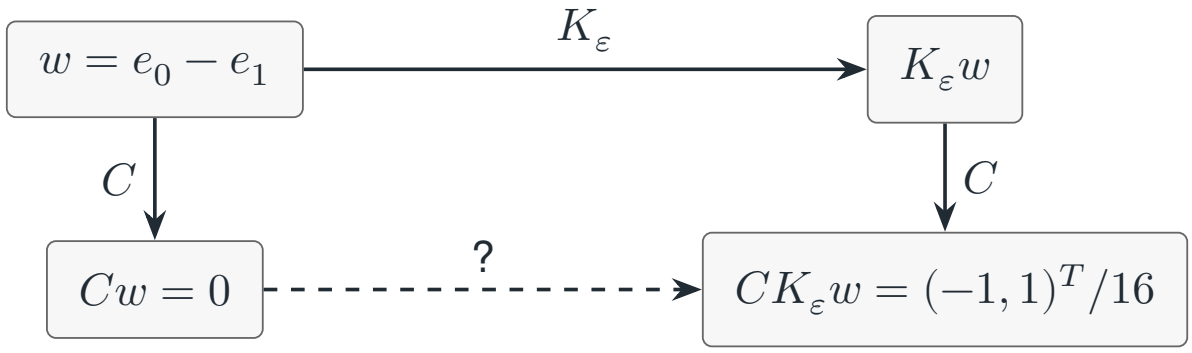
Traditional p. 54 / Part II: Theorem E5.3 and Corollary E5.4. A quotient criterion with a concrete witness.

Return to the four-state model of V–4 and the perturbed stochastic matrix $K_\varepsilon$ of V–6. Let

$$C = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \qquad w = e_0 - e_1.$$

The vector $w$ changes the fine distribution inside the first packet without changing its mass: $Cw = 0$. Yet

$$CK_\varepsilon w = \begin{pmatrix} -1/16 \\ 1/16 \end{pmatrix} \neq 0.$$




No linear coarse map can make this step: $0$ must map to $0$.

The coarse input cannot see $w$. The next coarse output can. Therefore no linear coarse operator can reproduce this fine step for every input.

### Why one witness rules out every coarse matrix

If some $\bar{K}$ satisfied $CK_\varepsilon = \bar{K}C$, then $CK_\varepsilon w = \bar{K}Cw = 0$, contradicting the displayed vector. The failure is not repaired by fitting a different coarse transition matrix to selected priors.

More generally, for a surjective aggregation map $C$, a linear map $M$ descends exactly if $CM$ vanishes on $\ker C$. The descended map is defined by $\bar{M}(Cx) = CMx$; the kernel condition makes this independent of the representative $x$.

For a deterministic partition of a finite stochastic system, differences $e_i - e_j$ within the same source block span the forgotten directions. Testing them gives the familiar equality of aggregate columns. Nonnegativity and column sums of the descended stochastic matrix then follow from the fine stochastic matrix and the partition structure.

> **Read this as:** The canonical obstruction is the action on information that aggregation discards. A nonzero witness in that kernel rules out exact descent at once.

V33 / ORNSTEIN–UHLENBECK / HERMITE ENERGY

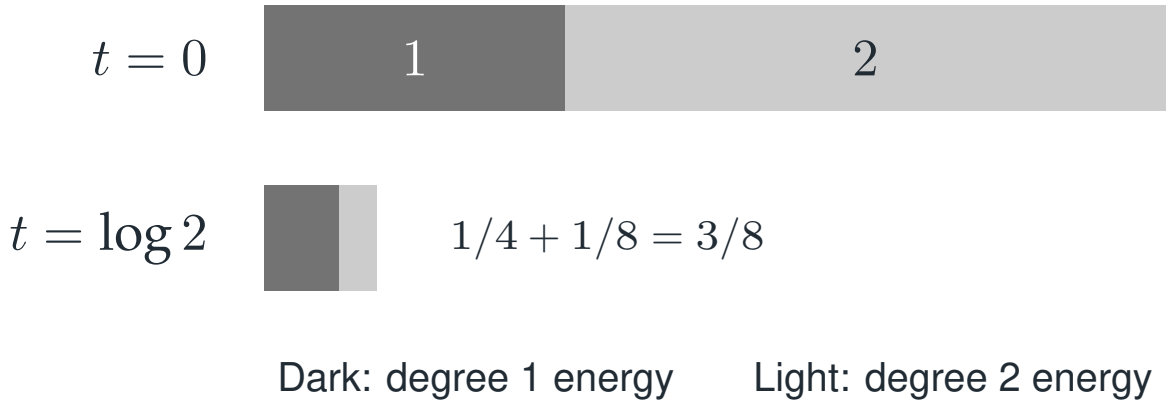
# Gaussian packets have separate clocks

Traditional pp. 58, 59, 60 / Part II: Theorem E7.1 and Corollaries E7.2, E7.4, and E7.5. Standard Gaussian measure.

In $L^2(\gamma)$ for the standard Gaussian law on $\mathbb{R}$, use the probabilists' Hermite polynomials

$$H_1(x) = x, \qquad H_2(x) = x^2 - 1, \qquad H_4(x) = x^4 - 6x^2 + 3.$$

The Ornstein–Uhlenbeck generator is $\mathcal{L}f = f'' - xf'$. Direct polynomial differentiation gives $\mathcal{L}H_n = -nH_n$ for these three polynomials. Their squared norms are $n!$ and different Hermite degrees are orthogonal.




For $f = H_1 + H_2$, the two bars are its separate squared-energy contributions. At $t = \log 2$, their sum is $3/8$, below the spectral-gap bound $3/4$.

### Reconstruction preserves the modes; time changes their weights

The semigroup sends $f = H_1 + H_2$ to

$$T(t)f = e^{-t}H_1 + e^{-2t}H_2, \qquad \|T(t)f\|_2^2 = e^{-2t} + 2e^{-4t}.$$

Its mean is zero. The Gaussian gap bounds this by $e^{-2t}\|f\|_2^2 = 3e^{-2t}$. It need not be an equality when multiple degrees are present.

For the nonnegative operator $-\mathcal{L}$, a square-root evolution replaces rate $n$ by $\sqrt{n}$. On $g = H_1 + H_4$, the coefficients at $t = \log 2$ are $1/2$ and $1/4$, whereas ordinary OU evolution gives $1/2$ and $1/16$. The eigenfunctions agree; the rates differ.

These calculations concern polynomial elements of Gaussian $L^2$. The general expansion and generator-domain statements belong to the traditional proof; a finite list of Hermite checks does not establish completeness of the basis.

> **Read this as:** Orthogonal packets make the energy accounting exact. A fractional generator alters the spectral clock while keeping the same Gaussian modes.

V34 / FIXED LIKELIHOODS / JOINT PROFILES

# Several observations determine one sufficient partition

Traditional p. 64 / Part II: Theorem E9.3. A finite family of strictly positive likelihoods.

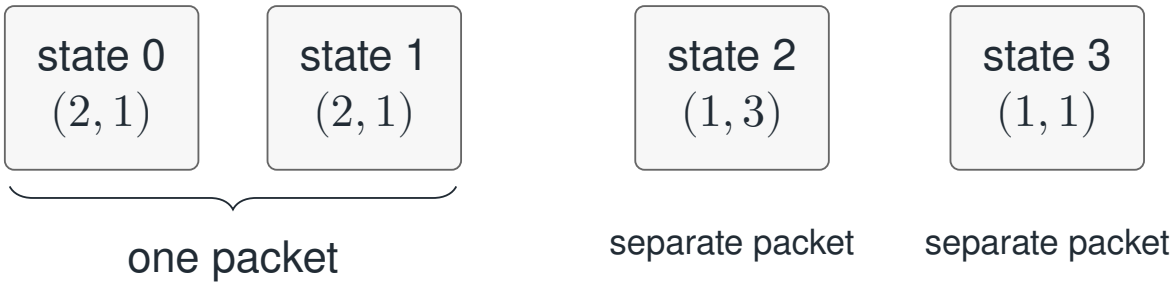
On four states, consider two likelihoods

$$\lambda^{(1)} = (2, 2, 1, 1), \qquad \lambda^{(2)} = (1, 1, 3, 1).$$

Two states may share a packet in the coarsest partition on which every listed likelihood is constant exactly when their entire likelihood profiles agree. The profiles are

$$(2, 1),\ (2, 1),\ (1, 3),\ (1, 1),$$

so the partition is $\{0, 1\}, \{2\}, \{3\}$.




$$\text{Joint likelihood profile} = (\lambda_i^{(1)}, \lambda_i^{(2)})$$

The first observation alone allows the last two states to be merged. The second observation distinguishes them. The shared partition must respect both likelihoods.

### A complete two-stage calculation

Start from the uniform prior. The coarse prior is $(1/2, 1/4, 1/4)$. After the first likelihood, it becomes $(2/3, 1/6, 1/6)$; after the second, it is

$$(1/2, 3/8, 1/8).$$

The direct fine update multiplies the likelihoods componentwise, giving weights $(2, 2, 3, 1)$ and posterior $(2, 2, 3, 1)/8$. Its packet masses are precisely the displayed coarse posterior.

The individual evidences are $Z_1 = 3/2$ and $Z_2 = 4/3$, and their product is $2$, the direct evidence of the product likelihood. These are relative likelihood weights, so evidence larger than one is allowed.

This partition supplies fixed coarse likelihoods that work for every prior. A single-packet aggregation always gives posterior $(1)$; that mass-only commutation does not identify likelihood profiles. A prior-dependent conditional average can summarize one update on a coarser partition, but requires the within-packet information that a fixed sufficient rule avoids.

> **Read this as:** Equal full likelihood profiles determine the coarsest partition on which every listed likelihood is constant. This partition admits fixed sufficient update rules.

V35 / RELATIVE ENTROPY / COARSE AND CONDITIONAL TERMS

# Information loss has an exact location

Traditional p. 66 / Part II: Theorem E9.8 and Corollary E9.9. Positive reference masses in this example.

Take

$$p = (1/2, 1/4, 1/8, 1/8), \qquad q = (1/4, 1/4, 1/4, 1/4)$$

and group the first two and last two states. Their coarse laws are $P = (3/4, 1/4)$ and $Q = (1/2, 1/2)$. Define $D(p\|q) = \sum_i p_i \log(p_i/q_i)$ with natural logarithms.

$$D(p\|q) = \tfrac{1}{4}\log 2$$



coarse information
$\frac{3}{4}\log 3 - \log 2$

conditional information
$\frac{5}{4}\log 2 - \frac{3}{4}\log 3$

The bar lengths use the displayed exact formulas evaluated for illustration. The split is an identity: coarse information plus posterior-weighted conditional information. It is not a fitted allocation.

### The chain rule identifies exactly what aggregation removes

Inside packet A, the conditional laws are $(2/3, 1/3)$ and $(1/2, 1/2)$; inside B they agree. Hence

$$\begin{aligned} D(p\|q) &= \tfrac{1}{4}\log 2, \\ D(P\|Q) &= \tfrac{3}{4}\log 3 - \log 2, \\ \sum_a P_a D(p(\cdot \mid a)\|q(\cdot \mid a)) &= \tfrac{5}{4}\log 2 - \tfrac{3}{4}\log 3. \end{aligned}$$

Adding the last two lines recovers the first. Every conditional term is nonnegative, so aggregation cannot increase relative entropy. Equality holds precisely when the occupied packets have the same conditional laws, under the stated support convention.

The weighting is by the packet mass $P_a$ from the first argument of $D(p\|q)$. Replacing it by $Q_a$ generally breaks the identity. Likewise, forgetting zero-mass conventions or allowing a positive $p_i$ over a zero $q_i$ changes which expressions are finite.

> **Read this as:** Coarse relative entropy measures distinguishability that survives aggregation. The conditional term records exactly the distinguishability removed inside the packets.

V36 / BAYESIAN FLOW / EVIDENCE CUMULANTS

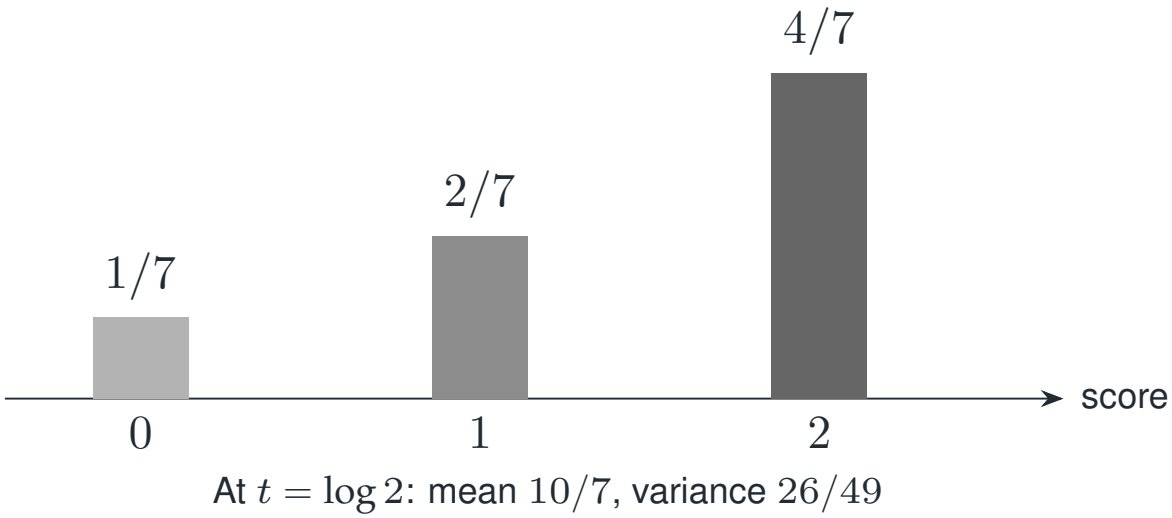
# The score rises because it has variance

Traditional p. 68 / Part II: Theorems E10.1 and E10.2. Exponential tilting of a fixed finite prior.

Start with equal prior masses on scores $u = (0, 1, 2)$ and likelihood $e^{tu}$. Put $q = e^t$. The evidence and tilted distribution are

$$Z(t) = \frac{1 + q + q^2}{3}, \qquad p(t) = \frac{(1, q, q^2)}{1 + q + q^2}.$$

At $t = \log 2$, they are $Z = 7/3$ and $p = (1, 2, 4)/7$.




The score values stay fixed. Their probabilities change. The rising mean score is a property of this tilted family, not an increase in the individual scores.

### Two derivatives have different statistical meanings

The first derivative of log evidence is the current mean score, and the second is its variance:

$$\frac{d}{dt} \log Z(t) = \mathbb{E}_{p(t)} u, \qquad \frac{d^2}{dt^2} \log Z(t) = \mathrm{Var}_{p(t)}(u) \geq 0.$$

At $q = 2$ the exact values are

$$\mathbb{E}u = \frac{10}{7}, \qquad \mathbb{E}u^2 = \frac{18}{7}, \qquad \mathrm{Var}(u) = \frac{26}{49}.$$

Thus the derivative of the mean is $26/49$ at this time. The flow equation $p_i' = (u_i - \mathbb{E}u)p_i$ gives derivatives $(-10, -6, 16)/49$, which sum to zero.

As $t \to \infty$, mass concentrates on the largest score present in the initial support. A score with zero initial prior mass is never created by multiplication. If several occupied states tie for the maximum, their relative prior proportions remain unchanged within that maximal packet.

> **Read this as:** Log evidence is a cumulant generator for the tilted law. Its curvature is variance, so the mean score is monotone even though the flow is nonlinear.

V37 / MEMORY COMPRESSION / AN OPTIMIZATION WITH A DEFINED COST

# Keep blocks by their certified contribution

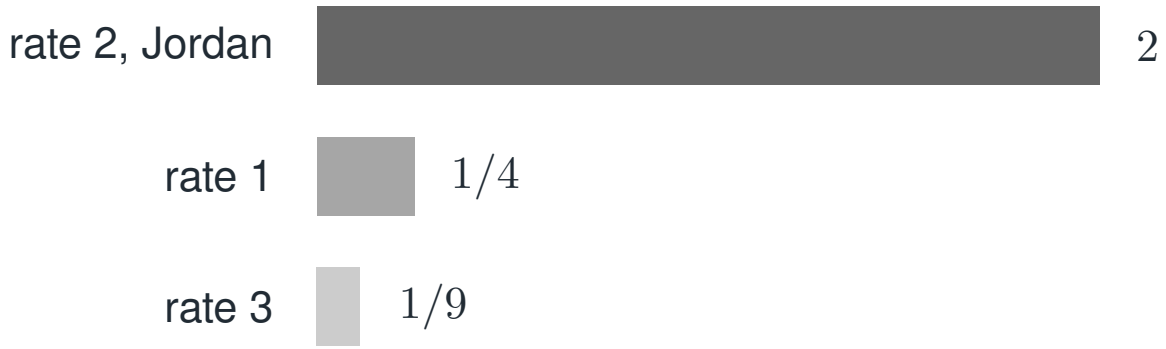
Traditional pp. 80, 81 / Part II: Theorem E15.1 and Corollary E15.2. Stable primary blocks; block count is the budget.

Consider the scalar memory kernel

$$\mathcal{K}(t) = \tfrac{1}{4}e^{-t} + (3+2t)e^{-2t} + \tfrac{1}{3}e^{-3t}.$$

It has three primary blocks. The middle block carries a Jordan term and must be retained or omitted as a whole for this selection rule. Their certified integrated weights are

$$w_1 = \frac{1}{4}, \qquad w_2 = \frac{3}{2} + \frac{2}{2^2} = 2, \qquad w_3 = \frac{1/3}{3} = \frac{1}{9}.$$




With one primary block retained, choose the middle block and certify omitted weight $13/36$. With two retained, add the rate-one block and the remaining certificate is $1/9$.

### The slowest eigenvalue is not necessarily the largest effect

The rate-two block has the greatest weight because its couplings and nilpotent contribution dominate. Retaining the slowest rate alone would omit weight $2 + 1/9$, a much larger certificate.

The descending order is $2, 1/4, 1/9$. A target tolerance $1/8$ needs two blocks according to this bound: one block leaves $13/36 > 1/8$, while two leave $1/9 \le 1/8$. Every term in this particular scalar kernel is nonnegative, so its omitted integral equals the certificate.

In general matrix kernels, cancellations can make the true error smaller. The theorem optimizes the sum of omitted block weights, not every possible approximation error. Moreover, a Jordan block may cost several state dimensions. A fixed budget in total hidden dimension is a different optimization problem from a fixed number of primary blocks.

> **Read this as:** The largest-weight rule is exact for its stated certificate and block-count budget. It is not a claim of globally optimal model reduction.

V38 / VOLTERRA STABILITY / SAME INITIAL DATA AND FORCING

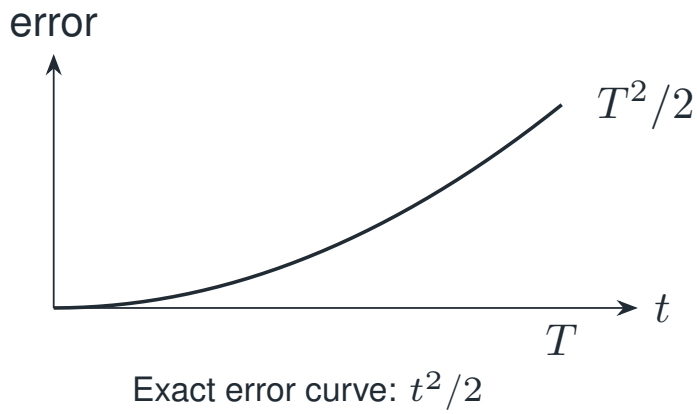
# A kernel certificate becomes a trajectory certificate

Traditional p. 83 / Part II: Theorem E15.6. Norms and a finite time horizon are explicit.

Let $N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, so $N^2 = 0$ and $\|N\|_2 = 1$. On $[0, T]$, compare the memory kernels $K(t) = N$ and $\widetilde{K}(t) = 0$, with identity free semigroup, no forcing, and initial state $x_0 = e_1 = (0, 1)^\mathsf{T}$. The equations are

$$x'(t) = \int_0^t N x(r)\, dr, \qquad \tilde{x}'(t) = 0.$$

Their exact solutions are $x(t) = (t^2/2, 1)^\mathsf{T}$ and $\tilde{x}(t) = (0, 1)^\mathsf{T}$.




interval certificate
$T^2 e^{T^2}$

The first-coordinate error is exactly $t^2/2$. The certificate bounds its maximum on the entire interval. The bound is deliberately shown separately from the trajectory.

### Track every constant before applying the estimate

Here $M_T = 1$, $B = \sup \|\tilde{x}\| = 1$, and

$$L = \|K\|_{L^1(0,T)} = T, \qquad \varepsilon = \|K - \widetilde{K}\|_{L^1(0,T)} = T.$$

The theorem yields

$$\sup_{0 \le t \le T} \|x(t) - \tilde{x}(t)\| = \frac{T^2}{2} \le T^2 e^{T^2}.$$

At $T = 1/2$, the exact error is $1/8$ and the certificate is $e^{1/4}/4$. Nilpotence makes this example polynomial; it does not remove the time horizon from the general theorem.

The two solutions share initial state, free semigroup, and forcing. If those are changed as well, their errors need additional terms. An integrated kernel error alone does not automatically control a trajectory on an infinite horizon.

> **Read this as:** The compression estimate transports an $L^1$ kernel budget into a finite-horizon state bound, using an explicit bound on the comparison trajectory.

V39 / FINITE REALIZATION / AN EXACT DIFFERENTIAL MODEL

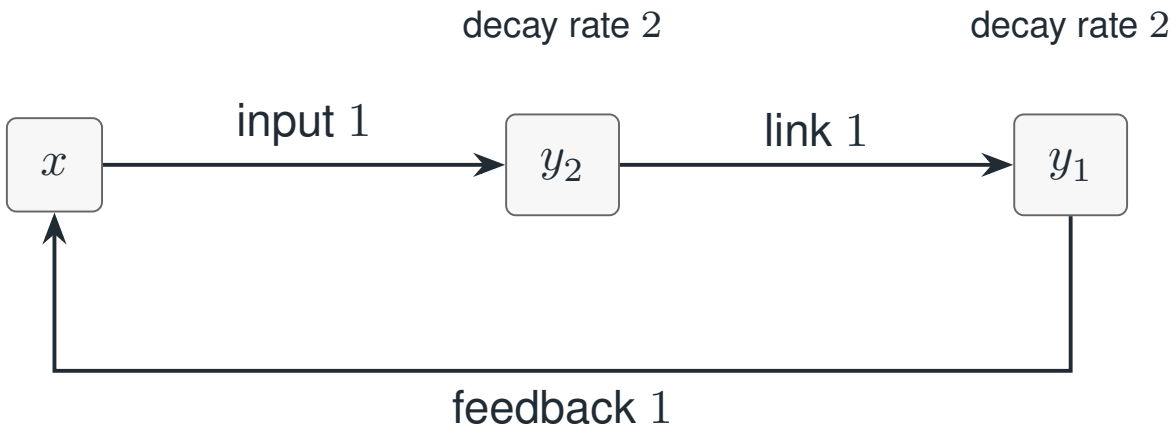
# Two auxiliary states realize a repeated pole

Traditional p. 82 / Part II: Corollaries E15.4 and E15.5. This is a supplied rational realization.

The kernel $K(t) = te^{-2t}$ from V–23 has transform $1/(z+2)^2$. It is realized by two auxiliary states:

$$\begin{aligned} \dot{y}_1 &= -2y_1 + y_2, \quad y_1(0) = 0, \\ \dot{y}_2 &= -2y_2 + x, \quad y_2(0) = 0, \\ \dot{x} &= ax + y_1. \end{aligned}$$

With $D = \begin{pmatrix} -2 & 1 \\ 0 & -2 \end{pmatrix}$, $B = (1, 0)$, and $C = (0, 1)^{\mathsf{T}}$, the feedback kernel is $Be^{tD}C = te^{-2t}$.




An input first enters $y_2$, transfers to $y_1$, and then returns to $x$. The two equal decay rates produce a repeated pole through the intervening link.

### Elimination recovers the convolution, including its preparation

Variation of constants gives

$$y(t) = e^{tD}y(0) + \int_0^t e^{(t-r)D}Cx(r)\,dr,$$

so the retained equation is

$$\dot{x}(t) = ax(t) + Be^{tD}y(0) + \int_0^t (t-r)e^{-2(t-r)}x(r)\,dr.$$

The convolution-only version is exact for $y(0) = 0$. For a nonzero preparation, $Be^{tD}y(0) = e^{-2t}(y_1(0) + ty_2(0))$ is an additional forcing.

The highest pole coefficient is nonzero in this example. A hidden Jordan block can become invisible to a transfer function if its entrance or exit coupling annihilates that coefficient. Pole order is therefore bounded by hidden Jordan depth, with equality requiring a visible highest term.

> **Read this as:** A polynomial-exponential kernel has a concrete finite state realization when the matrices are supplied. This does not realize every possible memory law with finitely many states.

V40 / COMPETING RISKS / RESIDUAL LIFE

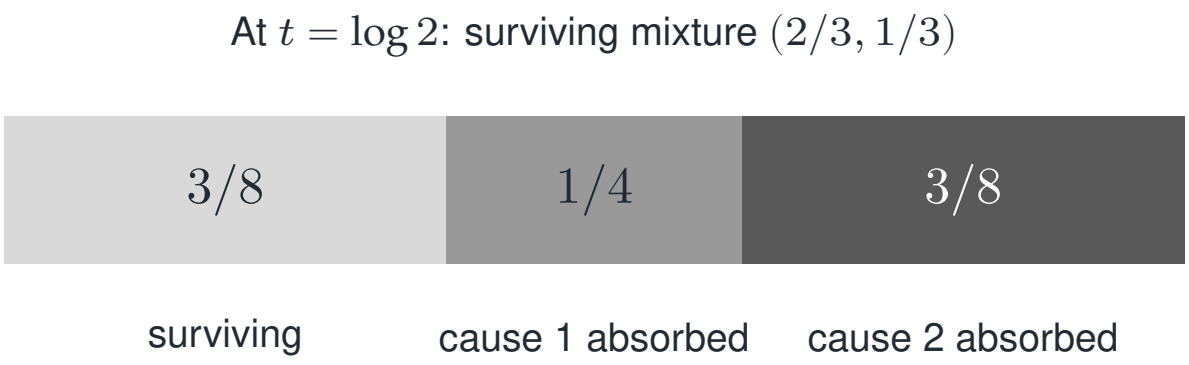
# Survivors have a different mixture of causes

Traditional p. 86 / Part II: Proposition E16.4 and Theorem E16.5. This example is stable and absorbs surely.

Two transient states have no internal transitions. Their exit rates are $1$ and $2$, their initial masses are both $1/2$, and the two states exit by distinct causes. Thus

$$A = \operatorname{diag}(-1, -2), \qquad S(t) = \tfrac{1}{2}e^{-t} + \tfrac{1}{2}e^{-2t},$$

with cause densities $f_1(t) = e^{-t}/2$ and $f_2(t) = e^{-2t}$.




At $t = \log 2$, surviving mass is $3/8$. Cause 1 has accumulated probability $1/4$ and cause 2 has accumulated probability $3/8$; the three pieces sum to one.

### Condition on survival before restarting the clock

At this time the transient masses are $(1/4, 1/8)$, so the conditional distribution of survivors is $(2/3, 1/3)$. Consequently

$$\frac{S(t+s)}{S(t)} = \frac{2}{3}e^{-s} + \frac{1}{3}e^{-2s}, \qquad \mathbb{E}[\tau - t \mid \tau > t] = \frac{2}{3} + \frac{1}{6} = \frac{5}{6}.$$

The original mean lifetime was $\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2} = 3/4$. Surviving longer changes the mixture toward the slower state; it does not restart the original mixture.

The cause densities at $t = \log 2$ both equal $1/4$, so the cause-specific hazards are both $2/3$ and total hazard is $4/3$. Hazards are rates and need not lie in $[0, 1]$. Each eventual cause probability is $1/2$, obtained by integrating its density. Stability of the subgenerator is part of that eventual-absorption statement.

> **Read this as:** Residual life restarts the same transient semigroup from the conditional surviving distribution. Cause probabilities, densities, and hazards answer different questions.

V41 / FUNCTIONAL CALCULUS / AN EXACT RATIONAL TEST

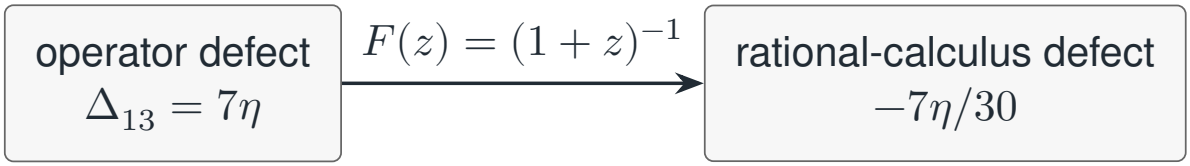
# An operator substitution transports the defect too

Traditional pp. 32, 47 / Part II: Theorem E3.4; compare Traditional Part I, Theorem 15.1, and V–25.

Use the operators from V–25, but perturb the rectangular comparison map:

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 9 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}, \quad C_\eta = \begin{pmatrix} 1 & 0 & \eta \\ 0 & 1 & 0 \end{pmatrix}.$$

The generator defect is $\Delta = C_\eta A - BC_\eta$, with only one nonzero entry, $\Delta_{13} = 7\eta$ (matrix indices begin at one).




Left inverse factor $1/3$ Right inverse factor $1/10$ Sign $-$

At $\eta = 0$: both defects vanish exactly

The perturbation lets the discarded eigenvalue enter the comparison. Exact naturality holds at $\eta = 0$; a nonzero defect is transformed, not erased, by the rational calculus.

### A resolvent identity supplies the sign and both inverse factors

For $F(z) = 1/(1+z)$, all required inverses exist, and

$$C_\eta F(A) - F(B)C_\eta = -(I+B)^{-1}\Delta(I+A)^{-1} = \begin{pmatrix} 0 & 0 & -7\eta/30 \\ 0 & 0 & 0 \end{pmatrix}.$$

This follows by multiplying the difference by $I+B$ on the left and $I+A$ on the right. The middle numerator is $(I+B)C_\eta - C_\eta(I+A) = -\Delta$.

The retained Jordan block still contributes $F'(2) = -1/9$, as on V–25. The new third-column discrepancy compares $F(9) = 1/10$ with $F(2) = 1/3$, multiplied by $\eta$.

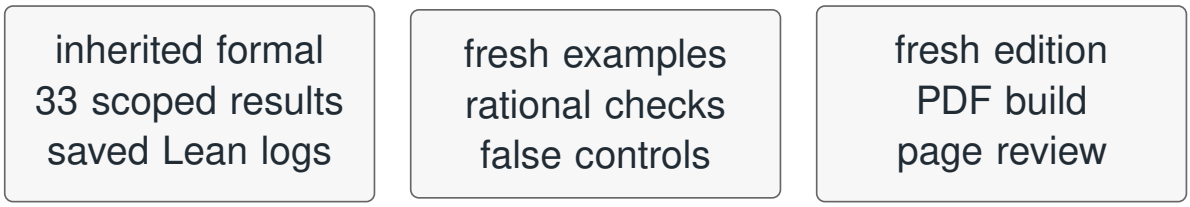
The general holomorphic defect formula integrates the corresponding resolvent defect under its contour and spectral assumptions. This rational example verifies an actual matrix inverse identity. It does not remove those analytic hypotheses or establish a calculus for arbitrary unbounded operators.

> **Read this as:** Exact intertwining gives exact naturality. When intertwining fails, the functional calculus has a structured defect whose factors can be computed.

V42 / EVIDENCE MAP / INHERITED RECORDS AND FRESH EXAMPLES

# What the final visual edition verifies

The final package retains original sources, new checks, build logs, and the traditional volume.




Separate claims, separate evidence records. No combined proof count.

The three evidence tracks have distinct scopes. The formal contract is inherited; the new exact examples and the present PDF build are fresh editorial checks.

### Read the records without adding incompatible counts

The September visual release records 204 exact or symbolic checks and 11 rejected false controls for its 28-page atlas. The scripts and saved report remain available. The monograph reruns the original symbolic suite with SymPy and mpmath. Its recorded inputs, outputs, resource limits, and execution logs accompany the current result.

The monograph also reruns the separate standard-library exact checker for V–29 through V–41. It uses rational matrices, finite polynomial algebra, Gaussian moments, and rational coefficients of logarithms. It checks the stated finite examples and rejects deliberately false alternatives. Its report records 120 assertions and 8 rejected false controls. A separate current verification also reruns 182 rational assertions and 4 rejected false controls for selected inherited models. That narrower check overlaps the full symbolic replay and remains separate from its count.

The new checks do not numerically certify improper integrals, all time horizons, or every analytic theorem. For example, the integrated memory weights use the stated exponential-integral formula; the checker verifies its coefficients and finite selection problem. Gaussian polynomial checks do not prove Hermite completeness. Rational entropy coefficient identities do not replace the general support-sensitive chain rule.

### The formal and editorial boundaries stay separate

The inherited Lean record covers 33 scoped results with 341 supporting theorem declarations in 47 proof modules. This guide neither enlarges that denominator nor claims a fresh kernel build. Traditional Part II retains broader written proofs; Part III retains earlier complementary results and the operator-substitution material migrated from Paper IV. The Lean companion maps their relationship to the formal baseline.

All 40 plates are editable vector sources: the 26 original plates plus 14 editorial plates. The portable build compiles the shipped geometry without needing symbolic regeneration. The supplied models and Python generators were used to regenerate all forty plates. Geometry samples remain drawing coordinates rather than additional analytic proofs.

> **Read this as:** A final edition can preserve strong earlier evidence while reporting exactly which checks were rerun. Treat each evidence count as a count of its own declared scope.