OPERATION KOMOREBI

REALIZATION CALCULUS

CORE SERIES VI

Transfer Semigroups and Memory

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

Evolution, conditioning, and hidden-state elimination share a packet language. The same transfer defect measures what a change of resolution preserves and what it loses.

Komorebi P.

Modern Monograph • 9 October 2026

How to use this volume

Part I. The complete released 33-result formal core, including nonnormal memory and canonical finite holomorphic naturality.

Part II. The complete strengthened mathematical development: bounded refinement towers, Bayesian information, Hermite transfer, quantitative memory compression, and competing risks.

Part III. Distinct retained results from the earlier human edition, exact examples, and the bounded-operator reciprocal calculus received from Paper IV.

Part IV. Forty-two visual pages work through exact finite models, analytical examples, and the boundaries of the formal contract.

Reading this edition

Paper VI has several substantial source branches. This edition preserves the latest formal-core manuscript in Part I and the complete strengthened standalone manuscript in Part II. Part III retains twenty-four earlier results with distinct statements or hypotheses, five exact examples, and the holomorphic operator-substitution section migrated from Paper IV. Overlap is intentional and is not a claim of independent novelty for each proof environment.

The formal denominator is exactly the thirty-three scoped results of Part I. The inherited September 6, 2026 release records 341 explicit Lean theorem declarations across 47 proof modules and an audit of 782 namespace declarations. These are different units. No fresh Lean compilation was available for this October edition, and the formal count does not automatically include Parts II or III. The Lean guide gives the exact targets and specialization boundaries.

Part I includes three explicitly open conjectures. Part II supplies bounded or finite theorems for several related questions; it does not settle arbitrary continuum limits, unbounded-generator convergence, general statistical sufficiency, or globally optimal nonnormal model reduction. The reader should keep each result's declared space, topology, positivity, and stability hypotheses. Part I's mass-only Bayesian commutation criterion assumes at least two coarse blocks; Part II's stronger sufficiency convention also preserves within-block conditionals, so its one-block case still requires a constant likelihood.

The current checks, historical compiler evidence, and numerical regressions are recorded separately in the accompanying package. Finite checks and illustrations have bounded scopes. They do not replace proofs. The September 12 worked visual appendix, its exact models, and additional guides to the strengthened and retained material are supplied in the visual companion.

Read Part I for the common operator language and the exact released formal contract. Read Part II for bounded Banach-space development and the broader probabilistic and analytical realizations. Read Part III for product-topology limits, reversible coarse dynamics, prior-restricted and coarse filtering, killed-transfer perturbations, rational realization, and the transferred operator reciprocal family. Section numbers in Part II have prefix E; those in Part III have prefix R. Original labels and statement locations are mapped in the editable sources.

The final mathematical audit makes explicit the hypotheses needed for contour winding, positive memory decay, finite moment closure, and eventual absorption. It also clarifies summability across varying spaces. Every substantive change is recorded against preserved sources. Historical source test counts retain their historical dates.

Contents

Reading this edition i

Part I The released formal core 1

Part I: The released formal core 1

1 The transfer-semigroup realization 2

2 Packet blocks of a semigroup 3

3 Refinement defects and the Duhamel identity 5
3.1 Resolvent form . . . 6

4 Markov seeds and stochastic transfer 7

5 Markov lumping as exact refinement 8

6 Conditional expectation as a stochastic projector 9

7 Bayesian conditioning as a normalized nonlinear operator 11
7.1 Packet evidence and reconstruction . . . 11

8 Exact Bayesian refinement and information-loss packets 13

9 Bayesian flows, conjugacy, and hidden states 18
9.1 Hidden Markov filtering . . . 19

10 Fractional powers as spectral packet transfer 20
10.1 Square-root subordination . . . 21
10.2 Fractional naturality under a bounded refinement defect . . . 22

11 Green operators and the Poisson seed 25
11.1 Low-mode convergence . . . 25

12 Finite-state Feynman–Kac transfer 27

13 Exact elimination of latent packets 28
13.1 Time-domain memory . . . 29

14 Survival and hazard semigroups 30

15 One common operator language 31

16 Numerical certificates 32

17 Remaining conjectural extensions 34

18 Computational implementation 34
18.1 Representative algorithms . . . 35

19 Conclusion 35

A Lean verification and scope 37

Part II Extended mathematical development 39

Part II: Extended mathematical development 39

E1 Introduction 39

E2 Packet resolved semigroups 41
E2.1 Directed path expansion 45

E3 Generator defects and functional calculus 46

E4 Refinement towers and limiting evolution 48
E4.1 Exact inverse limits 48
E4.2 Summable adjacent defects 49

E5 Finite Markov transfer and lumping 53

E6 Conditional expectation and martingale packets 56

E7 Gaussian transfer and Hermite packets 58

E8 Bayesian transfer and evidence packets 61

E9 Bayesian coarsening and finite sufficiency 63

E10 Bayesian flows and hidden Markov recursion 67
E10.1 Prediction and update 69

E11 Fractional functional calculus 70

E12 Green operators and finite difference transfer 73

E13 Finite state Feynman Kac transfer 76

E14 Schur Feshbach elimination 77
E14.1 Time domain memory 78

E15 Memory modes and controlled compression 80

E16 Survival, moments, and competing risks 84

E17 Finite illustrations and computational checks 87

E18 Limits and further questions 88

E19 Conclusion 89

Part III Retained results and operator calculus **90**

Part III: Retained results and operator calculus **90**

R1 Transfer paths and product-topology limits 90

R2 Reversible coarse dynamics 92

R3 Prior-restricted sufficiency and filtering 93

R4 Fractional modes and killed transfer 96

R5 Memory compression, rational realization, and phase-type consequences 97

R6 Earlier exact examples 99

R7 Holomorphic operator substitution 100

Part IV Worked visual atlas **102**

Part IV: Worked visual atlas **102**

List of Figures

1 The common operator language used throughout the paper. Semigroups, resolvents, normalized Bayesian maps, and fractional powers are distinct realizations of the same transfer stage. . . . 2
2 Two rational transfer steps resolved through all three intermediate states. The displayed products are computed exactly before plotting. This finite illustration explains the intermediate-packet sum. . . . 4
3 Norms of the packet blocks $P_\beta e^{tA} P_\alpha$. Off-diagonal entries quantify transfer between packets. . . . 5
4 Exact generator intertwining gives a vanishing semigroup defect. A controlled perturbation produces a nonzero defect propagated by the Duhamel identity. . . . 6
5 The point mass δ_0 propagated by repeated symmetric random-walk transfer. . . . 7
6 Exact lumping agrees at every horizon. A small generator-level or one-step defect creates a controlled coarse discrepancy. . . . 9
7 Hermite–Gaussian packets, which diagonalize the Ornstein–Uhlenbeck semigroup. The finite numerical release verifies their Gaussian orthogonality through quadrature. The five displayed modes run from n = 0 at the bottom to n = 4 at the top. Vertical offsets separate the curves. . . . 10
8 A finite prior and posterior split into two state packets. Evidence changes both the local packet shape and the global packet mass. . . . 12
9 Two rational priors with the same two coarse masses. A blockwise constant likelihood gives the same coarse update, while a likelihood varying within a block separates the posteriors. The exact fractions are calculated before plotting. . . . 13
10 For strictly positive likelihoods and at least two coarse packets, the Bayesian refinement square commutes for every prior exactly when the likelihood is constant on each coarsening fiber, up to one global scale. . . . 16
11 When packet sufficiency fails, the exact defect factors through the difference between the conditional effective likelihood and the declared coarse likelihood. . . . 17
12 Continuous likelihood accumulation produces a nonlinear semigroup on the probability simplex. . . . 18
13 A conjugate Bayesian update viewed as exact transfer of exponent packets. . . . 19
14 A hidden Markov filter alternates a linear prediction semigroup with nonlinear Bayesian normalization. . . . 20
15 Different spectral exponents produce different transfer profiles from the same point seed. 21
16 A generator-level refinement defect transfers quantitatively to fractional powers and then to fractional semigroups. The two upper curves are theorem certificates, not fitted error bars. . . . 24
17 A section of the continuum Green kernel and the scaled discrete inverse. They agree at the grid nodes. . . . 26
18 The first finite-difference eigenvalue converges with observed order approximately two. 27
19 Potential-weighted transfer values computed by the finite Feynman–Kac operator. Exhaustive path enumeration agrees with the matrix formula in the supplied tests. . . . 28
20 The norm of the exact memory kernel produced by eliminated packets. . . . 30
21 Survival and instantaneous hazard produced by one transient transfer semigroup. . . . 31
E1 Packet block norms at three times. Transfer appears in the off diagonal blocks as soon as the generator couples the two packet spaces. The three panels are shown in consecutive figures. . . . 44

E2 Packet block norms at time 0.8. The same fixed packet spaces and norm scale are used in the preceding panels. . . . 44
E3 Packet block norms at time 2.0. The same fixed packet spaces and norm scale are used in the preceding panels. . . . 45
E4 A finite semigroup defect and a direct Duhamel bound. The bound is designed to be uniform rather than sharp in this example. . . . 48
E5 Exact lumping agrees at every horizon. A small violation of the block transition condition produces a visible but controlled discrepancy. . . . 56
E6 Energy carried by the mean packet and three successive martingale difference packets in a finite weighted example. . . . 58
E7 Hermite packet multipliers for the Ornstein Uhlenbeck semigroup and two fractional powers of its number operator. . . . 61
E8 A prior and posterior on two coarse packets. The likelihood is constant in the first packet and varies inside the second, so only the second local conditional distribution changes. . . 63
E9 Relative information separated into a coarse packet contribution and conditional contributions inside the packets. . . . 67
E10 A finite Bayesian flow. Mass concentrates on the states with the largest score, while every relative odds ratio follows an exact exponential law. . . . 69
E11 Fractional semigroup multipliers at time one. Smaller exponents preserve more of the high spectral range. . . . 73
E12 The continuum Dirichlet Green section and the scaled discrete inverse. The nodal values agree exactly. . . . 75
E13 A two-coordinate Jordan hidden block has an exponential decay factor and a polynomial factor. The nilpotent polynomial is checked exactly at the displayed rational times. The exponential curve is a plotted illustration of the formula. . . . 80
E14 A stable Jordan block produces a polynomial exponential memory mode. A diagonal hidden block produces only pure exponentials. . . . 82
E15 Full and truncated memory kernels with their pointwise difference. The bound uses integrated error. . . . 84
E16 Survival and two cause specific event densities from one transient semigroup. . . . 87

List of Tables

1 The transfer-semigroup dictionary in the common operator language. . . . 31
2 Selected results from the machine-readable numerical record. . . . 33
E1 The transfer operations studied in the article and the condition that makes each coarse description exact. . . . 41
E2 Numerical tolerances in the finite verification suite. The proofs do not depend on these computations. . . . 88

Part I The released formal core

Projector-Resolved Seed Calculus VI

Transfer Semigroups: Stochastic, Bayesian, Fractional, and PDE Realizations

Overview

This paper develops the transfer-semigroup realization of the projector-resolved seed calculus. A seed is propagated by a one-parameter semigroup, a Markov kernel, a normalized Bayesian transfer, a fractional spectral power, a Green operator, or a potential-weighted Feynman–Kac operator. Exact packet projectors resolve every transfer into input-output blocks, and the semigroup law becomes block convolution over intermediate packets. Coarse and fine generators are compared by a refinement defect. A Duhamel identity proves that this generator-level defect is exactly the source of the corresponding semigroup defect and gives a quantitative stability bound. In the stochastic realization, point-mass seeds generate random walks and continuous-time Markov evolution; exact lumpability is the vanishing of the same refinement defect. Conditional expectation is treated as a weighted orthogonal packet projector, with nested partitions producing martingale-difference packets. Bayesian conditioning is a normalized nonlinear operator. Its independent-data composition law, packet-evidence decomposition, stability bound, conjugate updates, and continuous likelihood flow are proved explicitly. The Bayesian refinement conjecture is resolved in finite dimensions: for strictly positive likelihoods and at least two coarse packets, exact commutation for every prior is equivalent to blockwise-constant likelihoods up to global scale, and every failure factors through an explicit information-loss packet. Fractional powers are constructed packetwise through the spectral projectors of a positive operator, and the square-root semigroup is recovered by subordination. The fractional-naturality conjecture is also resolved for bounded nonnegative self-adjoint refinement systems: the fractional intertwining defect has an exact resolvent integral, a gap-free Hölder certificate, a sharper common-gap estimate, and a Duhamel semigroup bound. Green kernels, finite-difference Poisson problems, and second-order low-mode convergence provide the PDE realization. A finite-state Feynman–Kac formula is proved by path enumeration. Exact Schur–Feshbach elimination shows that discarded latent packets return as a frequency-dependent self-energy and, in time coordinates, as a memory kernel. Survival and hazard functions are generated by transient semigroups. Nineteen code-generated figures, machine-readable results, the original thirty-seven automated tests, and seventeen added boundary-case checks accompany this reviewed edition. The inherited September 6 release records five checked Lean blocks covering all 33 scoped theorem-like results in Appendix A. Finite-dimensional theorems, classical infinite-dimensional extensions, numerical certificates, and open continuum claims are kept separate.

Scope and claim boundary. The paper proves the stated finite-dimensional semigroup, packet, Markov, Bayesian, conditional-expectation, fractional spectral, Green-kernel, Feynman–Kac, Schur–Feshbach, and survival identities. It also proves quantitative refinement-defect formulas under explicit norm hypotheses, including exact transfer of a bounded generator defect to every fractional power $0 < \alpha < 1$ for bounded nonnegative self-adjoint generators. It does not claim that an arbitrary infinite-dimensional operator generates a semigroup, that every stochastic process admits a finite packet representation, that every posterior family is conjugate, that finite Bayesian sufficiency automatically extends to infinite disintegrations, or that every PDE continuum limit follows from a finite test. Domain, closability, sectoriality, measurability, tightness, and compactness hypotheses remain necessary in their respective infinite-dimensional settings. Classical sources are retained in the separate end-of-series reference ledger rather than reproduced in this paper.

1 The transfer-semigroup realization

Let X be a finite-dimensional real or complex vector space, let $A \in \mathrm{End}(X)$, and let $s \in X$ be a primitive seed. The basic continuous transfer is

$$T(t) = e^{tA}, \qquad t \geq 0,$$

while a positive operator L is more naturally written

$$H(t) = e^{-tL}.$$

Both satisfy the semigroup law. The sign convention changes with the domain: Markov generators and evolution equations are usually written with e^{tA}; positive diffusion and elliptic operators are usually written with e^{-tL}.

The common operator chain is

$$\boxed{\text{seed} \longrightarrow \text{transfer} \longrightarrow \text{project} \longrightarrow \text{condition} \longrightarrow \text{refine} \longrightarrow \text{complete} \longrightarrow \text{reconstruct}.}$$

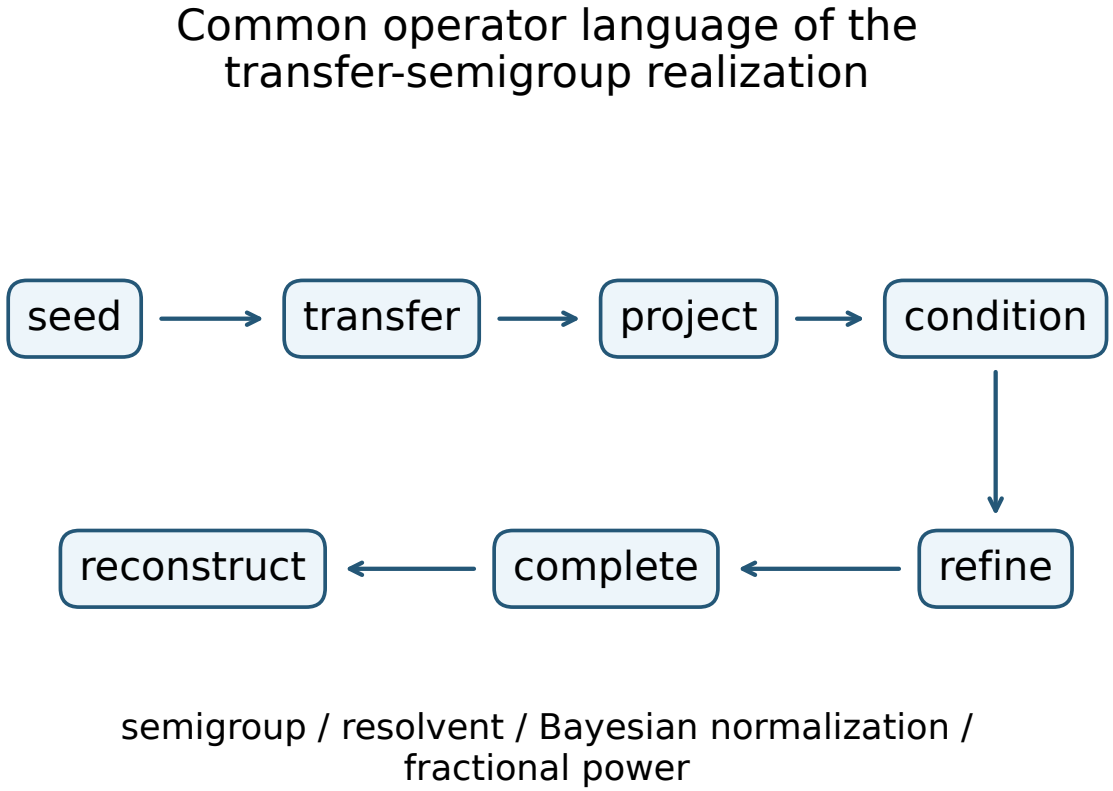


Figure 1. The common operator language used throughout the paper. Semigroups, resolvents, normalized Bayesian maps, and fractional powers are distinct realizations of the same transfer stage.

Definition 1.1 (Transfer-semigroup seed system)**.** A finite transfer-semigroup seed system is a tuple

$$\mathfrak{S} = (X, s, A, \{P_\alpha\}_{\alpha \in I}, \tau),$$

where X is the state space, s is the seed, A is a transfer generator, the P_α are pairwise orthogonal idempotents summing to the identity, and τ specifies the topology in which a completion or limit is interpreted.

Theorem 1.2 (Finite semigroup law)**.** *For every matrix A and all $s, t \in \mathbb{R}$,*

$$\boxed{e^{(s+t)A} = e^{sA}e^{tA}.}$$

Moreover, $T(0) = I$ and $T'(t) = AT(t) = T(t)A$.

Proof. Both exponentials are absolutely convergent matrix power series. Since every power of A commutes with every other power, the Cauchy product gives

$$e^{sA}e^{tA} = \sum_{m,n\geq 0} \frac{s^m t^n}{m!n!} A^{m+n} = \sum_{r\geq 0} \frac{(s+t)^r}{r!} A^r.$$

Termwise differentiation is valid in finite dimension and gives the derivative identities. □

Boundary 1.3 (Infinite-dimensional generators). For an unbounded operator, writing e^{tA} is not a definition unless a semigroup-generation theorem applies. The finite matrix identities in this paper remain algebraically correct models, but a continuum realization must state the operator domain, closure, growth bound, and topology of convergence.

2 Packet blocks of a semigroup

The central idea

A packet block remembers both the origin and the destination. Composition must sum over every intermediate packet, because an unobserved intermediate coordinate can carry a real contribution. The original formulas below express this bookkeeping exactly.

Let

$$P_\alpha P_\beta = \delta_{\alpha\beta} P_\alpha, \qquad \sum_{\alpha\in I} P_\alpha = I.$$

For any operator B, define

$$B_{\beta\alpha} = P_\beta B P_\alpha.$$

Then

$$B = \sum_{\alpha,\beta} B_{\beta\alpha}.$$

For the semigroup, write

$$T_{\beta\alpha}(t) = P_\beta e^{tA} P_\alpha.$$

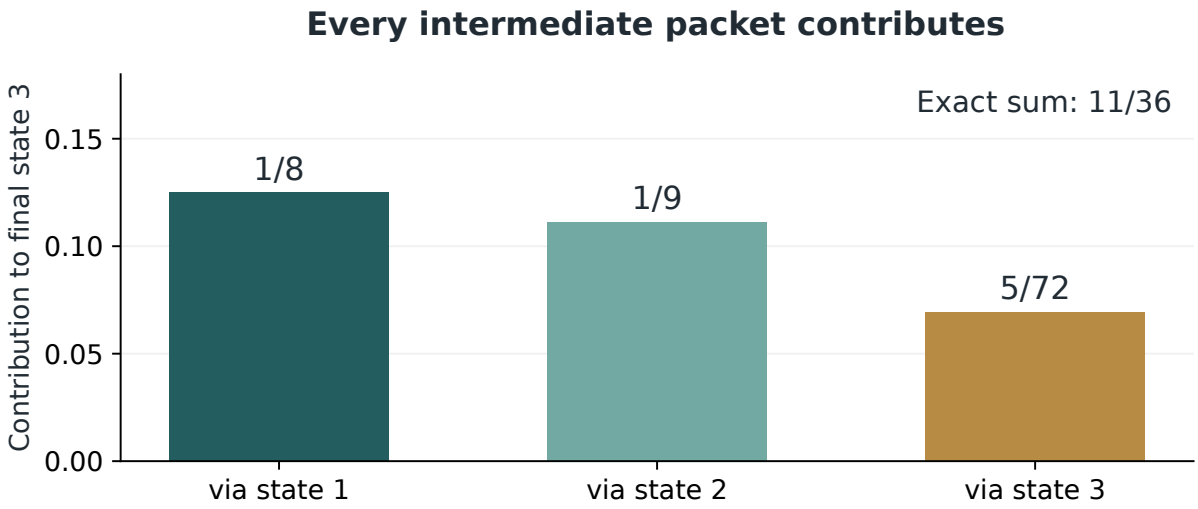


Figure 2. Two rational transfer steps resolved through all three intermediate states. The displayed products are computed exactly before plotting. This finite illustration explains the intermediate-packet sum.

Theorem 2.1 (Packet semigroup convolution)**.** *For every s, t and every pair of packets α, β,*

$$\boxed{T_{\beta\alpha}(s+t) = \sum_{\gamma \in I} T_{\beta\gamma}(s) T_{\gamma\alpha}(t).}$$

Thus the semigroup law becomes ordinary matrix multiplication in the packet indices.

Proof. Insert the resolution of the identity between the two semigroup factors:

$$\begin{aligned} P_\beta e^{(s+t)A} P_\alpha &= P_\beta e^{sA} \left(\sum_\gamma P_\gamma \right) e^{tA} P_\alpha \\ &= \sum_\gamma (P_\beta e^{sA} P_\gamma)(P_\gamma e^{tA} P_\alpha). \end{aligned}$$

□

Corollary 2.2 (Reducing packets)**.** *If $P_\beta A P_\alpha = 0$ for every $\beta \neq \alpha$, then*

$$P_\beta e^{tA} P_\alpha = 0 \quad (\beta \neq \alpha)$$

for every t. Each packet evolves independently under $P_\alpha A P_\alpha$.

Proof. The hypothesis makes A block diagonal. Every power A^m is block diagonal with respect to the same projectors, and therefore so is the exponential series. □

Remark 2.3 (Path expansion). Expanding the exponential and inserting $I = \sum_\alpha P_\alpha$ between every pair of generators gives

$$P_{\alpha_m} A^m P_{\alpha_0} = \sum_{\alpha_1, \ldots, \alpha_{m-1}} P_{\alpha_m} A P_{\alpha_{m-1}} \cdots P_{\alpha_1} A P_{\alpha_0}.$$

Every coefficient is therefore a sum over packet-transfer paths. The numerical library verifies that truncating this path series converges to the ordinary matrix exponential.

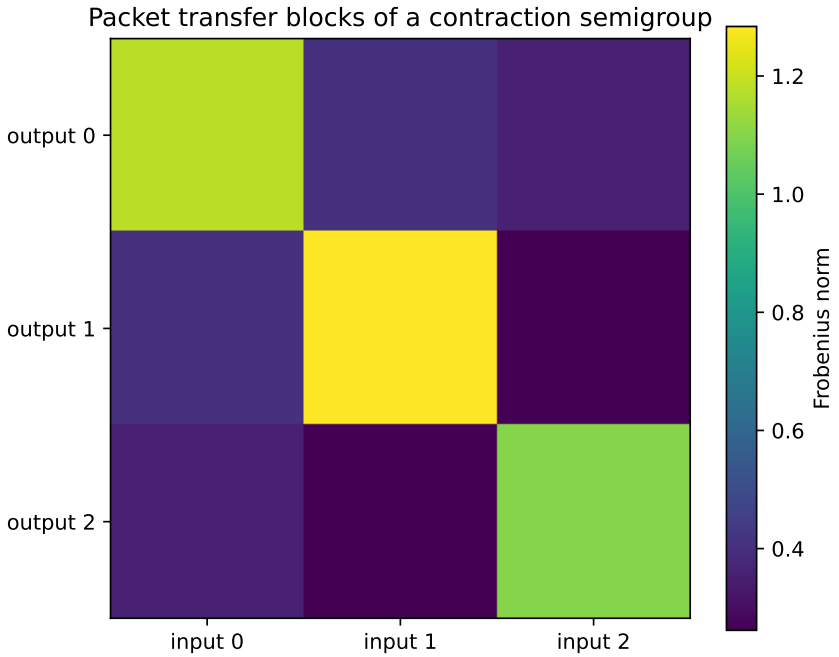


Figure 3. Norms of the packet blocks $P_\beta e^{tA} P_\alpha$. Off-diagonal entries quantify transfer between packets.

3 Refinement defects and the Duhamel identity

> **How to read the result**
>
> Read the generator defect as a mismatch between two declared paths: evolve and then coarsen, or coarsen and then evolve. The Duhamel identity transports that mismatch through time. Its estimate depends on the stated operator norms and growth bounds.

Let X_f and X_c be fine and coarse spaces, with generators A_f and A_c, and let

$$C : X_f \to X_c$$

be a coarsening map. Define the generator-level refinement defect

$$\boxed{\Delta = CA_f - A_c C.}$$

The corresponding semigroup defect is

$$\mathfrak{D}(t) = Ce^{tA_f} - e^{tA_c} C.$$

Theorem 3.1 (Exact Duhamel refinement identity). *For every $t \geq 0$,*

$$\boxed{Ce^{tA_f} - e^{tA_c} C = \int_0^t e^{(t-r)A_c} \Delta e^{rA_f}\, dr.}$$

In particular, $\Delta = 0$ implies exact semigroup intertwining at every time.

Proof. Set

$$F(r) = e^{(t-r)A_c} C e^{rA_f}.$$

Then

$$F'(r) = e^{(t-r)A_c}(CA_f - A_c C)e^{rA_f}.$$

Integrating from 0 to t gives $F(t) - F(0) = Ce^{tA_f} - e^{tA_c}C$ and proves the formula. □

Corollary 3.2 (Norm bound)**.** *Suppose*

$$\|e^{tA_f}\| \le M_f e^{\omega_f t}, \qquad \|e^{tA_c}\| \le M_c e^{\omega_c t}.$$

Then

$$\|\mathfrak{D}(t)\| \le M_f M_c \|\Delta\| \int_0^t e^{\omega_c(t-r)} e^{\omega_f r}\, dr.$$

For contraction semigroups,

$$\boxed{\|\mathfrak{D}(t)\| \le t\,\|\Delta\|\,.}$$

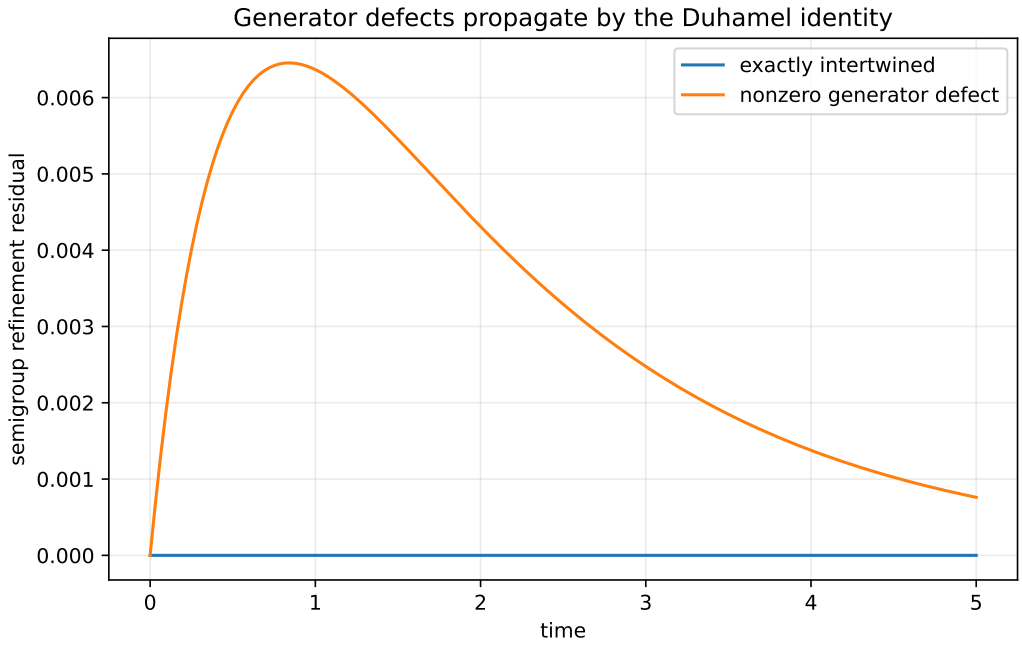


Figure 4. Exact generator intertwining gives a vanishing semigroup defect. A controlled perturbation produces a nonzero defect propagated by the Duhamel identity.

3.1 Resolvent form

For z in the common resolvent set, let

$$R_f(z) = (zI - A_f)^{-1}, \qquad R_c(z) = (zI - A_c)^{-1}.$$

Theorem 3.3 (Resolvent refinement identity)**.**

$$\boxed{CR_f(z) - R_c(z)C = R_c(z)\Delta R_f(z).}$$

Proof. Multiply the left side by $zI - A_c$ on the left and $zI - A_f$ on the right. The result is

$$(zI - A_c)C - C(zI - A_f) = CA_f - A_cC = \Delta.$$

Multiplication by the two resolvents recovers the stated identity. □

4 Markov seeds and stochastic transfer

Use column probability vectors. A finite continuous-time Markov generator Q has nonnegative off-diagonal entries and column sums zero. The stochastic semigroup is

$$K(t) = e^{tQ}.$$

Theorem 4.1 (Finite Markov semigroup). *If Q has nonnegative off-diagonal entries and column sums zero, then e^{tQ} has nonnegative entries and column sums one for every $t \geq 0$.*

Proof. Choose $\lambda \geq \max_i(-Q_{ii})$. Then $B = Q + \lambda I$ has nonnegative entries, so e^{tB} is entrywise nonnegative by its power series. Hence

$$e^{tQ} = e^{-\lambda t}e^{tB}$$

is nonnegative. Since $1^TQ = 0$, differentiation gives

$$\frac{d}{dt}1^Te^{tQ} = 1^TQe^{tQ} = 0,$$

so $1^Te^{tQ} = 1^T$. □

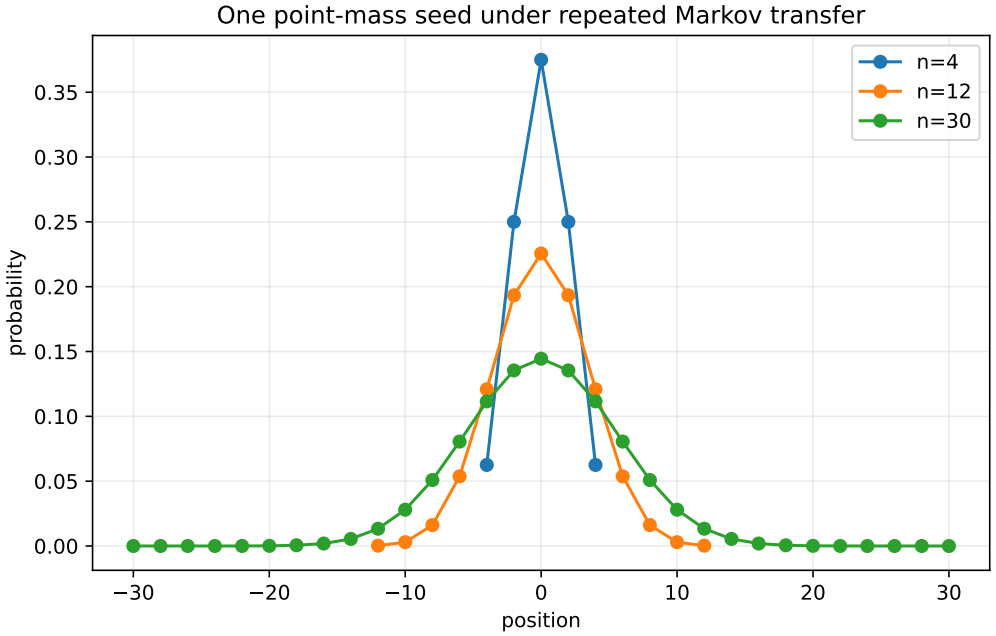


Figure 5. The point mass δ_0 propagated by repeated symmetric random-walk transfer.

Example 4.2 (Symmetric random walk)**.** Starting from δ_0 and applying the transfer

$$T = \frac{1}{2}(S + S^{-1})$$

for n steps gives

$$\mathbb{P}(X_n = k) = 2^{-n}\binom{n}{(n+k)/2}$$

when $k \equiv n \pmod 2$, and zero otherwise. Hence

$$\mathbb{E}[X_n] = 0, \qquad \mathrm{Var}(X_n) = n.$$

The parity packet is exact: all mass remains in the one residue class permitted by n.

5 Markov lumping as exact refinement

Let K be a column-stochastic fine transition matrix, let C aggregate fine states into coarse packets, and let $\bar{K}$ be a coarse transition matrix. Define

$$\Delta = CK - \bar{K}C.$$

Theorem 5.1 (Exact lumpability)**.** *If $\Delta = 0$, then*

$$\boxed{CK^n = \bar{K}^n C}$$

for every $n \geq 0$. Thus fine evolution followed by coarse reconstruction equals coarse evolution from the beginning.

Proof. The result is immediate for $n = 0$. If $CK^n = \bar{K}^n C$, then

$$CK^{n+1} = \bar{K}^n CK = \bar{K}^{n+1} C.$$

□

Theorem 5.2 (Finite-horizon defect formula)**.** *For arbitrary $\Delta = CK - \bar{K}C$,*

$$\boxed{CK^n - \bar{K}^n C = \sum_{j=0}^{n-1} \bar{K}^{n-1-j} \Delta K^j.}$$

If K and $\bar{K}$ are contractions in the chosen norm, then

$$\left\| CK^n - \bar{K}^n C \right\| \leq n \left\| \Delta \right\|.$$

Proof. The identity follows by telescoping:

$$\bar{K}^{n-1-j} CK^{j+1} - \bar{K}^{n-j} CK^j = \bar{K}^{n-1-j}(CK - \bar{K}C)K^j.$$

Summing from $j = 0$ to $n - 1$ leaves $CK^n - \bar{K}^n C$. The norm bound follows termwise. □

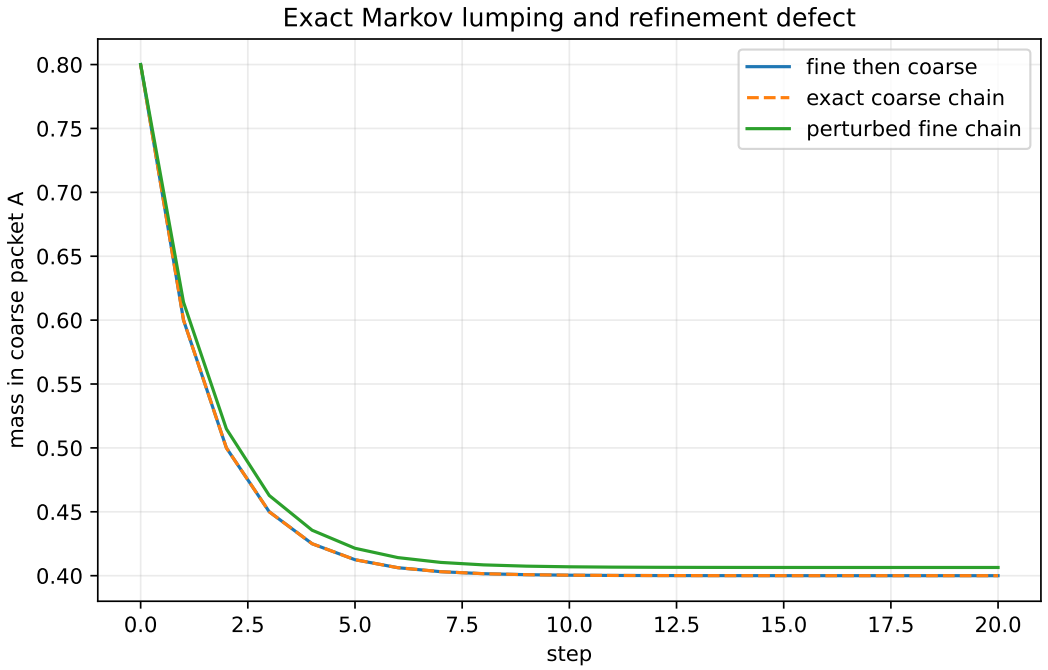


Figure 6. Exact lumping agrees at every horizon. A small generator-level or one-step defect creates a controlled coarse discrepancy.

6 Conditional expectation as a stochastic projector

A route through this section

Conditional expectation is a projector in a weighted Hilbert space. Nested partitions therefore split information into orthogonal increments. Keep the underlying measure fixed while reading the projector identities. A change of measure is a different operation.

Let $\Omega = \{1, \dots, n\}$ carry strictly positive probabilities π_i, and let $\mathcal{Q} = \{B_1, \dots, B_m\}$ be a partition. Define

$$(P_{\mathcal{Q}} f)_i = \frac{\sum_{j \in B(i)} \pi_j f_j}{\sum_{j \in B(i)} \pi_j},$$

where $B(i)$ is the block containing i.

Theorem 6.1 (Finite conditional-expectation projector)**.** *With the weighted inner product*

$$\langle f, g \rangle_\pi = \sum_i \pi_i f_i \overline{g_i},$$

the operator $P_{\mathcal{Q}}$ satisfies

$$P_{\mathcal{Q}}^2 = P_{\mathcal{Q}}, \qquad \langle P_{\mathcal{Q}} f, g \rangle_\pi = \langle f, P_{\mathcal{Q}} g \rangle_\pi.$$

Its range is the space of functions constant on each partition block.

Proof. After one application, the result is constant on every block, so applying the same block average again changes nothing. Weighted self-adjointness follows by expanding both inner products block by

block; both equal

$$\sum_{B\in\mathcal{Q}} \frac{1}{\pi(B)} \left(\sum_{i\in B} \pi_i f_i\right) \overline{\left(\sum_{j\in B} \pi_j g_j\right)}.$$

□

Theorem 6.2 (Tower law and martingale packets). *If $\mathcal{Q}_c$ is coarser than $\mathcal{Q}_f$, then*

$$\boxed{P_c P_f = P_f P_c = P_c.}$$

For a nested tower $P_0 \le P_1 \le \cdots \le P_r$, define

$$D_0 = P_0, \qquad D_j = P_j - P_{j-1}.$$

Then

$$D_i D_j = \delta_{ij} D_j, \qquad P_r = \sum_{j=0}^{r} D_j.$$

Proof. A coarse-block average of a fine-block average is the coarse-block average, and a fine-block average of a coarse-constant function leaves it unchanged. The difference identities follow from $P_i P_j = P_{\min(i,j)}$. □

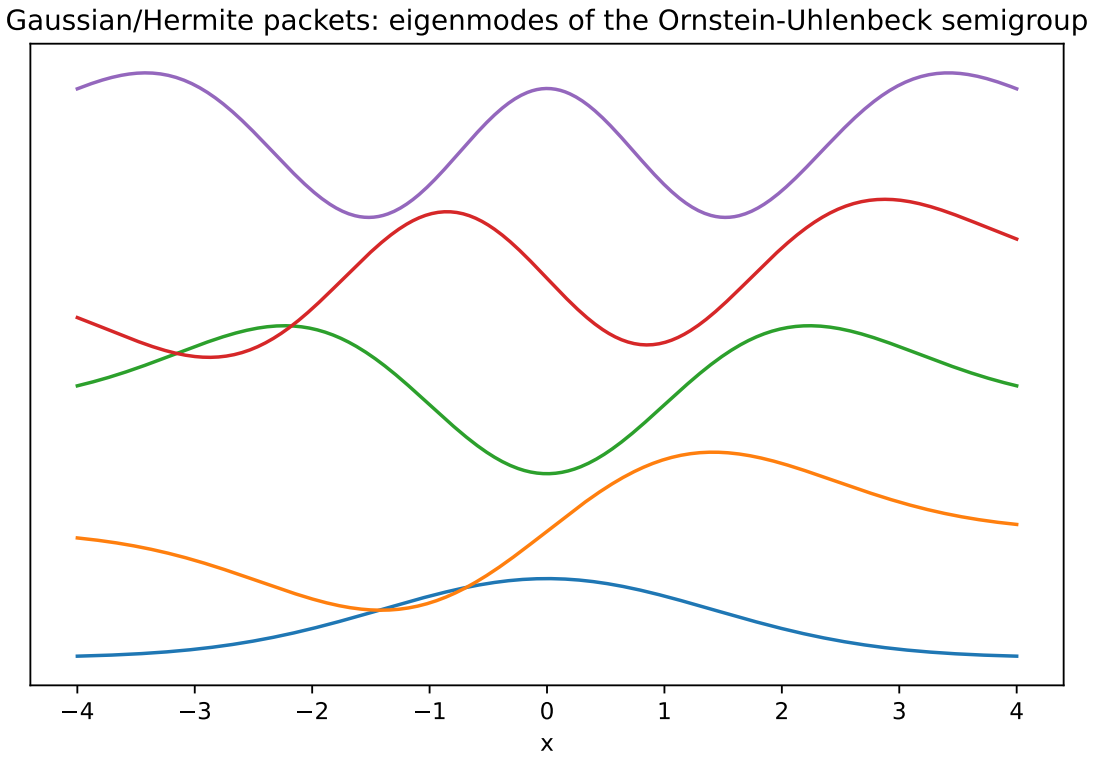


Figure 7. Hermite–Gaussian packets, which diagonalize the Ornstein–Uhlenbeck semigroup. The finite numerical release verifies their Gaussian orthogonality through quadrature. The five displayed modes run from n = 0 at the bottom to n = 4 at the top. Vertical offsets separate the curves.

Classical Input 6.3 (Gaussian chaos realization). For the standard Gaussian measure, the normalized probabilists' Hermite polynomials form orthonormal chaos packets. The Ornstein–Uhlenbeck semigroup acts diagonally:

$$T_t H_n = e^{-nt} H_n.$$

Thus conditional expectation, martingale differences, and Gaussian chaos are all projector realizations of stochastic refinement.

7 Bayesian conditioning as a normalized nonlinear operator

Let $p = (p_i)$ be a probability vector and let $\lambda = (\lambda_i)$ be a nonnegative likelihood with positive evidence

$$Z_\lambda(p) = \sum_i \lambda_i p_i.$$

Define

$$\boxed{\mathcal{B}_\lambda(p) = \frac{\lambda \odot p}{Z_\lambda(p)}.}$$

This map is nonlinear solely because of its scalar reconstruction factor.

Theorem 7.1 (Independent-likelihood composition)**.** *For nonnegative likelihoods λ and μ, assume $Z_\lambda(p) > 0$ and $Z_{\lambda\odot\mu}(p) > 0$. Equivalently, both successive updates have positive evidence. Then*

$$\boxed{\mathcal{B}_\mu(\mathcal{B}_\lambda(p)) = \mathcal{B}_{\lambda\odot\mu}(p).}$$

Proof. The first normalization appears in both numerator and denominator of the second update and cancels:

$$\frac{\mu_i(\lambda_i p_i / Z_\lambda)}{\sum_j \mu_j \lambda_j p_j / Z_\lambda} = \frac{\mu_i \lambda_i p_i}{\sum_j \mu_j \lambda_j p_j}.$$

□

7.1 Packet evidence and reconstruction

Let the state indices be partitioned into packets I_α. Define

$$Z_\alpha = \sum_{i \in I_\alpha} \lambda_i p_i, \qquad m_\alpha = \frac{Z_\alpha}{\sum_\beta Z_\beta}, \qquad q_{\alpha,i} = \begin{cases} \lambda_i p_i / Z_\alpha, & Z_\alpha > 0,\ i \in I_\alpha, \\ 0, & \text{otherwise.} \end{cases}$$

Theorem 7.2 (Bayesian packet reconstruction)**.**

$$\boxed{\mathcal{B}_\lambda(p) = \sum_\alpha m_\alpha q_\alpha.}$$

Only packets with positive evidence carry a conditional probability vector. A zero-evidence packet has zero posterior mass and contributes the zero vector. The nonlinear coupling between packets occurs only through the total evidence used to normalize the packet evidences.

Proof. If $Z_\alpha = 0$, nonnegativity implies $\lambda_i p_i = 0$ for every $i \in I_\alpha$, so the reconstruction identity holds with both sides zero. For $Z_\alpha > 0$ and $i \in I_\alpha$,

$$m_\alpha q_{\alpha,i} = \frac{Z_\alpha}{Z} \frac{\lambda_i p_i}{Z_\alpha} = \frac{\lambda_i p_i}{Z}.$$

The packet supports are disjoint, so summing reconstructs the full posterior. □

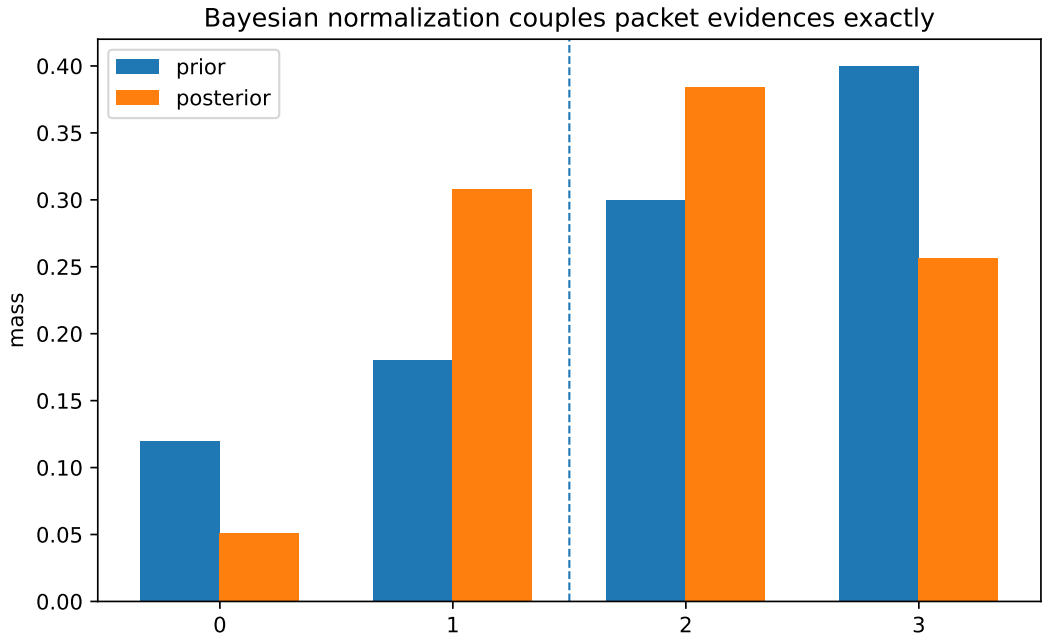


Figure 8. A finite prior and posterior split into two state packets. Evidence changes both the local packet shape and the global packet mass.

Theorem 7.3 (Likelihood perturbation bound). *Let p be fixed and let $\lambda, \mu \geq 0$ have positive evidences Z_λ, Z_μ. Then*

$$\boxed{\left\|\mathcal{B}_\lambda(p) - \mathcal{B}_\mu(p)\right\|_1 \leq \frac{2\sum_i |\lambda_i - \mu_i| p_i}{\min(Z_\lambda, Z_\mu)}.}$$

Proof. Add and subtract $\mu \odot p / Z_\lambda$ and use the triangle inequality:

$$\left\|\frac{\lambda \odot p}{Z_\lambda} - \frac{\mu \odot p}{Z_\mu}\right\|_1 \leq \frac{\|(\lambda - \mu) \odot p\|_1}{Z_\lambda} + \|\mu \odot p\|_1 \left|\frac{1}{Z_\lambda} - \frac{1}{Z_\mu}\right|.$$

The second term equals $|Z_\lambda - Z_\mu| / Z_\lambda$ and is bounded by the same weighted L^1 difference divided by Z_λ. Interchanging λ and μ gives the minimum-evidence form. □

8 Exact Bayesian refinement and information-loss packets

How to read the result

The all-prior criterion here tests coarse posterior masses for every prior and assumes at least two coarse blocks. Part II uses a stronger convention that also preserves each conditional distribution inside a block. These are different contracts, so the one-block boundary must be read with its own definition.

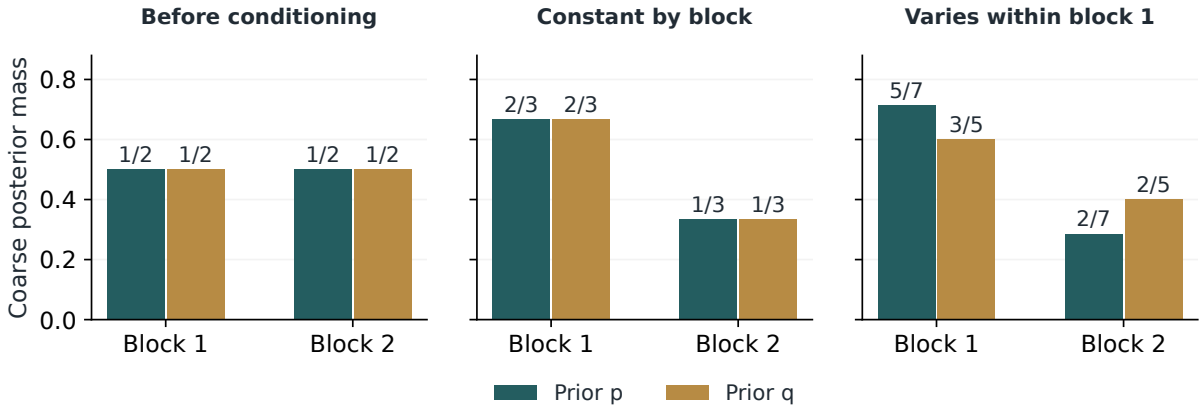


Figure 9. Two rational priors with the same two coarse masses. A blockwise constant likelihood gives the same coarse update, while a likelihood varying within a block separates the posteriors. The exact fractions are calculated before plotting.

Let the fine state set I be partitioned into nonempty coarse packets I_a, $a \in A$. The coarsening map on probability vectors is

$$(C_\pi p)_a = P_a = \sum_{i \in I_a} p_i.$$

For a fine likelihood $\lambda = (\lambda_i)_{i \in I}$ and a prior p, define the effective coarse likelihood

$$r_a(p) = \frac{\sum_{i \in I_a} \lambda_i p_i}{P_a} \qquad (P_a > 0).$$

Its value on a zero-mass packet is irrelevant and may be set to zero.

Theorem 8.1 (Exact effective-likelihood factorization)**.** *Whenever the fine evidence is positive,*

$$\boxed{C_\pi \mathcal{B}_\lambda(p) = \mathcal{B}_{r(p)}(C_\pi p).}$$

Thus the only fine information needed for the coarse posterior, beyond the coarse prior masses, is the vector of conditional packet evidences $r(p)$.

Proof. For $P = C_\pi p$ and $Z = \sum_i \lambda_i p_i = \sum_a P_a r_a(p)$,

$$\left(C_\pi \mathcal{B}_\lambda(p)\right)_a = \frac{\sum_{i \in I_a} \lambda_i p_i}{Z} = \frac{P_a r_a(p)}{\sum_b P_b r_b(p)} = \left(\mathcal{B}_{r(p)}(P)\right)_a.$$

□

The next result gives the exact necessary and sufficient condition for a coarse likelihood to work for every prior.

Theorem 8.2 (Bayesian refinement criterion)**.** *Assume that the partition contains at least two coarse packets and that the fine and coarse likelihoods are strictly positive. The identity*

$$\boxed{C_\pi \mathcal{B}_\lambda(p) = \mathcal{B}_{\bar\lambda}(C_\pi p)}$$

holds for every fine prior p if and only if there exists one scalar $c > 0$ such that

$$\boxed{\lambda_i = c\,\bar\lambda_a \qquad (i \in I_a).}$$

Equivalently, the likelihood is constant on each coarse packet, with the coarse values determined up to the globally invisible Bayesian scaling c.

Proof. If $\lambda_i = c\bar\lambda_a$ on I_a, then $P_a r_a(p) = c\bar\lambda_a P_a$ for every prior, including zero-mass packets, and the preceding theorem gives the desired identity because Bayesian normalization removes the common scalar c.

Conversely, fix two fine states $i, i' \in I_a$ and one state $j \in I_b$ in another coarse packet. Vary the distribution of a fixed coarse mass inside I_a between i and i' while leaving every coarse packet mass unchanged. The right side depends only on those coarse masses and is therefore unchanged. The left-side posterior mass of packet a depends on the conditional average of λ_i and $\lambda_{i'}$. It can be independent of that variation only if $\lambda_i = \lambda_{i'}$. Hence λ is constant on every packet; write the value on I_a as λ_a.

Now use priors supported on one state of packet a and one state of packet b. Equality of the two coarse posterior odds gives

$$\frac{\lambda_a}{\lambda_b} = \frac{\bar\lambda_a}{\bar\lambda_b}.$$

This holds for every pair of coarse packets, so $\lambda_a = c\bar\lambda_a$ for one common $c > 0$. □

Remark 8.3 (The one-packet boundary)*.* If the coarse partition has only one packet, every posterior coarsens to the scalar 1, so commutation is automatic and no within-packet likelihood information can be identified. The nontrivial criterion begins with at least two coarse packets.

When sufficiency fails, the defect still has an exact finite factorization. Fix a declared coarse likelihood $\bar\lambda$, put

$$P = C_\pi p, \qquad \delta(p) = r(p) - \bar\lambda, \qquad \bar b = \mathcal{B}_{\bar\lambda}(P), \qquad Z_r = \langle r(p), P\rangle.$$

Theorem 8.4 (Bayesian information-loss packet)**.** *Assume nonnegative fine and declared coarse likelihoods, with both $Z_r > 0$ and $Z_{\bar\lambda} > 0$. The coarsening defect is*

$$\boxed{C_\pi \mathcal{B}_\lambda(p) - \mathcal{B}_{\bar\lambda}(P) = \frac{1}{Z_r}\left(\operatorname{diag}(P) - \bar b P^{\mathsf{T}}\right)\delta(p).}$$

The factor matrix annihilates $\bar\lambda$. This records the common-scale null direction at the declared coarse likelihood; the full exact defect remains homogeneous under simultaneous likelihood rescaling because the numerator

and evidence denominator scale together. Moreover,

$$\boxed{\left\|C_\pi \mathcal{B}_\lambda(p) - \mathcal{B}_{\bar{\lambda}}(P)\right\|_1 \le \frac{2\sum_a P_a|\delta_a(p)|}{\min\{Z_r, Z_{\bar{\lambda}}\}}}$$

and

$$\sum_a P_a|\delta_a(p)| \le \sum_i p_i \left|\lambda_i - \bar{\lambda}_{\pi(i)}\right|.$$

Proof. The effective-likelihood theorem reduces the defect to

$$\mathcal{B}_{\bar{\lambda}+\delta}(P) - \mathcal{B}_{\bar{\lambda}}(P).$$

Since $Z_r = Z_{\bar{\lambda}} + P^\mathsf{T}\delta$, direct subtraction gives

$$\frac{1}{Z_r}\left(\operatorname{diag}(P) - \bar{b}P^\mathsf{T}\right)\delta.$$

Also,

$$\left(\operatorname{diag}(P) - \bar{b}P^\mathsf{T}\right)\bar{\lambda} = P \odot \bar{\lambda} - \bar{b}Z_{\bar{\lambda}} = 0.$$

The first norm bound is the likelihood perturbation theorem applied on the coarse simplex. For a zero-mass packet, both sides of the following estimate vanish. For $P_a > 0$,

$$P_a|\delta_a| = P_a\left|\sum_{i\in I_a}\frac{p_i}{P_a}(\lambda_i - \bar{\lambda}_a)\right| \le \sum_{i\in I_a} p_i|\lambda_i - \bar{\lambda}_a|,$$

and summing over a proves the second inequality. □

Corollary 8.5 (Nested Bayesian packet tower)**.** *Consider a nested finite packet tower with strictly positive likelihoods at every level. At each adjacent coarsening with at least two coarse packets, commutation for every prior is equivalent to the fine likelihood being blockwise proportional to the declared coarse likelihood, with one common positive proportionality factor. An adjacent coarsening to a single packet commutes automatically and imposes no such condition. Under these conditions, all composite coarsening squares commute.*

Proof. At each adjacent coarsening with at least two coarse packets, apply the refinement criterion. For a single coarse packet, both normalized posteriors equal 1. Composition of these commuting squares gives the full tower statement. □

Corollary 8.6 (Canonical coarsest likelihood-resolving partition)**.** *For a finite family of likelihoods* $\{\lambda^{(y)}\}_{y\in Y}$*, define*

$$i \sim j \quad\Longleftrightarrow\quad \lambda_i^{(y)} = \lambda_j^{(y)} \text{ for every } y \in Y.$$

The equivalence classes form the unique coarsest likelihood-resolving partition: every likelihood is constant on each packet. Every update therefore closes on this partition whenever its evidence is positive. If all likelihoods are strictly positive and there are at least two profile classes, this is also the unique coarsest nontrivial partition on which every update closes for every prior. The one-packet partition always closes, independently of the profiles, and is excluded from this last assertion.

Proof. Equality of likelihood profiles is an equivalence relation. Its classes are likelihood-resolving by definition, and every likelihood-resolving partition refines them. Exact closure follows by summing

the weighted masses in each fiber, including zero-mass fibers. Under strict positivity, the refinement criterion implies that every nontrivial all-prior closed partition is likelihood-resolving. When every profile is identical, all partitions close; there need not be a unique coarsest nontrivial partition. □

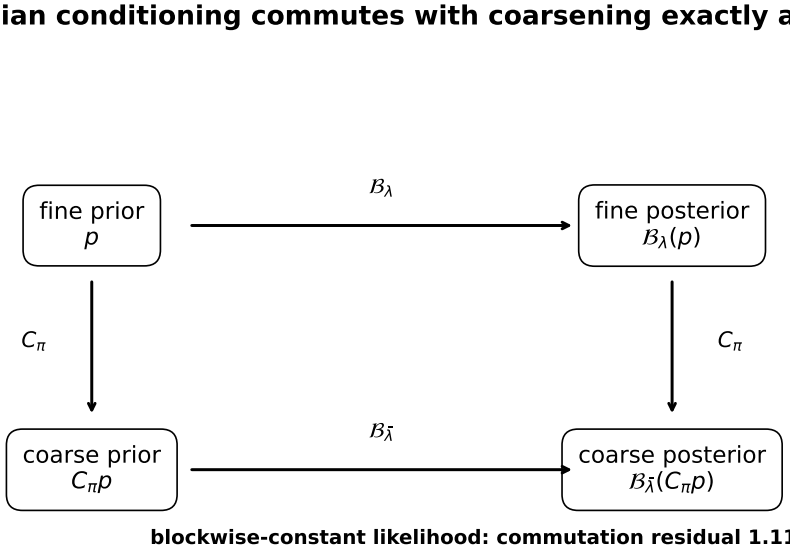


Figure 10. For strictly positive likelihoods and at least two coarse packets, the Bayesian refinement square commutes for every prior exactly when the likelihood is constant on each coarsening fiber, up to one global scale.

Numerical Certificate 8.7 (Bayesian refinement). For the finite test partition $\{0,1\} \sqcup \{2,3\}$, a blockwise-constant likelihood gave commutation residual 1.11×10^{-16}. A nonconstant likelihood produced an exact defect-factorization residual 7.29×10^{-17}. Its measured L^1 defect was 4.8787×10^{-2}, below the certified bound 1.1016×10^{-1}. Two priors with identical coarse masses but different hidden within-packet shapes produced coarse posteriors differing by 1.2400×10^{-1}, which is a direct necessity witness.

Boundary 8.8 (Finite exact theorem). The conjectured Bayesian refinement law is therefore resolved for finite packet systems and arbitrary priors. Infinite-dimensional disintegration, posterior consistency, and stability of data-dependent partitions require their own measure-theoretic and asymptotic hypotheses.

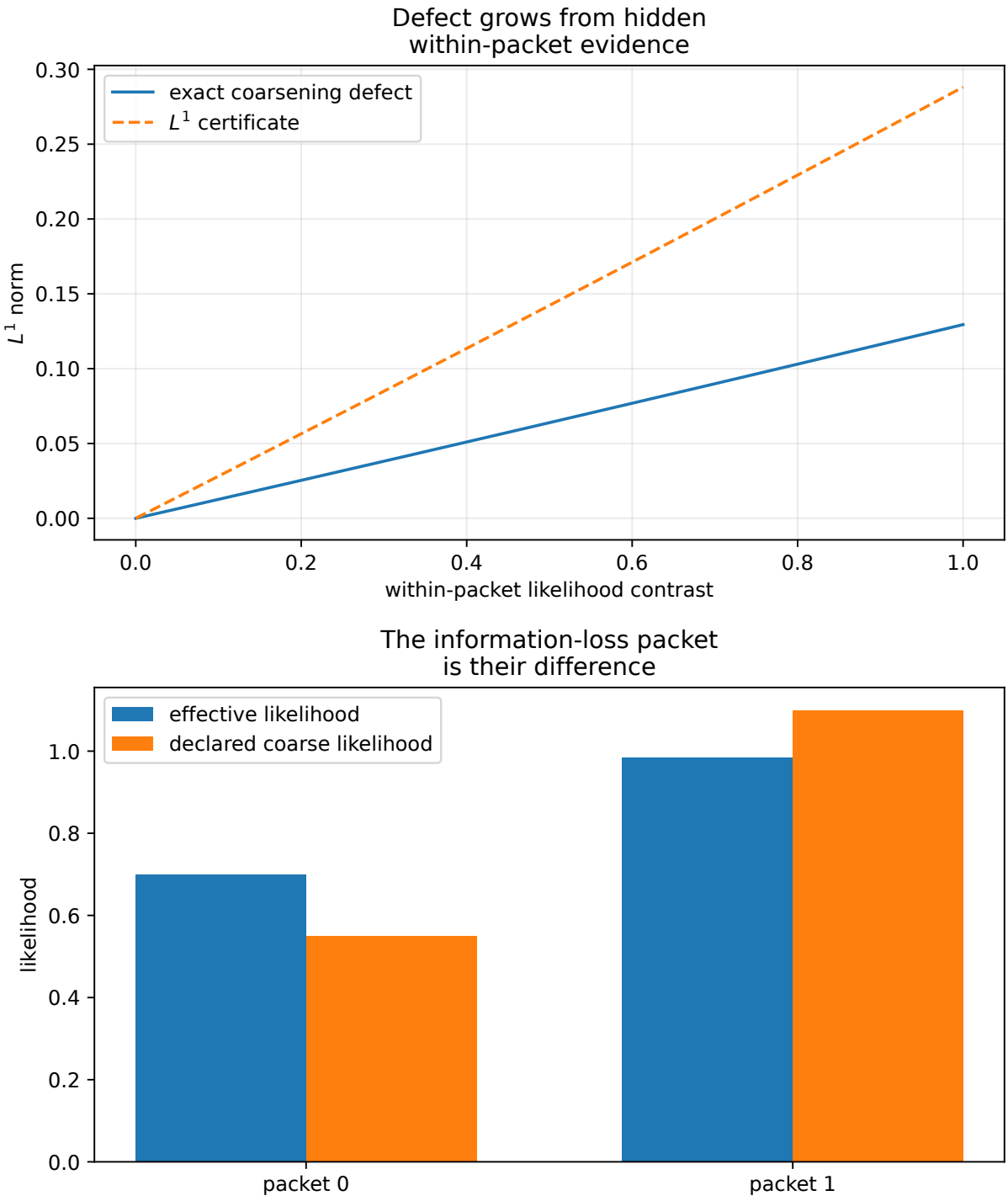


Figure 11. When packet sufficiency fails, the exact defect factors through the difference between the conditional effective likelihood and the declared coarse likelihood.

9 Bayesian flows, conjugacy, and hidden states

Fix a score vector g. The likelihood flow is

$$\lambda_i(t) = e^{tg_i},$$

and the Bayesian flow is

$$\Phi_t(p)_i = \frac{e^{tg_i}p_i}{\sum_j e^{tg_j}p_j}.$$

Theorem 9.1 (Bayesian nonlinear semigroup)**.**

$$\boxed{\Phi_{s+t} = \Phi_s \circ \Phi_t.}$$

Its generator is the centered multiplicative field

$$\boxed{\frac{d}{dt}p_i(t) = (g_i - \langle g, p(t)\rangle)p_i(t).}$$

Proof. The semigroup law is the likelihood-composition theorem with $e^{sg}e^{tg} = e^{(s+t)g}$. Differentiating the normalized exponential formula gives

$$\dot{p}_i = g_i p_i - p_i \frac{\sum_j g_j e^{tg_j}p_j(0)}{\sum_j e^{tg_j}p_j(0)}.$$

The quotient is $\langle g, p(t)\rangle$. □

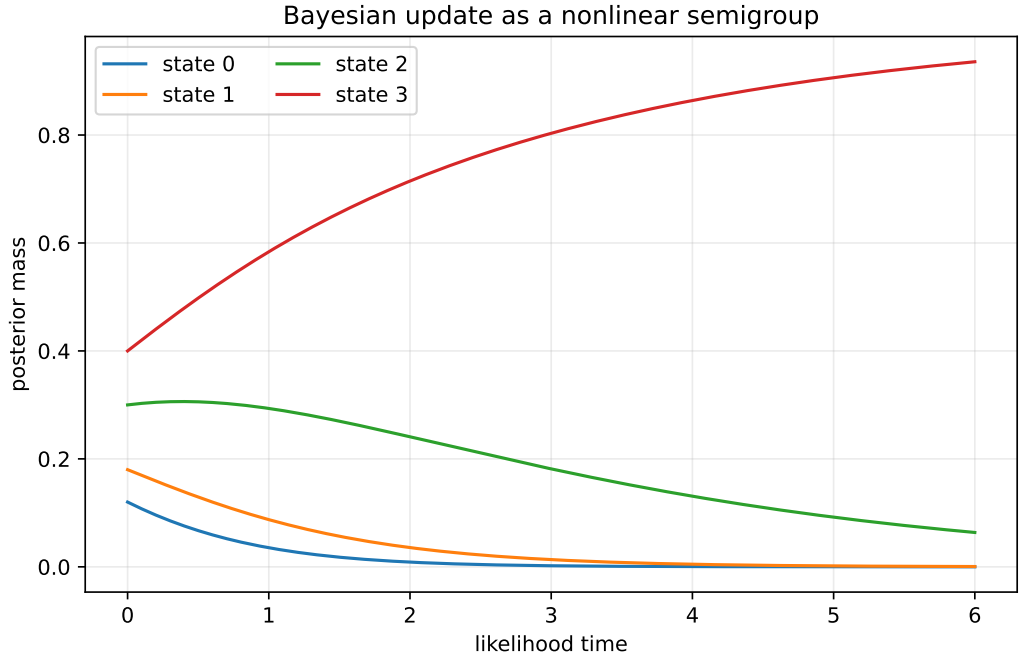


Figure 12. Continuous likelihood accumulation produces a nonlinear semigroup on the probability simplex.

Example 9.2 (Beta–Binomial conjugate packet)**.** A Beta(a, b) prior and s successes in n trials produce

$$\text{Beta}(a + s, b + n - s).$$

The prior exponents are additive packets: the likelihood increments them by $(s, n-s)$. In the numerical example,

$$(a, b) = (2, 3), \quad (s, n) = (7, 10),$$

so the posterior is Beta$(9, 6)$ with mean 0.6.

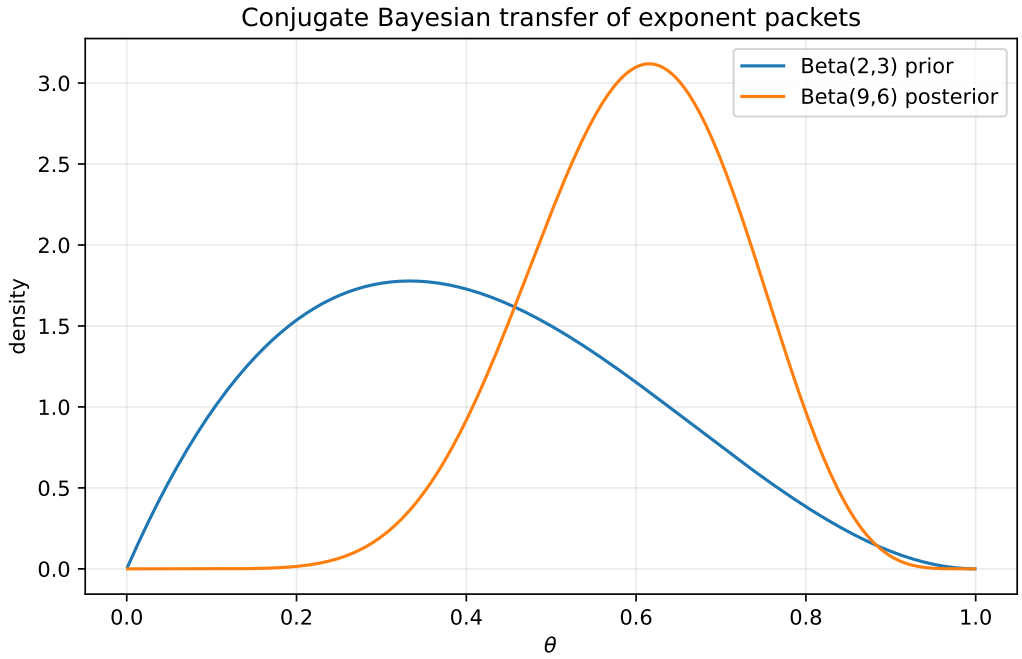


Figure 13. A conjugate Bayesian update viewed as exact transfer of exponent packets.

9.1 Hidden Markov filtering

Let K be column stochastic and let $\lambda^{(r)}$ be the likelihood associated with the rth observation. The filter recursion is

$$\hat{p}_r = K p_{r-1}, \qquad p_r = \mathcal{B}_{\lambda^{(r)}}(\hat{p}_r).$$

Proposition 9.3 (Prediction-update reconstruction)**.** *Every hidden Markov filtering step is the composition of a linear Markov transfer and a normalized nonlinear likelihood transfer. Assume a probability initial vector, a column-stochastic transition matrix, and nonnegative likelihoods with positive evidence at every step. Then every p_r is a probability vector.*

Proof. Column stochasticity gives $\hat{p}_r \geq 0$ and $1^T \hat{p}_r = 1$. Bayesian normalization then gives $p_r \geq 0$ and $1^T p_r = 1$. □

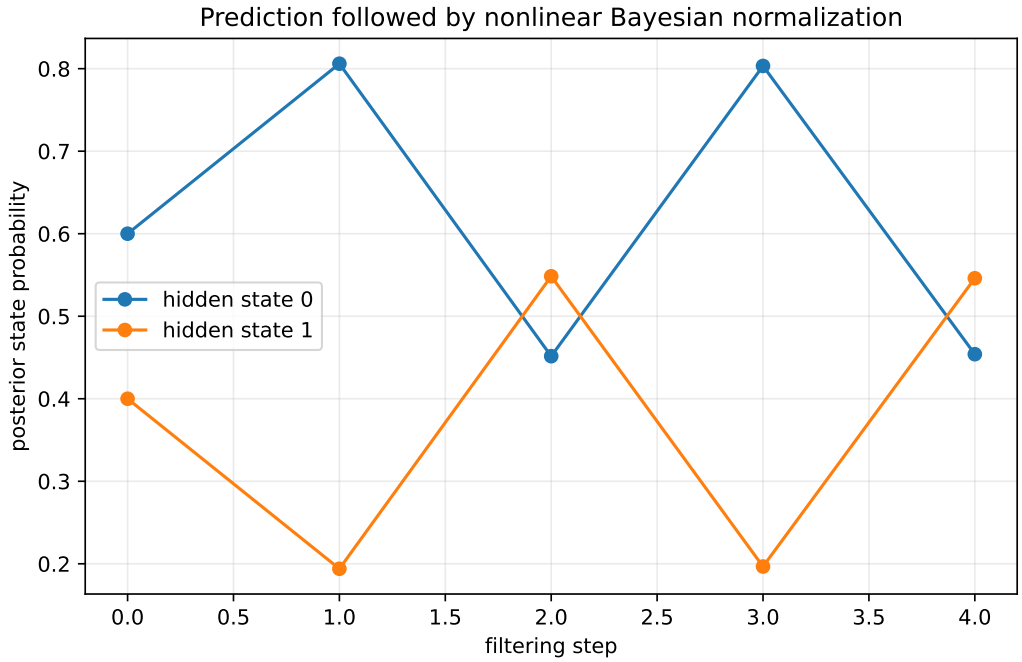


Figure 14. A hidden Markov filter alternates a linear prediction semigroup with nonlinear Bayesian normalization.

10 Fractional powers as spectral packet transfer

A route through this section

First identify the positive operator and its spectral packets. Then distinguish an exact spectral identity from a bound for a refinement mismatch. The gap-free certificate and the common-gap estimate have different hypotheses and different strengths.

Let $L = L^* \geq 0$ on a finite-dimensional Hilbert space. Its spectral resolution is

$$L = \sum_j \lambda_j P_j.$$

For $\alpha > 0$, define

$$\boxed{L^\alpha = \sum_j \lambda_j^\alpha P_j.}$$

The associated fractional semigroup is

$$T_\alpha(t) = e^{-tL^\alpha} = \sum_j e^{-t\lambda_j^\alpha} P_j.$$

Theorem 10.1 (Fractional packet calculus). *For $\alpha, \beta > 0$,*

$$L^\alpha L^\beta = L^{\alpha+\beta},$$

and for $s, t \geq 0$,

$$\boxed{T_\alpha(s)T_\alpha(t) = T_\alpha(s+t).}$$

Every spectral packet P_j reduces L^α and $T_\alpha(t)$.

Proof. All statements hold coefficientwise on the mutually orthogonal spectral projectors. For example,

$$L^\alpha L^\beta = \sum_{j,k} \lambda_j^\alpha \lambda_k^\beta P_j P_k = \sum_j \lambda_j^{\alpha+\beta} P_j.$$

□

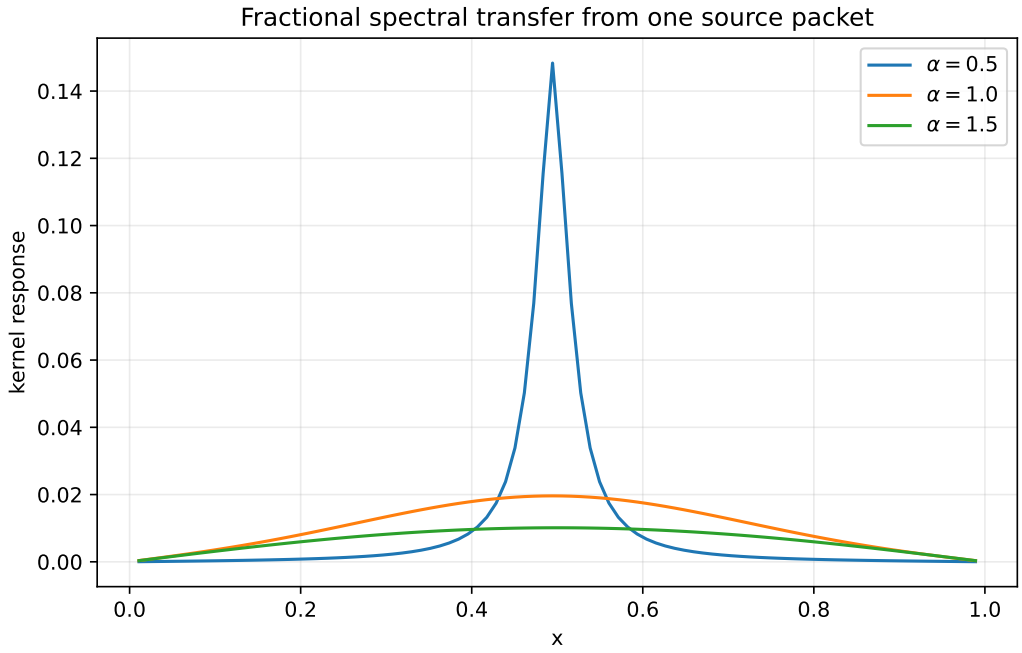


Figure 15. Different spectral exponents produce different transfer profiles from the same point seed.

10.1 Square-root subordination

Theorem 10.2 (Poisson subordination). *For $t > 0$ and $L \geq 0$,*

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

The integral converges in operator norm in finite dimension.

Proof. By the spectral theorem it suffices to prove the scalar identity

$$e^{-t\sqrt{\lambda}} = \frac{t}{2\sqrt{\pi}} \int_0^\infty s^{-3/2} e^{-t^2/(4s) - \lambda s}\, ds$$

for $\lambda \geq 0$. This is the classical Gaussian integral after the substitution $u = t/(2\sqrt{s})$. Applying the identity to every eigenvalue and recombining the spectral projectors proves the operator formula. □

10.2 Fractional naturality under a bounded refinement defect

Let $A \geq 0$ act on a fine Hilbert space, let $B \geq 0$ act on a coarse Hilbert space, and let C be a bounded coarsening map. Set

$$D = CA - BC, \qquad D_\alpha = CA^\alpha - B^\alpha C, \qquad 0 < \alpha < 1.$$

Theorem 10.3 (Fractional intertwining theorem)**.** *For bounded nonnegative self-adjoint A and B,*

$$\boxed{D_\alpha = \frac{\sin(\pi\alpha)}{\pi} \int_0^\infty t^\alpha (tI + B)^{-1} D (tI + A)^{-1}\, dt.}$$

The improper integral converges in operator norm. Moreover,

$$\boxed{\|D_\alpha\| \leq K_\alpha \|C\|^{1-\alpha} \|D\|^\alpha, \qquad K_\alpha = \frac{2^{1-\alpha}\sin(\pi\alpha)}{\pi\alpha(1-\alpha)}.}$$

If $A, B \geq mI$ for some $m > 0$, then the stronger linear estimate holds:

$$\boxed{\|D_\alpha\| \leq \alpha m^{\alpha-1} \|D\|.}$$

In particular, exact generator intertwining $CA = BC$ implies exact fractional intertwining $CA^\alpha = B^\alpha C$.

Proof. For $0 < \alpha < 1$, the Balakrishnan formula gives

$$A^\alpha = \frac{\sin(\pi\alpha)}{\pi} \int_0^\infty t^{\alpha-1} A (tI + A)^{-1}\, dt,$$

and likewise for B. Using

$$A(tI + A)^{-1} = I - t(tI + A)^{-1}$$

and the rectangular resolvent identity

$$(tI + B)^{-1}C - C(tI + A)^{-1} = (tI + B)^{-1}(CA - BC)(tI + A)^{-1},$$

the difference of the two Balakrishnan integrals becomes the displayed resolvent-defect formula.

For small t,

$$\left\| CA(tI + A)^{-1} - B(tI + B)^{-1}C \right\| \leq 2\|C\|,$$

because the two positive resolvent factors are contractions. For large t, the resolvent identity gives the bound $\|D\|/t$. Splitting the integral at

$$r = \frac{\|D\|}{2\|C\|}$$

when C and D are nonzero yields

$$\begin{aligned} \|D_\alpha\| &\leq \frac{\sin(\pi\alpha)}{\pi} \left(2\|C\| \int_0^r t^{\alpha-1}\, dt + \|D\| \int_r^\infty t^{\alpha-2}\, dt \right) \\ &= K_\alpha \|C\|^{1-\alpha} \|D\|^\alpha. \end{aligned}$$

The zero cases are immediate.

If $A, B \geq mI$, then both resolvents have norm at most $(t+m)^{-1}$, so

$$\|D_\alpha\| \leq \frac{\sin(\pi\alpha)}{\pi} \|D\| \int_0^\infty \frac{t^\alpha}{(t+m)^2}\, dt.$$

After $t = mu$, the beta integral equals

$$m^{\alpha-1} \frac{\alpha\pi}{\sin(\pi\alpha)},$$

which gives the linear estimate. □

Corollary 10.4 (Fractional semigroup refinement)**.** *For $s \geq 0$,*

$$\boxed{Ce^{-sA^\alpha} - e^{-sB^\alpha}C = -\int_0^s e^{-(s-r)B^\alpha} D_\alpha e^{-rA^\alpha}\, dr.}$$

Hence

$$\left\|Ce^{-sA^\alpha} - e^{-sB^\alpha}C\right\| \leq s \left\|D_\alpha\right\|.$$

Under the common gap $A, B \geq mI$,

$$\boxed{\left\|Ce^{-sA^\alpha} - e^{-sB^\alpha}C\right\| \leq se^{-sm^\alpha} \alpha m^{\alpha-1} \left\|D\right\|.}$$

Proof. Differentiate

$$r \longmapsto e^{-(s-r)B^\alpha} Ce^{-rA^\alpha}$$

and integrate from 0 to s. Positive fractional semigroups are contractions; under the gap, the product of their two norm bounds is e^{-sm^α}. □

Corollary 10.5 (Fractional refinement towers)**.** *Let $D_\ell = C_\ell A_{\ell+1} - A_\ell C_\ell$ in a bounded nonnegative self-adjoint tower, with $\sup_\ell \|C_\ell\| \leq M$. If*

$$\sum_\ell \|D_\ell\|^\alpha < \infty,$$

then the fractional defects

$$C_\ell A_{\ell+1}^\alpha - A_\ell^\alpha C_\ell$$

have summable operator norms. If the tower has a uniform gap $m > 0$, it is enough that $\sum_\ell \|D_\ell\| < \infty$. On every bounded time interval, the series of corresponding fractional-semigroup defect norms converges uniformly. The maps may have different domain and codomain spaces; no operator sum across unrelated spaces is asserted.

Boundary 10.6 (Unbounded and nonnormal generators). The theorem resolves fractional naturality for bounded nonnegative self-adjoint refinement systems and gives the exact estimate needed for finite discretization towers. For unbounded positive generators, the same resolvent argument requires a declared common-domain condition and a bounded extension of $CA - BC$. Sectorial or nonnormal generators require their own contour calculus and resolvent bounds. Those domain questions are not inferred from the finite certificates.

Boundary 10.7 (Non-self-adjoint fractional powers). The spectral definition above is complete for finite self-adjoint positive operators. Sectorial and nonnormal operators require a resolvent or contour functional calculus; the same packet language can be used only after those analytic hypotheses are established.

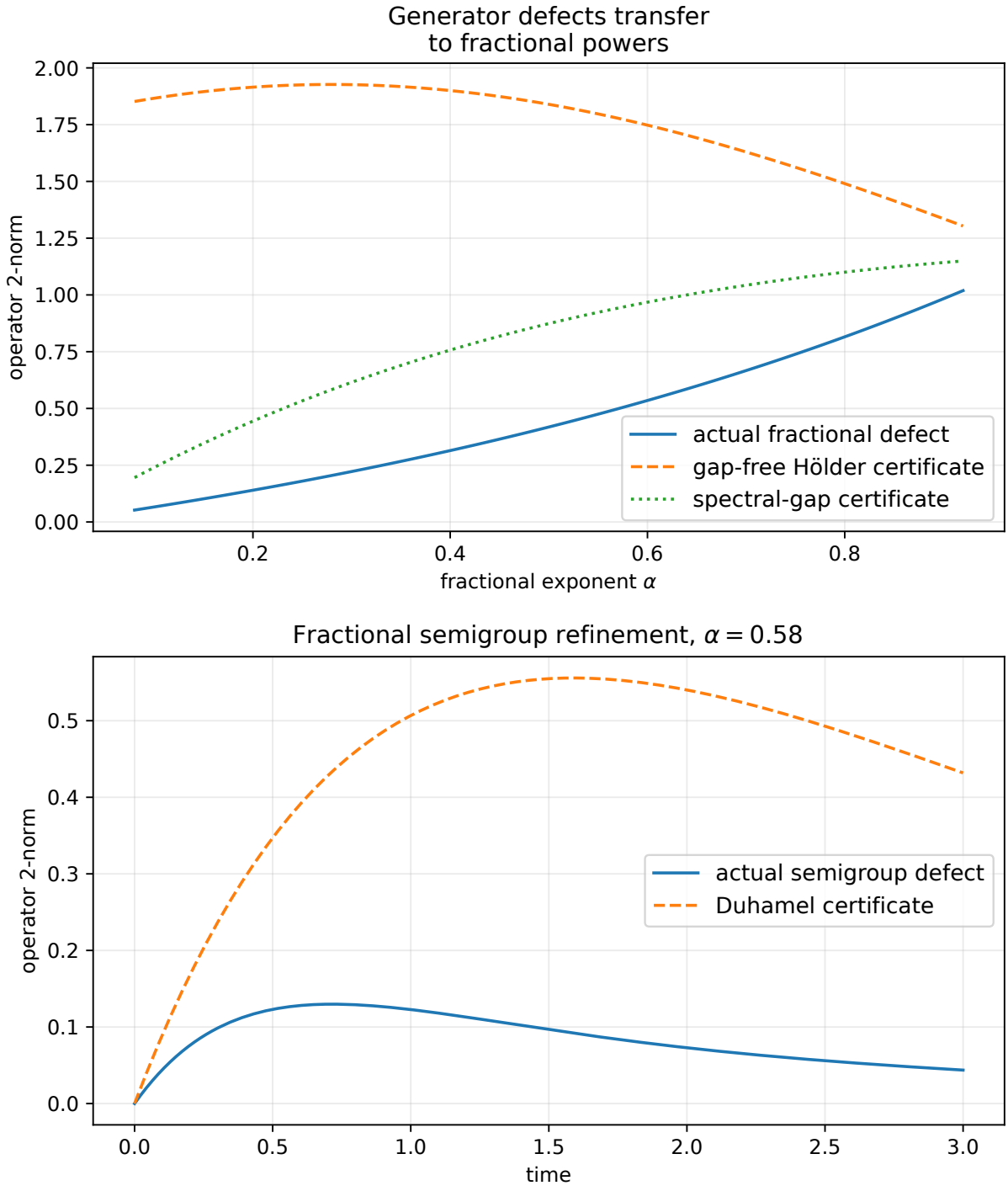


Figure 16. A generator-level refinement defect transfers quantitatively to fractional powers and then to fractional semigroups. The two upper curves are theorem certificates, not fitted error bars.

11 Green operators and the Poisson seed

On the interval $(0,1)$ with Dirichlet boundary conditions, the Green kernel for $-u'' = f$ is

$$\boxed{G(x,y) = \min(x,y) - xy.}$$

Thus the point source δ_y generates the solution kernel, and a general source is reconstructed by

$$u(x) = \int_0^1 G(x,y) f(y)\, dy.$$

For n interior points with spacing $h = 1/(n+1)$, define

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

Theorem 11.1 (Exact discrete Green reconstruction at grid nodes). *Let $x_i = ih$. Then*

$$\boxed{\frac{1}{h}(L_h^{-1})_{ij} = G(x_i, x_j).}$$

Thus the scaled discrete inverse reproduces the continuum Dirichlet Green kernel exactly at the grid nodes.

Proof. The inverse of the unscaled tridiagonal matrix has entries

$$\frac{\min(i,j)(n+1-\max(i,j))}{n+1}.$$

Multiplying by h^2 and then dividing by h gives

$$h \min(i,j) \left(1 - \frac{\max(i,j)}{n+1}\right) = \min(x_i, x_j) - x_i x_j.$$

□

11.1 Low-mode convergence

The kth discrete eigenvalue is

$$\lambda_{k,h} = \frac{4}{h^2} \sin^2\left(\frac{k\pi h}{2}\right),$$

whereas the continuum eigenvalue is $(k\pi)^2$.

Theorem 11.2 (Second-order low-mode convergence). *For fixed k,*

$$\lambda_{k,h} = (k\pi)^2 + O(h^2).$$

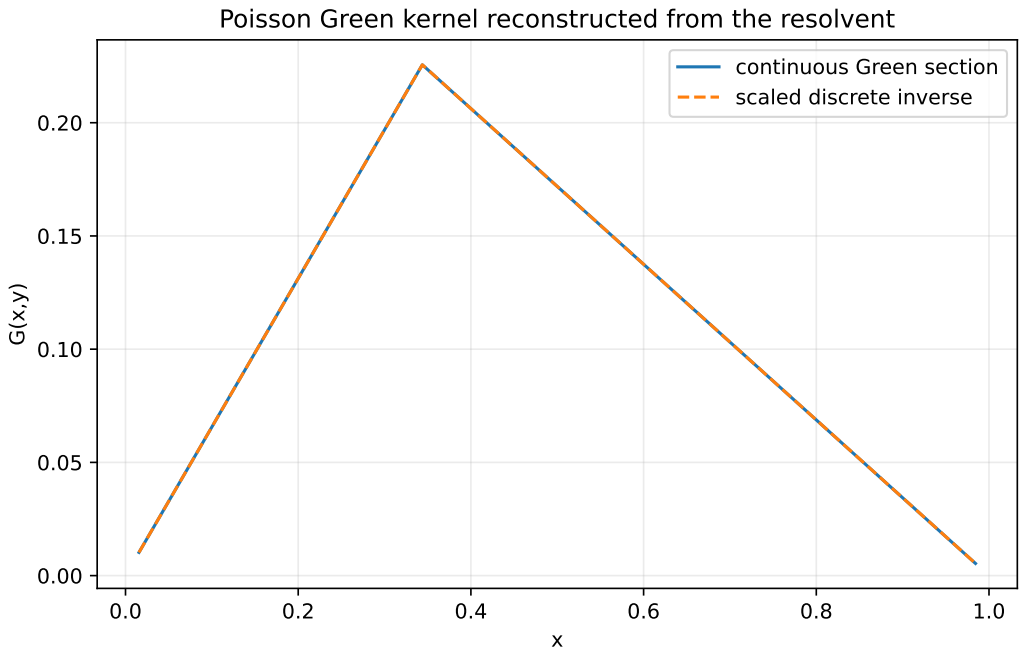


Figure 17. A section of the continuum Green kernel and the scaled discrete inverse. They agree at the grid nodes.

More precisely, for sufficiently small h,

$$0 \le (k\pi)^2 - \lambda_{k,h} \le \frac{(k\pi)^4}{12} h^2.$$

Proof. For $x \ge 0$, $0 \le x^2 - \sin^2 x \le x^4/3$. Set $x = k\pi h/2$ and multiply by $4/h^2$. □

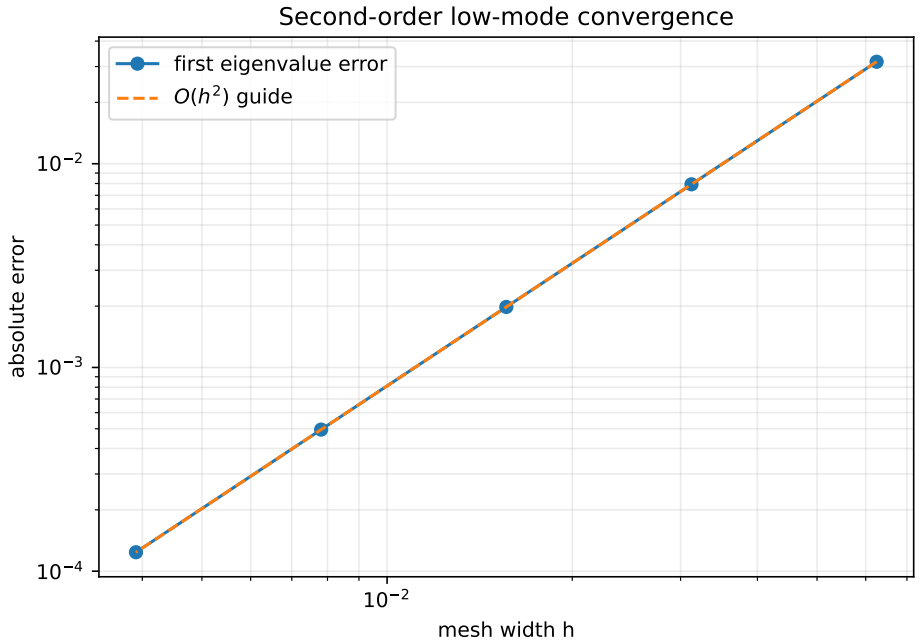


Figure 18. The first finite-difference eigenvalue converges with observed order approximately two.

12 Finite-state Feynman–Kac transfer

Let K be a column-stochastic matrix, let V be a potential on the states, and let f be terminal data. Define

$$M_V = \operatorname{diag}(e^{-V})K^T.$$

The backward Feynman–Kac value after n steps is

$$u_n = M_V^n f.$$

Theorem 12.1 (Discrete Feynman–Kac path formula). *For an initial state i,*

$$\boxed{(u_n)_i = \mathbb{E}_i\left[\exp\left(-\sum_{r=0}^{n-1} V(X_r)\right) f(X_n)\right].}$$

Proof. For $n = 1$,

$$(M_V f)_i = e^{-V(i)} \sum_j K_{ji} f_j,$$

which is the stated expectation. Iterating and expanding the matrix product sums over every path $(i = i_0, i_1, \dots, i_n)$ with probability $\prod_r K_{i_{r+1}, i_r}$ and weight $\prod_r e^{-V(i_r)}$. □

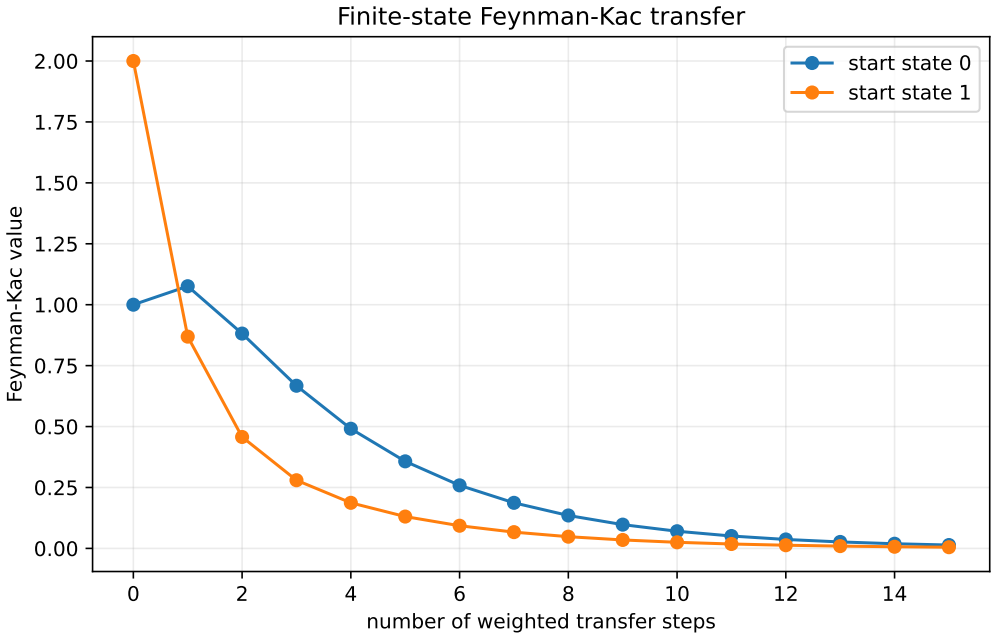


Figure 19. Potential-weighted transfer values computed by the finite Feynman–Kac operator. Exhaustive path enumeration agrees with the matrix formula in the supplied tests.

13 Exact elimination of latent packets

The central idea

Elimination removes hidden coordinates from the displayed state, but their future influence survives. The resolvent stores that influence as a self energy. The time description stores it as a memory kernel. Both descriptions refer to the same declared block system.

Let $X = PX \oplus QX$ and write

$$A = \begin{pmatrix} A_{PP} & A_{PQ} \\ A_{QP} & A_{QQ} \end{pmatrix}.$$

Theorem 13.1 (Schur–Feshbach projected resolvent). *Assume both $zI - A$ and $zI - A_{QQ}$ are invertible. Define*

$$\boxed{F_P(z) = zI - A_{PP} - A_{PQ}(zI - A_{QQ})^{-1}A_{QP}.}$$

Then

$$\boxed{P(zI - A)^{-1}\big|_{\operatorname{Ran} P} = F_P(z)^{-1}.}$$

Proof. Solve the hidden block equation

$$(zI - A_{QQ})x_Q = A_{QP}x_P + y_Q$$

for x_Q and substitute it into the retained equation. The resulting retained coefficient is $F_P(z)$. Invertibility of the full operator and of the hidden block implies invertibility of this Schur complement. Setting $y_Q = 0$ gives the stated identity as an operator on $\operatorname{Ran} P$. Equivalently, this is the block inverse formula. □

The self-energy

$$\Sigma_P(z) = A_{PQ}(zI - A_{QQ})^{-1}A_{QP}$$

contains the exact effect of the eliminated packets. A bare truncation A_{PP} discards this term and is generally not exact.

13.1 Time-domain memory

For the homogeneous system $\dot{x} = Ax$, variation of constants gives

$$x_Q(t) = e^{tA_{QQ}}x_Q(0) + \int_0^t e^{(t-r)A_{QQ}}A_{QP}x_P(r)\,dr.$$

Substitution yields

$$\dot{x}_P(t) = A_{PP}x_P(t) + A_{PQ}e^{tA_{QQ}}x_Q(0) + \int_0^t \mathcal{K}(t-r)x_P(r)\,dr,$$

where

$$\boxed{\mathcal{K}(t) = A_{PQ}e^{tA_{QQ}}A_{QP}.}$$

Theorem 13.2 (Memory/self-energy correspondence)**.** *For finite complex block matrices, assume* $\operatorname{Re} z > \operatorname{Re}\lambda$ *for every* $\lambda \in \sigma(A_{QQ})$*. Then the following improper integral is absolutely convergent and*

$$\boxed{\int_0^\infty e^{-zt}\mathcal{K}(t)\,dt = A_{PQ}(zI - A_{QQ})^{-1}A_{QP} = \Sigma_P(z).}$$

Proof. Put $B = A_{QQ}$ and $T_z(t) = e^{-zt}e^{tB} = e^{t(B-zI)}$. The generalized-eigenspace decomposition in finite dimensions writes T_z as a finite sum of constant matrix coefficients times $t^k e^{(\lambda-z)t}$. This follows by expanding the nilpotent part on each generalized eigenspace; equivalently, apply the decomposition to left multiplication by B on the matrix algebra and evaluate the exponential at I. The strict spectral inequality makes every such term absolutely integrable on $(0,\infty)$ and forces it to tend to zero. Hence T_z is Bochner integrable and $T_z(t) \to 0$, including when B is nonnormal.

Let $J = \int_0^\infty T_z(t)\,dt$. Integrating both derivative identities $T_z' = (B - zI)T_z = T_z(B - zI)$ and using $T_z(0) = I$ gives

$$(zI - B)J = J(zI - B) = I.$$

Thus $J = (zI - B)^{-1}$. Continuous left and right multiplication carry the Bochner integral through the two rectangular transfer blocks, yielding the claim. □

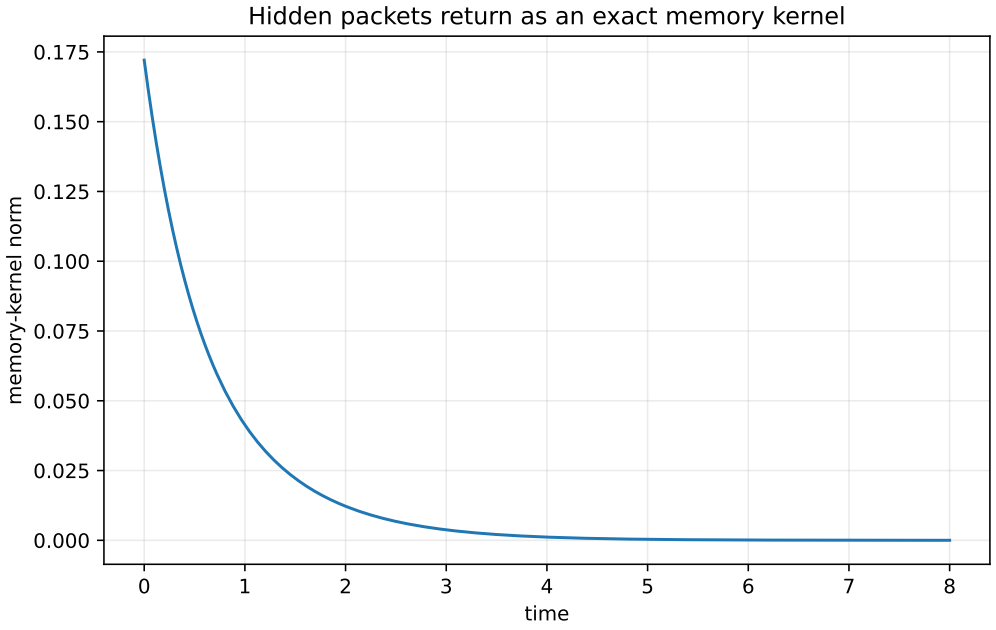


Figure 20. The norm of the exact memory kernel produced by eliminated packets.

14 Survival and hazard semigroups

Let A be a finite subgenerator: its off-diagonal entries are nonnegative and its column sums are nonpositive. These sign conditions alone allow nonabsorbing states; eventual absorption is a separate hypothesis. Let p_0 be an initial transient-state distribution. Define

$$p(t) = e^{tA}p_0, \qquad S(t) = 1^T p(t).$$

The exit-rate row vector is

$$a^T = -1^T A \geq 0.$$

Theorem 14.1 (Semigroup survival identities). *For $t \geq 0$, the finite-time exit density and, wherever $S(t) > 0$, the hazard are*

$$\boxed{f(t) = a^T e^{tA} p_0 = -S'(t), \qquad h(t) = \frac{f(t)}{S(t)}.}$$

Consequently, S is nonincreasing and

$$S(t) = \exp\left(-\int_0^t h(r)\,dr\right)$$

whenever $S(t) > 0$. The density may be defective: its total mass is $1 - \lim_{t\to\infty} S(t)$. It is a probability density when absorption is eventual, so that this limit is zero.

Proof. Differentiate:

$$S'(t) = 1^T A e^{tA} p_0 = -a^T e^{tA} p_0.$$

The hazard identity is the definition of conditional instantaneous exit rate. Integrating $(\log S)' = -h$ gives the final formula. □

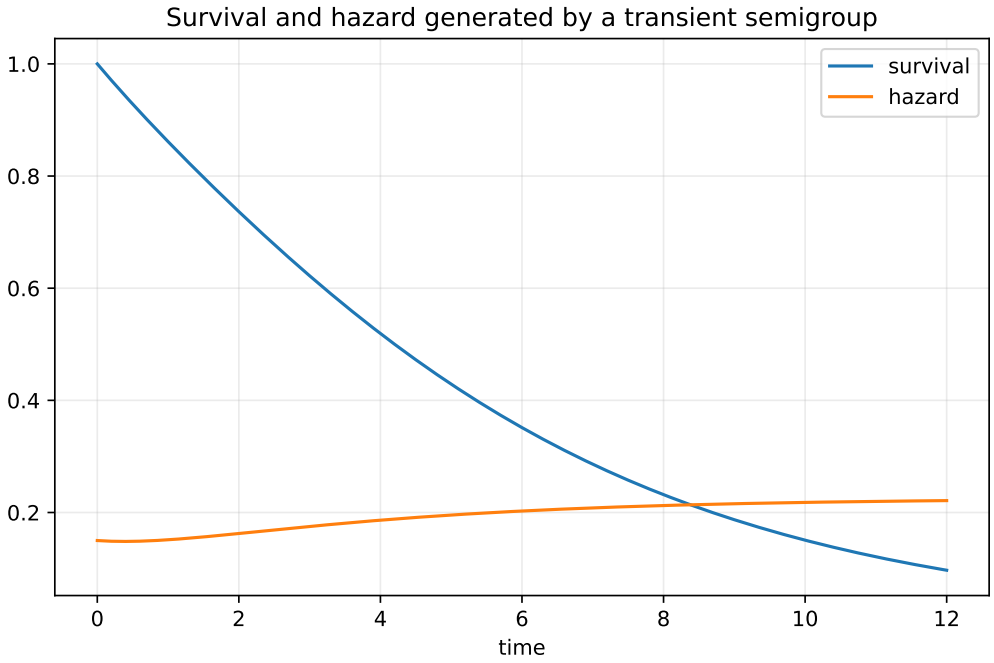


Figure 21. Survival and instantaneous hazard produced by one transient transfer semigroup.

15 One common operator language

The preceding constructions are not identical, but they use the same packet operations. The domain dictionary is summarized in table 1.

Realization	Seed	Transfer	Packet/reconstruction law
Markov	point mass δ_i	K^n or e^{tQ}	state lumping and $CK^n = \bar{K}^n C$
Conditional	random variable	P_Q	nested partitions and tower law
Bayesian	prior p	$\mathcal{B}_\lambda$	packet evidences and normalized synthesis
Fractional	spectral source	e^{-tL^α}	spectral projectors P_j
PDE	source f or δ_y	resolvent/Green operator	eigenmode or spatial packet synthesis
Feynman–Kac	terminal data f	$(\mathrm{diag}(e^{-V})K^T)^n$	weighted path reconstruction
Latent system	retained state	Feshbach/self-energy	hidden packets return as memory
Survival	initial transient law	e^{tA}	sum over transient packets

Table 1. The transfer-semigroup dictionary in the common operator language.

For clarity, the finite-dimensional holomorphic calculus used below has the following canonical definition. For A acting on $\mathbb{C}^d$, its generalized eigenspaces $E_\mu = \ker(A - \mu I)^d$ form a direct-sum decompo-

sition over the eigenvalues. Define the unique linear operator $F(A)$ by

$$F(A)|_{E_\mu} = \sum_{k=0}^{d-1} \frac{F^{(k)}(\mu)}{k!}(A-\mu I)^k\Big|_{E_\mu}.$$

The formula is independent of a basis and of increasing the truncation index: the extra nilpotent powers vanish. It depends only on the finite Taylor jets of F at the spectrum. If F is holomorphic on an open neighborhood U of the spectrum, each eigenvalue admits a radius $r_\mu > 0$ with $\overline{D(\mu, r_\mu)} \subset U$, and Cauchy's formula gives

$$\frac{F^{(k)}(\mu)}{k!} = \frac{1}{2\pi i}\oint_{|z-\mu|=r_\mu} \frac{F(z)}{(z-\mu)^{k+1}}\,dz.$$

Consequently the same canonical operator is obtained from any such local Cauchy circles, even when U is disconnected. This is the standard finite Taylor-jet definition of the holomorphic matrix function.

Theorem 15.1 (Naturality under exact intertwining)**.** *For finite complex matrices, let F be holomorphic on a neighborhood of the union of the spectra of A_f and A_c. Real matrices are treated by complexification. If*

$$CA_f = A_cC,$$

then

$$\boxed{CF(A_f) = F(A_c)C.}$$

In particular, exact generator refinement propagates simultaneously to semigroups, resolvents, integer powers, and analytic transfer functions.

Proof. The intertwining identity implies $C(A_f-\mu I)^k = (A_c-\mu I)^kC$ for every $k \geq 0$. It therefore sends the generalized μ-eigenspace of A_f into that of A_c. Use a common truncation index at least as large as both dimensions in the canonical formula above. Multiplying its finite sum by C now gives the desired identity on every generalized eigenspace. The direct-sum decomposition gives it on the full space. The argument covers nilpotent Jordan parts and uses only local spectral Taylor jets; it does not require one Taylor series to converge on both spectra. The zero-dimensional case is immediate. □

Remark 15.2 (Where nonlinearity enters). Bayesian conditioning does not fit the preceding linear functional calculus because normalization depends on the current state. It nevertheless obeys an exact composition law and has a closed packet reconstruction. Paper III supplies the general nonlinear coordinate theorem; the Bayesian map is a particularly transparent probabilistic realization.

16 Numerical certificates

All calculations in this release were run on CPU. No GPU timing or performance claim is made. Floating-point residuals are implementation checks and are not substitutes for the symbolic proofs.

Numerical Certificate 16.1 (Test suite). The September 6 release records the following automated-test result:

$$\boxed{54/54 \text{ automated tests.}}$$

The tests cover projector reconstruction, packet-block semigroup composition, exact and approximate Markov lumping, Bayesian composition and flow, the exact Bayesian refinement criterion, information-

Certificate	Value
Packet-block semigroup composition residual	2.554×10^{-15}
Operator reconstruction residual	0
Random-walk mass error	0
Exact Markov generator-lumping defect	5.551×10^{-17}
Exact finite-horizon lumping error	2.220×10^{-16}
Perturbed ten-step lumping error	6.523×10^{-3}
Linear defect bound for that example	2.000×10^{-1}
Bayesian composition residual	5.551×10^{-17}
Bayesian packet reconstruction residual	0
Bayesian refinement commutation residual	1.110×10^{-16}
Bayesian defect-factorization residual	7.286×10^{-17}
Bayesian flow generator residual	1.959×10^{-11}
Conditional-expectation tower residual	0
Hermite Gram residual	9.992×10^{-15}
Fractional semigroup residual	2.082×10^{-17}
Maximum square-root subordination error	1.665×10^{-16}
Exact fractional-naturality residual	0
Fractional-power defect / gap certificate	$0.51059/0.95079$
Fractional-semigroup defect / certificate	$0.12969/0.44480$
Observed low-mode PDE convergence order	1.999854
Feynman–Kac enumeration residual	1.665×10^{-16}
Feshbach projected-resolvent residual	5.551×10^{-17}
Memory Laplace-transform residual	6.939×10^{-18}
Duhamel refinement identity residual	2.108×10^{-16}
Survival monotonicity violation	0
Hidden-Markov normalization error	1.110×10^{-16}

Table 2. Selected results from the machine-readable numerical record.

loss factorization, likelihood-profile partitions, conditional-expectation towers, Hermite orthogonality, fractional semigroups, square-root subordination, fractional intertwining and semigroup-defect bounds, discrete Green kernels, PDE convergence, Feynman–Kac enumeration, Schur–Feshbach elimination, memory kernels, refinement defects, hidden Markov filtering, and survival curves.

17 Remaining conjectural extensions

The finite theory closes several operator identities but does not settle the corresponding universal continuum questions.

Conjecture 17.1 (Transfer-semigroup realization)**.** Every sufficiently regular realization of the seed calculus whose transfer stage is local in time admits a closed generator on a completed state space, and the packet analysis maps form a compatible system for its semigroup or evolution family.

Conjecture 17.2 (Uniform refinement control)**.** For a nested family of physically or analytically meaningful discretizations, there exist norms and constants independent of resolution such that

$$\left\|C_\ell e^{tA_{\ell+1}} - e^{tA_\ell} C_\ell\right\| \le M\, t\, \left\|C_\ell A_{\ell+1} - A_\ell C_\ell\right\|$$

on every bounded time interval. Summable generator defects would then produce a coherent inverse-limit evolution.

Conjecture 17.3 (Latent-packet closure)**.** For a stable infinite packet system, exact elimination of unresolved packets produces a causal memory kernel whose Laplace transform is the limiting packet self-energy. A Markovian closure exists precisely when this self-energy is constant, or can be represented by finitely many additional state packets.

Boundary 17.4 (What remains outside Paper VI). The following issues are not proved here:

1. existence and uniqueness for arbitrary stochastic differential equations;
2. posterior consistency or asymptotic normality in infinite statistical models;
3. domain invariance for unbounded fractional operators;
4. tightness and compactness required for stochastic continuum limits;
5. nonlinear PDE well-posedness beyond the stated finite and linear models;
6. an optimal packet choice for a general Markov, Bayesian, or PDE problem.

These belong to later realization-specific work or to external analytic theory.

18 Computational implementation

The accompanying Python package implements the formulas and finite models used in the paper.

`projector_transfer_semigroups`

The computational organization mirrors the paper:

`packets.py`	projectors, blocks, reconstruction, block composition
`semigroup.py`	matrix semigroups, defects, Feshbach operators, memory
`markov.py`	random walks, exact lumping, survival
`bayes.py`	normalized updates, packet evidence, conjugacy, flows, HMM
`conditional.py`	conditional expectation and Gaussian chaos
`fractional.py`	spectral powers, subordination, and fractional refinement certificates
`pde.py`	Dirichlet Laplacians, Green kernels, eigenvalue error
`feynman_kac.py`	weighted transfer and exhaustive path enumeration.

18.1 Representative algorithms

Listing 1. Normalized finite Bayesian update.

```
def bayes_update(prior, likelihood):
    p = normalize(prior)
    evidence = likelihood @ p
    if evidence <= 0:
        raise ValueError("positive evidence required")
    return likelihood * p / evidence
```

Listing 2. Packet block composition.

```
def block_composition(left, right):
    result = []
    for beta in range(len(left)):
        row = []
        for alpha in range(len(left)):
            value = sum(
                left[beta][gamma] @ right[gamma][alpha]
                for gamma in range(len(left))
            )
            row.append(value)
        result.append(row)
    return result
```

Listing 3. Finite Schur-Feshbach effective operator.

```
def feshbach_effective_operator(A, retained, z):
    App, Apq, Aqp, Aqq = split_blocks(A, retained)
    return z*I - App - Apq @ inv(z*I - Aqq) @ Aqp
```

The full implementation, tests, tables, figure-generation code, and exact numerical record are included with the release rather than embedded in the typeset paper.

19 Conclusion

The transfer-semigroup realization gives one common operator language to several subjects that are often presented separately. Markov propagation is linear transfer of a probability seed. Conditional expectation is stochastic projection. Bayesian conditioning is normalized nonlinear transfer, and its exact finite coarsening law is now characterized by packetwise likelihood sufficiency. Fractional calculus is

spectral packet reweighting, and bounded positive refinement defects now transfer to fractional powers and fractional semigroups by explicit resolvent and Duhamel estimates. A Green operator reconstructs a field from source packets. Feynman–Kac propagation adds path weights to stochastic transfer. Schur–Feshbach elimination shows that hidden packets are not destroyed; they return through self-energy and memory. Survival and hazard are scalar reconstructions of a transient semigroup.

The central structural identity is

$$\boxed{\text{generator defect} \longrightarrow \text{Duhamel semigroup defect} \longrightarrow \text{controlled refinement error.}}$$

Together with exact packet reconstruction, this makes the semigroup paper the natural bridge from the foundational calculus to the quantum, computational, and arithmetic realizations that follow.

A Lean verification and scope

Fifth checked block, 6 September 2026. The release contains 341 explicit Lean theorem declarations across 47 proof modules, including 31 additional declarations since the fourth checked block. The September 6 release records a successful full build and namespace audit with Lean 4.34.0-rc2 and pinned mathlib. The audit inspects 782 declarations, including definitions and generated auxiliaries; only `propext`, `Classical.choice` and `Quot.sound` are allowed. These are different counting units from the paper's 33 theorem-like results.

The ledger has **33 complete, zero partial and zero pending results**. The last two results, 13.2 and 15.1, are now complete. The three separately stated conjectures remain open and are excluded from this count.

"Complete" means the stated scope follows from the named checked targets and the ordinary specializations recorded in the ledger. Prose-to-Lean correspondence is manually reviewed. The fractional targets use complex Hilbert spaces; complexification for real operators is an ordinary mathematical specialization, not a separately checked Lean construction. Completion of the theorem ledger does not certify every prose sentence, numerical figure or open conjecture.

Result	Manuscript statement	Lean scope
1.2	Finite semigroup law	Complete
2.1	Packet semigroup convolution	Complete
2.2	Reducing packets	Complete
3.1	Exact Duhamel refinement identity	Complete
3.2	Norm bound	Complete
3.3	Resolvent refinement identity	Complete
4.1	Finite Markov semigroup	Complete
5.1	Exact lumpability	Complete
5.2	Finite-horizon defect formula	Complete
6.1	Finite conditional-expectation projector	Complete
6.2	Tower law and martingale packets	Complete
7.1	Independent-likelihood composition	Complete
7.2	Bayesian packet reconstruction	Complete
7.3	Likelihood perturbation bound	Complete
8.1	Exact effective-likelihood factorization	Complete
8.2	Bayesian refinement criterion	Complete
8.4	Bayesian information-loss packet	Complete
8.5	Nested Bayesian packet tower	Complete
8.6	Canonical coarsest likelihood-resolving partition	Complete
9.1	Bayesian nonlinear semigroup	Complete
9.3	Prediction-update reconstruction	Complete
10.1	Fractional packet calculus	Complete
10.2	Poisson subordination	Complete
10.3	Fractional intertwining theorem	Complete
10.4	Fractional semigroup refinement	Complete
10.5	Fractional refinement towers	Complete
11.1	Exact discrete Green reconstruction at grid nodes	Complete
11.2	Second-order low-mode convergence	Complete
12.1	Discrete Feynman–Kac path formula	Complete
13.1	Schur–Feshbach projected resolvent	Complete
13.2	Memory/self-energy correspondence	Complete
14.1	Semigroup survival identities	Complete

15.1	Naturality under exact intertwining	Complete

General nonnormal memory (13.2). The actual damped exponential is decomposed into finitely many terms of the form $t^k e^{(\mu-z)t}$. Integrability and convergence to zero follow from $\operatorname{Re}\mu < \operatorname{Re} z$. The formal proof uses generalized eigenspaces of left multiplication on the matrix algebra, so nilpotent Jordan parts are included. Both inverse identities follow by integrating the two exponential derivative formulas. A continuous linear sandwich map gives the literal rectangular matrix memory integral and the actual inverse. No normality or externally supplied growth bound is required.

Canonical holomorphic naturality (15.1). The standard finite Taylor-jet calculus defined before the theorem is constructed using the direct sum of generalized eigenspaces. Uniqueness, cutoff stability and exact rectangular naturality are proved. Cauchy's formula identifies the local coefficients with actual circle integrals; admissible radii exist in every stated open holomorphic neighborhood, and changing them leaves the canonical operator unchanged. The argument permits disconnected neighborhoods and different matrix dimensions. It makes no assertion about general Banach-algebra contour deformation.

Fractional results (10.3–10.5). The inherited fourth block proves the scalar beta normalization, the Lebesgue integral representation of actual positive fractional powers, the literal resolvent-defect formula, both printed norm constants and exact intertwining. The rectangular complex-Hilbert-space semigroup proof substitutes those constants into Duhamel's identity. The varying-space tower proof gives summability of fractional-defect norms and uniform convergence of time-defect norm series on bounded time intervals. Here "summable" means summable in norm. Operators with unrelated codomains are not added together; a separate fixed-space theorem proves operator-series convergence when such a sum is defined.

Verification and numerical evidence. The September 6 release records that the full pinned build, project-wide axiom audit, all 54 companion tests and six verifier controls passed. No fresh Lean build is asserted for this October edition. The 54 tests comprise the original 37 and 17 added boundary and validation cases. The 19 numerical figure pairs are unchanged. The earlier 79-field numerical comparison remains historical evidence, not a rerun in that September 6 block. The September 12 visual release subsequently records a numerical replay; current October checks are reported separately in the final package. Machine-readable source contracts, exact theorem targets, full logs and previous checked ledgers accompany the source archive. Numerical tests are finite regressions and are not substitutes for the Lean proofs.

Part II Extended mathematical development

Overview

A bounded generator and a finite packet system turn a semigroup into a matrix of transfer blocks. We prove convolution, differential, path, Laplace, resolvent, and defect formulas. Exact intertwining yields an inverse limit semigroup. Weighted summable adjacent defects give convergent reconstructed evolutions, while range leakage determines whether the limit remains a semigroup. For finite Markov chains we identify a canonical lumping obstruction and quantify its failure. Conditional expectations give orthogonal martingale packets, and the Ornstein Uhlenbeck semigroup gives a countable Hermite model. Bayesian updating is treated as normalized transfer. We characterize exact coarse sufficiency, determine the coarsest nontrivial fixed sufficient partition for a likelihood family when the common level sets are nontrivial, and split relative information into coarse and local terms. Positive fractional powers are handled through spectral, Balakrishnan, and subordination formulas. Green operators, finite differences, and finite state Feynman Kac formulas provide analytic realizations. Schur Feshbach elimination produces a rational self energy and an exact memory kernel. Every stable finite hidden system yields a polynomial exponential expansion, including Jordan blocks. A certified optimal primary block selection and a Volterra estimate control reduced trajectories. State dependent killing, survival laws, moments, and competing risks complete the finite theory.

E1 Introduction

One parameter semigroups provide a common language for evolution equations, Markov processes, diffusion, and resolvent methods. Standard analytic foundations appear in [Paz83, EN00]. Finite Markov chains and their coarse descriptions are treated in [KS60, Nor97, RP81]. Conditional expectation and martingale difference decompositions are basic parts of probability theory [Wil91]. Gaussian transfer and Hermite expansions provide a spectral model for the Ornstein Uhlenbeck semigroup [Jan97]. Fractional powers of generators and subordinate semigroups belong to the functional calculus of positive and sectorial operators [Bal60, SSV12]. The Feynman Kac formula, projection methods, and phase type laws add three further views of transfer [Kac49, Fes58, Zwa61, Neu81].

These subjects share a simple operation. A state is propagated, resolved into declared components, compared across levels, and reconstructed. The same operation can be linear, as in a Markov semigroup, or normalized and nonlinear, as in Bayesian conditioning. Hidden components can be discarded from the state vector without disappearing from the dynamics. Their exact influence returns through a frequency dependent self energy or a time dependent memory kernel.

The first part of the article treats bounded semigroups on Banach spaces. A finite packet system resolves every operator into input and output blocks. The semigroup law becomes block convolution. Splitting the generator into its packet diagonal and transfer part gives a convergent Dyson expansion over directed packet paths. For a coarse map C, the generator defect

$$\Delta = CA_f - A_cC$$

produces the entire semigroup defect through a Duhamel integral. The same defect controls resolvents and holomorphic functions of the generators.

The next question concerns refinement towers. Exact intertwining gives a coordinatewise semigroup on the bounded inverse limit. Approximate refinement is more delicate. Even if every adjacent defect is small, reconstruction from a fine level can introduce a second error because a coarse state need not remain inside the reconstructed fine subspace. We separate these two effects. Summability of adjacent defects gives a uniform limit of reconstructed coarse evolutions. An asymptotic closure condition is then sufficient for the limiting family to satisfy the semigroup law.

The probabilistic part begins with finite Markov chains. Exact lumpability is the vanishing of a canonical obstruction on the kernel of aggregation. In coordinates, this is equality of the transition mass sent from any two states in the same source block into every target block. The usual finite horizon defect formula follows from telescoping. Conditional expectation is then treated as an orthogonal projector. A nested family of partitions produces mutually orthogonal martingale packets and an exact energy decomposition.

Bayesian updating requires more care because normalization depends on the current prior. Given a partition, every fine update has an exact two stage description. The block masses are updated by prior dependent average likelihoods, while the conditional distribution inside each block is updated by its centered likelihood variation. A fixed coarse Bayesian rule works for every prior precisely when the likelihood is constant inside each block. This gives a finite and exact form of Bayesian sufficiency. For a family of likelihoods with at least two common level sets, those level sets form the unique coarsest nontrivial fixed sufficient partition. Relative entropy then separates exactly into information retained by the coarse packet masses and information carried by the local conditional changes. The same part derives perturbation bounds, the replicator flow, a variance identity for accumulated evidence, and stability estimates for hidden Markov filtering.

For a positive operator L, spectral packets define L^{α} and $\mathrm{e}^{-tL^{\alpha}}$. Exact intertwining passes to every fractional power. When both fine and coarse operators have a common positive lower bound m, the Balakrishnan formula gives the quantitative estimate

$$\left\| CL_f^{\alpha} - L_c^{\alpha} C \right\| \leq \alpha m^{\alpha-1} \left\| CL_f - L_c C \right\|, \qquad 0 < \alpha < 1.$$

A Duhamel argument then controls the fractional semigroup defect. This closes a finite version of fractional naturality under explicit spectral hypotheses.

The final sections move from generators to source reconstruction and hidden variables. The one dimensional Dirichlet Green kernel is reproduced exactly by the scaled inverse of the second difference matrix at grid nodes. A discrete maximum principle gives a complete second order error estimate for smooth Poisson solutions. Finite state Feynman Kac transfer is proved in discrete and continuous time. Schur Feshbach elimination yields the exact projected resolvent, the retained memory equation, and the Laplace correspondence between memory and self energy. Every stable finite hidden system gives a polynomial exponential memory expansion. Primary blocks can be ranked optimally for the certified L^1 bound, and a Volterra estimate carries kernel error to the retained solution. Survival, event density, moments, and competing risks are then read directly from a transient semigroup.

Several claims are intentionally absent. No arbitrary unbounded operator is asserted to generate a semigroup. The inverse limit theorem uses bounded generators or an explicit uniform continuity condition. Bayesian sufficiency is proved for a finite partition and a fixed likelihood. The fractional defect estimate uses a common spectral gap. The Poisson estimate concerns the stated one dimensional scheme. Memory compression is controlled for the declared finite or integrable kernels. General stochastic differential equations, posterior consistency, nonlinear partial differential equations, and optimal model reduction remain outside the article.

Realization	Transfer	Exact coarse condition
Operator evolution	e^{tA}	generator intertwining
Markov chain	K^n or e^{tQ}	equality of aggregated transition masses
Conditional expectation	$P_{\mathcal{Q}}$	nesting of the underlying partitions
Bayesian update	$\mathcal{B}_\lambda$	likelihood constant inside each coarse packet
Fractional diffusion	e^{-tL^α}	spectral or resolvent intertwining
Hidden dynamics	projected resolvent	vanishing self energy, or an explicit memory state
Survival law	transient semigroup	summation over all transient states

Table E1. The transfer operations studied in the article and the condition that makes each coarse description exact.

The same words have different meanings in the realizations below. The table records the object that is transferred and the condition that makes its coarse description exact.

E2 Packet resolved semigroups

Let X be a complex Banach space. A finite packet system is a family of bounded idempotents

$$\mathcal{P} = (P_\alpha)_{\alpha \in I}$$

such that

$$P_\alpha P_\beta = \delta_{\alpha\beta} P_\alpha, \qquad \sum_{\alpha \in I} P_\alpha = I_X.$$

No orthogonality is needed in this section.

Definition E2.1 (Packet resolved transfer system)**.** A packet resolved transfer system is a triple

$$(X, A, \mathcal{P})$$

consisting of a Banach state space, a bounded generator, and a finite packet system. A marked transfer system also includes an initial state or source $s \in X$. The topology and all limiting operations are supplied by the ambient Banach space structure.

Definition E2.2 (Packet blocks)**.** For $B \in \mathcal{B}(X)$, define

$$B_{\beta\alpha} = P_\beta B P_\alpha.$$

If $A \in \mathcal{B}(X)$ and $T(t) = \mathrm{e}^{tA}$, write

$$T_{\beta\alpha}(t) = P_\beta T(t) P_\alpha.$$

Theorem E2.3 (Semigroup and packet convolution)**.** *For every bounded operator A, the family $T(t) = \mathrm{e}^{tA}$ is a uniformly continuous group and*

$$T(s+t) = T(s)T(t).$$

For every $\alpha, \beta \in I$,

$$T_{\beta\alpha}(s+t) = \sum_{\gamma \in I} T_{\beta\gamma}(s) T_{\gamma\alpha}(t).$$

Proof. The exponential series converges in operator norm. Since every power of A commutes with every other power, the Cauchy product gives the group law. Inserting $I_X = \sum_\gamma P_\gamma$ between the two factors gives the block formula. □

Proposition E2.4 (Block differential system)**.** *The packet blocks are differentiable in operator norm and satisfy*

$$\frac{\mathrm{d}}{\mathrm{d}t} T_{\beta\alpha}(t) = \sum_{\gamma \in I} A_{\beta\gamma} T_{\gamma\alpha}(t) = \sum_{\gamma \in I} T_{\beta\gamma}(t) A_{\gamma\alpha}.$$

Their initial values are

$$T_{\beta\alpha}(0) = \delta_{\beta\alpha} P_\alpha.$$

Proof. Differentiate $T(t)$ in operator norm and insert the packet resolution between A and $T(t)$, then between $T(t)$ and A. □

Theorem E2.5 (Laplace resolvent and packet transform)**.** *Let $A \in \mathcal{B}(X)$, let $T(t) = \mathrm{e}^{tA}$, and let*

$$\operatorname{Re} z > \|A\|.$$

Then the Bochner integral converges in operator norm and

$$(zI - A)^{-1} = \int_0^\infty \mathrm{e}^{-zt} T(t)\, \mathrm{d}t.$$

For every pair of packets,

$$P_\beta (zI - A)^{-1} P_\alpha = \int_0^\infty \mathrm{e}^{-zt} T_{\beta\alpha}(t)\, \mathrm{d}t.$$

Proof. The estimate $\|T(t)\| \le \mathrm{e}^{t\|A\|}$ gives absolute convergence of the operator valued integral. Let

$$R(z) = \int_0^\infty \mathrm{e}^{-zt} T(t)\, \mathrm{d}t.$$

Since $T'(t) = AT(t) = T(t)A$, integration by parts gives

$$(zI - A) R(z) = -\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t} (\mathrm{e}^{-zt} T(t))\, \mathrm{d}t = I.$$

The same calculation on the other side gives $R(z)(zI - A) = I$. Multiplication by P_β and P_α may be passed under the norm convergent integral, which gives the packet formula. □

Corollary E2.6 (Resolvent expansion over packet paths)**.** *With $A = D + N$ as in the directed path decomposition, let*

$$R_D(z) = (zI - D)^{-1}.$$

If

$$\|R_D(z)N\| < 1,$$

then

$$(zI - A)^{-1} = \sum_{m=0}^{\infty} \big(R_D(z)N\big)^m R_D(z).$$

After multiplication by P_β and P_α, the term of order m is a sum over directed packet paths of length m from α to β.

Proof. Factor

$$zI - A = (zI - D)(I - R_D(z)N)$$

and expand the second inverse by its norm convergent Neumann series. The diagonal resolvent preserves each packet and every factor N changes the packet along an active transfer edge. □

Theorem E2.7 (Reducing packet criterion)**.** *The following conditions are equivalent.*

1. *$AP_\alpha = P_\alpha A$ for every α.*
2. *$A_{\beta\alpha} = 0$ whenever $\beta \neq \alpha$.*
3. *$T_{\beta\alpha}(t) = 0$ whenever $\beta \neq \alpha$ and $t \in \mathbb{R}$.*
4. *Every packet space $P_\alpha X$ is invariant under $T(t)$ for every t.*

When these conditions hold,

$$T(t)|_{P_\alpha X} = \mathrm{e}^{tA_{\alpha\alpha}}.$$

Proof. The first two conditions are equivalent because the packet resolution gives

$$AP_\alpha = \sum_\beta A_{\beta\alpha}, \qquad P_\alpha A = \sum_\gamma A_{\alpha\gamma}.$$

A block diagonal generator has block diagonal powers, so the second condition implies the third. Differentiating the off diagonal blocks at zero proves the converse. The equivalence with invariance follows from the definition of the blocks. The final identity follows by restricting the exponential series. □

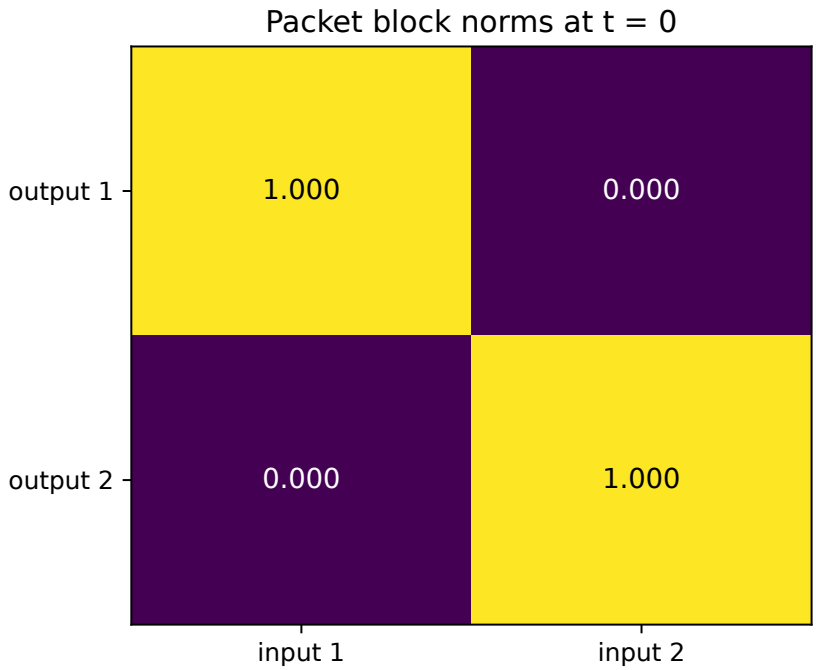


Figure E1. Packet block norms at three times. Transfer appears in the off diagonal blocks as soon as the generator couples the two packet spaces. The three panels are shown in consecutive figures.

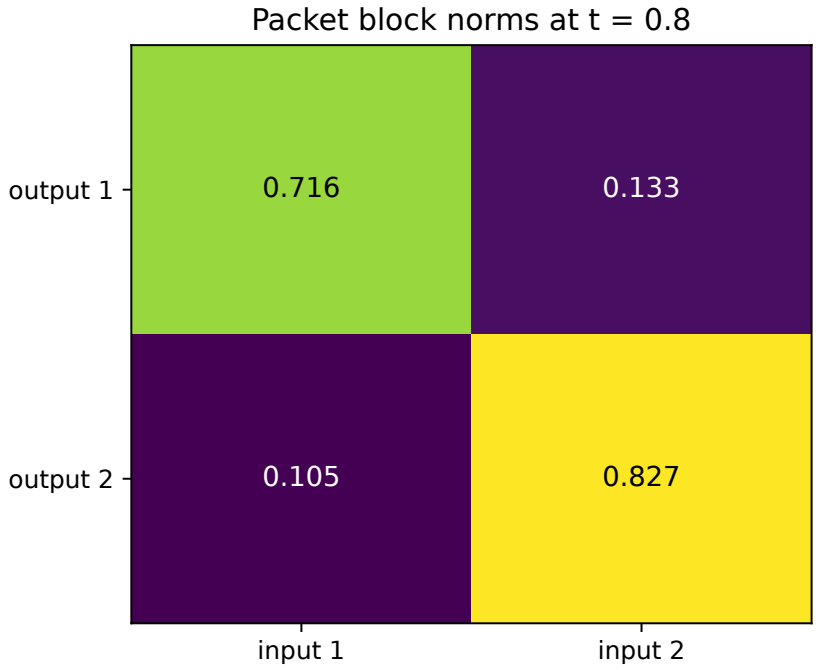


Figure E2. Packet block norms at time 0.8. The same fixed packet spaces and norm scale are used in the preceding panels.

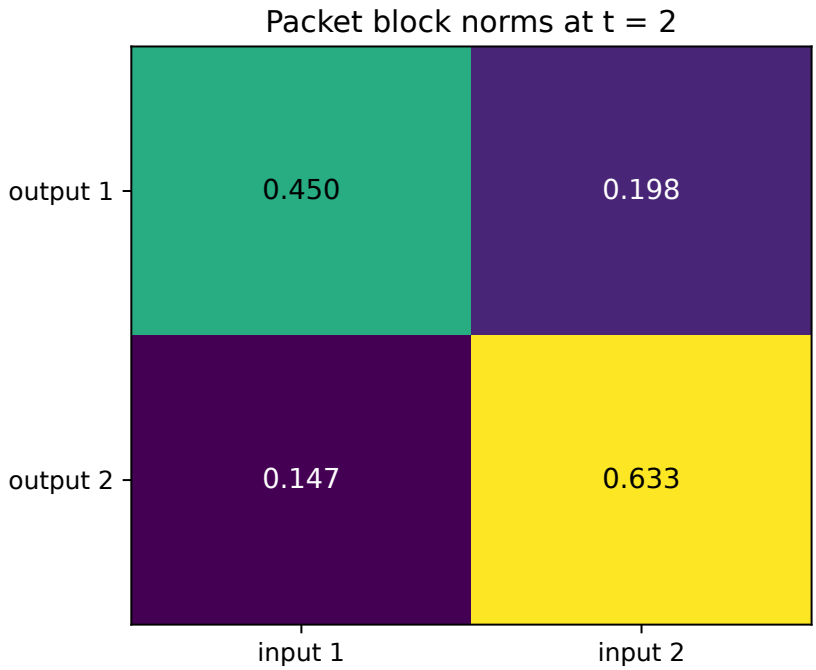


Figure E3. Packet block norms at time 2.0. The same fixed packet spaces and norm scale are used in the preceding panels.

E2.1 Directed path expansion

Set

$$D = \sum_{\alpha \in I} P_\alpha A P_\alpha, \qquad N = A - D.$$

The operator D preserves every packet and N contains only transfers between distinct packets.

Theorem E2.8 (Dyson expansion over packet paths)**.** *For $t \geq 0$,*

$$\mathrm{e}^{tA} = \mathrm{e}^{tD} + \sum_{m=1}^{\infty} \int_{0<s_1<\cdots<s_m<t} \mathrm{e}^{(t-s_m)D} N \mathrm{e}^{(s_m - s_{m-1})D} \cdots N \mathrm{e}^{s_1 D} \, \mathrm{d}s_1 \cdots \mathrm{d}s_m.$$

The series converges absolutely in operator norm. After multiplication by P_β and P_α, the term of order m is a sum over directed packet paths of length m from α to β.

Proof. Write the variation of constants identity

$$\mathrm{e}^{tA} = \mathrm{e}^{tD} + \int_0^t \mathrm{e}^{(t-s)D} N \mathrm{e}^{sA} \, \mathrm{d}s.$$

Iterating the identity gives the displayed series. The norm of the term of order m is at most

$$\mathrm{e}^{t\|D\|} \frac{(t\,\|N\|)^m}{m!}.$$

The majorant is summable. Inserting the packet resolution around every factor N leaves only products whose consecutive packet labels are joined by nonzero transfer blocks. □

Corollary E2.9 (Termination on an acyclic transfer graph)**.** *Suppose the directed graph with an edge $\alpha \to \beta$ whenever $P_\beta N P_\alpha \neq 0$ is acyclic. If its longest directed path has length q, then every Dyson term of order greater than q vanishes.*

Proof. Every nonzero term of order m contains a directed transfer path of length m. An acyclic graph has no path longer than q. □

E3 Generator defects and functional calculus

Let X_f and X_c be complex Banach spaces. Let $A_f \in \mathcal{B}(X_f)$, $A_c \in \mathcal{B}(X_c)$, and $C \in \mathcal{B}(X_f, X_c)$. Define

$$\Delta = CA_f - A_c C$$

and

$$\mathfrak{D}(t) = C\mathrm{e}^{tA_f} - \mathrm{e}^{tA_c}C.$$

Theorem E3.1 (Duhamel defect identity)**.** *For every $t \geq 0$,*

$$\mathfrak{D}(t) = \int_0^t \mathrm{e}^{(t-s)A_c}\Delta \mathrm{e}^{sA_f}\,\mathrm{d}s.$$

If

$$\left\|\mathrm{e}^{tA_f}\right\| \leq M_f \mathrm{e}^{\omega_f t}, \qquad \left\|\mathrm{e}^{tA_c}\right\| \leq M_c \mathrm{e}^{\omega_c t},$$

then

$$\|\mathfrak{D}(t)\| \leq M_f M_c \|\Delta\| \int_0^t \mathrm{e}^{\omega_c(t-s)+\omega_f s}\,\mathrm{d}s.$$

For contraction semigroups,

$$\|\mathfrak{D}(t)\| \leq t\,\|\Delta\|.$$

Proof. For fixed t, set

$$F(s) = \mathrm{e}^{(t-s)A_c} C \mathrm{e}^{sA_f}.$$

Then

$$F'(s) = \mathrm{e}^{(t-s)A_c}\Delta \mathrm{e}^{sA_f}.$$

Integrating from zero to t gives the identity. The estimates follow by taking norms under the integral. □

Proposition E3.2 (Infinitesimal recovery of the defect)**.** *The generator defect is the first order semigroup defect:*

$$\lim_{t\to 0} \frac{\mathfrak{D}(t)}{t} = \Delta$$

in operator norm. Consequently,

$$C\mathrm{e}^{tA_f} = \mathrm{e}^{tA_c}C \quad \textit{for every } t$$

if and only if $CA_f = A_c C$.

Proof. The operator norm expansions

$$\mathrm{e}^{tA_f} = I + tA_f + o(t), \qquad \mathrm{e}^{tA_c} = I + tA_c + o(t)$$

give the limit. The equivalence follows from this limit and theorem E3.1. □

Theorem E3.3 (Resolvent defect identity)**.** *Let z belong to the resolvent sets of A_f and A_c. With*

$$R_f(z) = (zI - A_f)^{-1}, \qquad R_c(z) = (zI - A_c)^{-1},$$

one has

$$CR_f(z) - R_c(z)C = R_c(z)\Delta R_f(z).$$

Proof. Multiply the proposed identity on the left by $zI - A_c$ and on the right by $zI - A_f$. Both sides reduce to

$$(zI - A_c)C - C(zI - A_f) = \Delta.$$

□

Theorem E3.4 (Holomorphic defect formula)**.** *Let Γ be the positively oriented rectifiable boundary of a finite union of pairwise disjoint bounded Jordan domains. Assume that both spectra lie inside this union and that Γ lies in the common resolvent set. Thus its winding number is one about every spectral point. If g is holomorphic on a neighborhood of the closure of the union, then*

$$Cg(A_f) - g(A_c)C = \frac{1}{2\pi\mathrm{i}} \int_\Gamma g(z) R_c(z) \Delta R_f(z)\, \mathrm{d}z.$$

In particular,

$$\left\| Cg(A_f) - g(A_c)C \right\| \le \frac{\operatorname{length}(\Gamma)}{2\pi} \sup_{z\in\Gamma} |g(z)| \sup_{z\in\Gamma} \|R_c(z)\| \sup_{z\in\Gamma} \left\|R_f(z)\right\| \|\Delta\|.$$

Proof. Apply the Dunford integral formula to both operators and use theorem E3.3 under the contour integral. The norm estimate is immediate. □

Corollary E3.5 (Exact naturality)**.** *If $CA_f = A_cC$, then*

$$Cg(A_f) = g(A_c)C$$

for every holomorphic function covered by theorem E3.4. The conclusion includes exponentials, resolvents, and polynomials.

Proof. Set $\Delta = 0$ in theorem E3.4. □

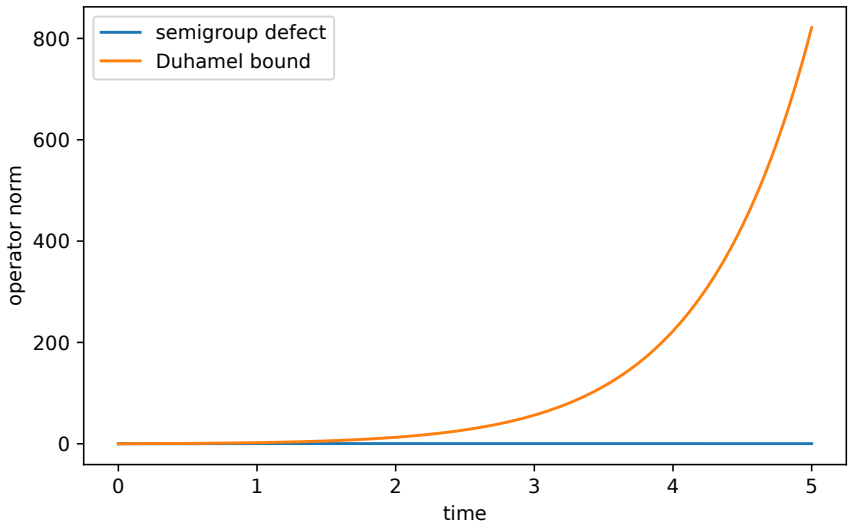


Figure E4. A finite semigroup defect and a direct Duhamel bound. The bound is designed to be uniform rather than sharp in this example.

E4 Refinement towers and limiting evolution

> **How to read the result**
>
> The limiting argument names its topology, reconstruction maps, and summability weights. Convergence of reconstructed trajectories is distinct from closure under the semigroup law. The range-leakage condition identifies the additional obstruction to that closure.

The exact inverse limit and the approximate completion problem require different hypotheses. Exact intertwining preserves the inverse limit coordinate by coordinate. Approximate intertwining gives a Cauchy family only after reconstruction maps are chosen.

E4.1 Exact inverse limits

Let

$$X_0 \xleftarrow{C_0} X_1 \xleftarrow{C_1} X_2 \leftarrow \cdots$$

be Banach spaces with bounded connecting maps. Define the bounded inverse limit

$$X_\infty = \left\{ (x_\ell)_{\ell \geq 0} : C_\ell x_{\ell+1} = x_\ell, \quad \sup_\ell \|x_\ell\| < \infty \right\}$$

with the supremum norm.

Lemma E4.1 (Completeness of the bounded inverse limit)**.** *The space X_∞ is Banach.*

Proof. The product of the spaces X_ℓ, equipped with the bounded supremum norm, is Banach. The equations $C_\ell x_{\ell+1} = x_\ell$ define a closed linear subspace. □

Theorem E4.2 (Exact projective limit semigroup). *For each ℓ, let $T_\ell(t)$ be a strongly continuous semigroup on X_ℓ. Assume*

$$C_\ell T_{\ell+1}(t) = T_\ell(t) C_\ell$$

for every ℓ and $t \geq 0$, and assume the uniform bound

$$\sup_\ell \|T_\ell(t)\| \leq M \mathrm{e}^{\omega t}.$$

Define

$$T_\infty(t)(x_\ell)_\ell = (T_\ell(t) x_\ell)_\ell.$$

Then $T_\infty(t)$ is a bounded semigroup on X_∞. It is strongly continuous if

$$\lim_{t \downarrow 0} \sup_\ell \|T_\ell(t) x_\ell - x_\ell\| = 0$$

for every $(x_\ell)_\ell \in X_\infty$.
If $T_\ell(t) = \mathrm{e}^{tA_\ell}$, the generators are bounded, and

$$\sup_\ell \|A_\ell\| < \infty, \qquad C_\ell A_{\ell+1} = A_\ell C_\ell,$$

then the continuity condition holds and

$$A_\infty(x_\ell)_\ell = (A_\ell x_\ell)_\ell$$

is the bounded generator of T_∞.

Proof. Exact intertwining shows that $(T_\ell(t)x_\ell)_\ell$ remains compatible. The uniform growth bound makes the coordinatewise operator bounded on the supremum norm. The semigroup law holds in every coordinate. The stated continuity condition is exactly strong continuity in the supremum norm.

For bounded generators with common bound M_A,

$$\sup_\ell \left\|(\mathrm{e}^{tA_\ell} - I)x_\ell\right\| \leq (\mathrm{e}^{tM_A} - 1) \sup_\ell \|x_\ell\|.$$

This tends to zero with t. Exact generator intertwining makes $(A_\ell x_\ell)_\ell$ compatible, and coordinatewise differentiation identifies the generator. □

Remark E4.3. Strong continuity on each level does not by itself imply strong continuity in the supremum norm of the inverse limit. The additional uniform condition in theorem E4.2 cannot be omitted without replacing the topology.

E4.2 Summable adjacent defects

Assume now that each $C_\ell : X_{\ell+1} \to X_\ell$ has a bounded right inverse

$$R_\ell : X_\ell \to X_{\ell+1}, \qquad C_\ell R_\ell = I_{X_\ell}.$$

For $m < n$, set

$$C_{m,n} = C_m C_{m+1} \cdots C_{n-1}, \qquad R_{n,m} = R_{n-1} \cdots R_m.$$

The empty products are

$$C_{m,m} = I_{X_m}, \qquad R_{m,m} = I_{X_m}.$$

Definition E4.4 (Reconstructed fine evolution)**.** For a semigroup $T_n(t)$ on X_n, define the evolution seen from level m by

$$U_{m,n}(t) = C_{m,n}T_n(t)R_{n,m}.$$

Theorem E4.5 (Weighted summable defect completion)**.** *Define the adjacent semigroup defect*

$$E_n(t) = C_nT_{n+1}(t) - T_n(t)C_n.$$

Fix a coarse level m and a time horizon $T > 0$. For $n \geq m$, set

$$w_{m,n} = \|C_{m,n}\| \, \|R_nR_{n,m}\|.$$

If

$$\sum_{n=m}^{\infty} w_{m,n} \sup_{0\leq t\leq T} \|E_n(t)\| < \infty,$$

then $U_{m,n}(t)$ converges in operator norm uniformly for $0 \leq t \leq T$. Its limit is denoted by $U_m(t)$, and

$$\sup_{0\leq t\leq T} \|U_m(t) - U_{m,n}(t)\| \leq \sum_{j=n}^{\infty} w_{m,j} \sup_{0\leq t\leq T} \|E_j(t)\|.$$

Proof. Since $C_nR_n = I$,

$$U_{m,n+1}(t) - U_{m,n}(t) = C_{m,n}(C_nT_{n+1}(t) - T_n(t)C_n)R_nR_{n,m}.$$

Therefore

$$\|U_{m,n+1}(t) - U_{m,n}(t)\| \leq w_{m,n} \|E_n(t)\|.$$

The weighted summability hypothesis makes the sequence uniformly Cauchy. Summing the tail estimates gives the stated bound. □

Corollary E4.6 (Contractive summable defect completion)**.** *Assume*

$$\|C_\ell\| \leq 1, \qquad \|R_\ell\| \leq 1, \qquad \|T_\ell(t)\| \leq 1$$

for every ℓ and $t \geq 0$. If

$$\sum_{\ell=0}^{\infty} \sup_{0\leq t\leq T} \|E_\ell(t)\| < \infty,$$

then the conclusion of theorem E4.5 holds, and

$$\sup_{0\leq t\leq T} \|U_m(t) - U_{m,n}(t)\| \leq \sum_{\ell=n}^{\infty} \sup_{0\leq t\leq T} \|E_\ell(t)\|.$$

Proof. Every weight $w_{m,n}$ is at most one. □

Corollary E4.7 (Generator defect criterion)**.** *Assume*

$$\|C_\ell\| \le 1, \qquad \|R_\ell\| \le 1,$$

and suppose $T_\ell(t) = \mathrm{e}^{tA_\ell}$ *are contraction semigroups with bounded generators. Set*

$$\Delta_\ell = C_\ell A_{\ell+1} - A_\ell C_\ell.$$

If

$$\sum_{\ell=0}^{\infty} \|\Delta_\ell\| < \infty,$$

then the conclusion of corollary E4.6 holds on every bounded time interval, and

$$\sup_{0\le t\le T} \left\|U_m(t) - U_{m,n}(t)\right\| \le T \sum_{\ell=n}^{\infty} \|\Delta_\ell\|\,.$$

Proof. Apply the contraction estimate in theorem E3.1 to every adjacent pair. □

The limit in theorem E4.5 need not satisfy the semigroup law. The obstruction comes from leaving the reconstructed fine subspace between two time steps.

Theorem E4.8 (Asymptotic closure criterion)**.** *Assume the weighted convergence in theorem E4.5 holds for every bounded time interval, and define*

$$P_{m,n} = R_{n,m} C_{m,n}$$

on X_n*. Suppose that for every* $T > 0$*,*

$$\rho_{m,n}(T) = \sup_{0\le s,t\le T} \left\|C_{m,n} T_n(s)(I - P_{m,n}) T_n(t) R_{n,m}\right\| \longrightarrow 0.$$

Then

$$U_m(s+t) = U_m(s)U_m(t)$$

for all $s, t \ge 0$*. If the convergence in theorem E4.5 is locally uniform and every* T_n *is norm continuous, then* U_m *is a norm continuous semigroup.*

Proof. At a finite level,

$$U_{m,n}(s)U_{m,n}(t) - U_{m,n}(s+t) = C_{m,n} T_n(s)(P_{m,n} - I) T_n(t) R_{n,m}.$$

Its norm is at most $\rho_{m,n}(T)$ for $s, t \le T$. Pass to the limit using uniform convergence. Norm continuity follows from a uniform limit of norm continuous functions on compact intervals. □

Remark E4.9. Summable generator defects control consistency between adjacent levels. The quantity $\rho_{m,n}(T)$ controls closure of the reconstructed fine range under the fine evolution. These are different requirements.

Example E4.10 (Summable defects do not force a semigroup limit)**.** Let $X_0 = \mathbb{R}$ and let $X_n = \mathbb{R}^2$ for every $n \ge 1$. Set

$$C_0(x, y) = x, \qquad R_0 x = (x, 0),$$

and use identity connecting maps and reconstructions above level one. Let $T_0(t) = I$ and let every higher level carry the rotation group

$$T_n(t) = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}.$$

Only the first adjacent defect is nonzero, so every defect sum is finite. Yet for $n \geq 1$,

$$U_{0,n}(t) = C_0 T_n(t) R_0 = \cos t.$$

The limiting family is not a semigroup because in general

$$\cos(s+t) \neq \cos s \cos t.$$

The failure is exactly the excursion from the reconstructed line into its orthogonal complement between two time steps.

Proposition E4.11 (Range leakage criterion)**.** *Under the contractive assumptions of corollary E4.6, define*

$$\eta_{m,n}(T) = \sup_{0 \leq t \leq T} \left\| (I - P_{m,n}) T_n(t) R_{n,m} \right\|.$$

Then

$$\rho_{m,n}(T) \leq \eta_{m,n}(T).$$

Consequently, $\eta_{m,n}(T) \to 0$ for every $T > 0$ is sufficient for the limiting family U_m to satisfy the semigroup law.

Proof. For $s, t \leq T$, the contraction assumptions give

$$\begin{aligned} &\left\| C_{m,n} T_n(s) (I - P_{m,n}) T_n(t) R_{n,m} \right\| \\ &\quad \leq \left\| (I - P_{m,n}) T_n(t) R_{n,m} \right\|. \end{aligned}$$

Take the supremum and apply theorem E4.8. □

Theorem E4.12 (Generator control of range leakage)**.** *Let $T(t) = \mathrm{e}^{tA}$ on a Banach space X, let P be a bounded projection, and set $Q = I - P$. Assume*

$$\|T(t)\| \leq M \mathrm{e}^{\omega t}$$

and

$$\left\| \mathrm{e}^{t(QAQ)|_{QX}} \right\| \leq M_Q \mathrm{e}^{\omega_Q t}.$$

Then

$$\|QT(t)P\| \leq M M_Q \|P\|^2 \|QAP\| \int_0^t \mathrm{e}^{\omega_Q (t-s) + \omega s} \, \mathrm{d}s.$$

In particular, for orthogonal projections and contraction bounds,

$$\sup_{0 \leq t \leq T} \|QT(t)P\| \leq T \|QAP\|.$$

Proof. For $x \in PX$, write

$$p(t) = PT(t)x, \qquad q(t) = QT(t)x.$$

The hidden component satisfies

$$q'(t) = QAQq(t) + QAPp(t), \qquad q(0) = 0.$$

Variation of constants gives

$$q(t) = \int_0^t \mathrm{e}^{(t-s)(QAQ)|_{QX}} QAPp(s)\,\mathrm{d}s.$$

Since $\|p(s)\| \le \|P\|\,\|T(s)\|\,\|x\|$ and $x = Px$, the first estimate follows. The contraction case is immediate. □

Corollary E4.13 (A usable closure condition for refinement towers)**.** *Assume the contractive summability hypothesis of corollary E4.6 holds on every bounded time interval. Suppose every $P_{m,n}$ is an orthogonal projection on a Hilbert space and $T_n(t) = \mathrm{e}^{tA_n}$ is contractive. If, for every fixed m,*

$$\left\|(I - P_{m,n})A_n P_{m,n}\right\| \longrightarrow 0,$$

then the limiting family U_m is a semigroup.

Proof. A bounded generator of a contraction semigroup on a Hilbert space is dissipative. If $Q_{m,n} = I - P_{m,n}$ and $y \in Q_{m,n}X_n$, then

$$\operatorname{Re}\left\langle Q_{m,n}A_nQ_{m,n}y, y\right\rangle = \operatorname{Re}\left\langle A_n y, y\right\rangle \le 0.$$

Hence the compressed hidden semigroup is contractive. Apply theorem E4.12 to $P_{m,n}$. Since $P_{m,n}R_{n,m} = R_{n,m}$,

$$\eta_{m,n}(T) \le T\left\|(I - P_{m,n})A_nP_{m,n}\right\|\left\|R_{n,m}\right\|.$$

The right side tends to zero. Use proposition E4.11. □

E5 Finite Markov transfer and lumping

Probability vectors are columns. A column stochastic matrix K has nonnegative entries and column sums one. A continuous time generator Q has nonnegative off diagonal entries and column sums zero.

Theorem E5.1 (Finite Markov semigroup)**.** *If Q is a finite Markov generator, then e^{tQ} is column stochastic for every $t \ge 0$.*

Proof. Choose $a \ge \max_i(-Q_{ii})$. The matrix $Q + aI$ is nonnegative, so its exponential is nonnegative. Hence

$$\mathrm{e}^{tQ} = \mathrm{e}^{-at}\mathrm{e}^{t(Q+aI)}$$

is nonnegative. Since $1^TQ = 0$, differentiation gives

$$1^T\mathrm{e}^{tQ} = 1^T.$$

□

Let $\{B_1, \dots, B_m\}$ be a partition of the fine state set. Its aggregation matrix is

$$C_{\alpha i} = 1_{\{i \in B_\alpha\}}.$$

Definition E5.2 (Canonical lumping obstruction)**.** For a linear fine transition operator K, define

$$\mathrm{Ob}_C(K) = (CK)|_{\mathrm{Ker}\, C} \in \mathcal{B}(\mathrm{Ker}\, C, \mathbb{R}^m).$$

Theorem E5.3 (Descent through an aggregation map)**.** *Let $C : X_f \to X_c$ be a surjective linear map between finite dimensional spaces. A unique operator $\overline{K} : X_c \to X_c$ satisfying*

$$CK = \overline{K}C$$

exists if and only if

$$\mathrm{Ob}_C(K) = 0.$$

For any right inverse R of C, the only possible coarse operator is

$$\overline{K} = CKR,$$

and

$$CK - \overline{K}C = CK(I - RC).$$

Proof. If $CK = \overline{K}C$, then CK vanishes on $\mathrm{Ker}\, C$. Conversely, suppose the obstruction vanishes and define

$$\overline{K}(Cx) = CKx.$$

If $Cx = Cy$, then $x - y \in \mathrm{Ker}\, C$, so the value does not depend on the representative. Surjectivity defines $\overline{K}$ on all of X_c, and the intertwining relation follows from the definition. It also gives uniqueness. If R is a right inverse, then

$$\overline{K} = \overline{K}CR = CKR.$$

Substitution gives the defect identity. □

Corollary E5.4 (Coordinate form of the lumping obstruction)**.** *For the partition aggregation above,* $\mathrm{Ker}\, C$ *is spanned by differences $e_i - e_j$ with i and j in one source block. Also,*

$$\mathrm{Ob}_C(K)(e_i - e_j) = \left(\sum_{r \in B_\alpha} (K_{ri} - K_{rj}) \right)_{\alpha=1}^{m}.$$

Proof. Inside each block, the vectors with coordinate sum zero are spanned by differences of coordinate vectors. Applying CK gives the displayed aggregated column differences. □

Theorem E5.5 (Characterization of exact discrete lumping)**.** *Let K be column stochastic. A coarse column stochastic matrix $\overline{K}$ satisfying*

$$CK = \overline{K}C$$

exists if and only if, for every source block B_β, every pair $i, j \in B_\beta$, and every target block B_α,

$$\sum_{r \in B_\alpha} K_{ri} = \sum_{r \in B_\alpha} K_{rj}.$$

When it exists, $\overline{K}$ is unique and

$$\overline{K}_{\alpha\beta} = \sum_{r \in B_\alpha} K_{ri}$$

for any $i \in B_\beta$.

Proof. The jth column of CK records the transition mass from state j into every target block. The jth column of $\overline{K}C$ is the column of $\overline{K}$ indexed by the source block containing j. Equality therefore holds exactly under the stated condition. The formula gives uniqueness. Summing over α gives one, so the coarse matrix is column stochastic. □

Corollary E5.6 (Continuous time lumping)**.** *A coarse Markov generator $\overline{Q}$ satisfying*

$$CQ = \overline{Q}C$$

exists if and only if the aggregated rates

$$\sum_{r \in B_\alpha} Q_{ri}$$

depend only on the source block of i. In that case

$$C\mathrm{e}^{tQ} = \mathrm{e}^{t\overline{Q}}C$$

for every $t \geq 0$.

Proof. The characterization is the same column argument as in theorem E5.5. The off diagonal entries of $\overline{Q}$ are nonnegative and its column sums are zero. Exact semigroup intertwining follows from proposition E3.2. □

Theorem E5.7 (Finite horizon lumping defect)**.** *For arbitrary fine and coarse operators K and $\overline{K}$, with the fixed aggregation map C, set*

$$\Delta_K = CK - \overline{K}C.$$

Then

$$CK^n - \overline{K}^n C = \sum_{j=0}^{n-1} \overline{K}^{n-1-j} \Delta_K K^j.$$

If K and $\overline{K}$ are column stochastic, then

$$\left\| CK^n - \overline{K}^n C \right\|_{1\to 1} \leq n \left\| \Delta_K \right\|_{1\to 1}.$$

For every probability vector p, the total variation discrepancy is at most

$$\frac{n}{2} \left\| \Delta_K \right\|_{1\to 1}.$$

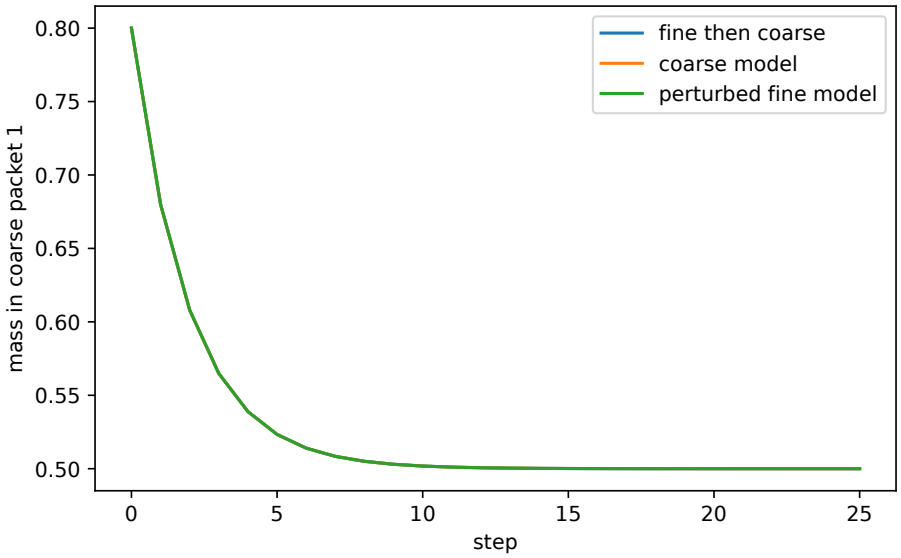


Figure E5. Exact lumping agrees at every horizon. A small violation of the block transition condition produces a visible but controlled discrepancy.

Proof. Sum the telescoping identities

$$\overline{K}^{n-1-j} C K^{j+1} - \overline{K}^{n-j} C K^{j} = \overline{K}^{n-1-j} \Delta_K K^{j}.$$

Stochastic matrices are contractions in the ℓ^1 norm. The last statement uses total variation equal to one half of the ℓ^1 distance. □

Corollary E5.8 (Stationary mass descends)**.** *If $CK = \overline{K}C$ and $Kp = p$, then*

$$\overline{K}(Cp) = Cp.$$

Proof. Apply the intertwining identity to p. □

E6 Conditional expectation and martingale packets

Let $\Omega = \{1, \dots, n\}$ have strictly positive probabilities π_i. For a partition $\mathcal{Q}$, let $B(i)$ denote the block containing i. Define

$$(P_{\mathcal{Q}} f)_i = \frac{\sum_{j \in B(i)} \pi_j f_j}{\sum_{j \in B(i)} \pi_j}.$$

Equip $\mathbb{C}^n$ with

$$\langle f, g \rangle_\pi = \sum_i \pi_i f_i \overline{g_i}.$$

Theorem E6.1 (Finite conditional expectation)**.** *The operator $P_{\mathcal{Q}}$ is the orthogonal projector onto the functions that are constant on every block of $\mathcal{Q}$. In particular,*

$$P_{\mathcal{Q}}^2 = P_{\mathcal{Q}}, \qquad P_{\mathcal{Q}}^* = P_{\mathcal{Q}}.$$

Proof. One application produces a block constant function, so a second application changes nothing. Expanding both weighted inner products block by block gives

$$\langle P_Q f, g\rangle_\pi = \sum_{B\in Q} \frac{\left(\sum_{i\in B}\pi_i f_i\right)\overline{\left(\sum_{j\in B}\pi_j g_j\right)}}{\pi(B)} = \langle f, P_Q g\rangle_\pi .$$

The range is exactly the block constant subspace. □

Theorem E6.2 (Tower law and martingale packets)**.** *Let $P_0, P_1, \dots, P_r$ come from increasingly fine partitions, so that*

$$\operatorname{Ran} P_0 \subseteq \operatorname{Ran} P_1 \subseteq \cdots \subseteq \operatorname{Ran} P_r.$$

Define

$$D_0 = P_0, \qquad D_j = P_j - P_{j-1} \quad (1 \le j \le r).$$

Then

$$P_i P_j = P_{\min(i,j)}, \qquad D_i D_j = \delta_{ij} D_j, \qquad P_r = \sum_{j=0}^{r} D_j.$$

Every D_j is an orthogonal projector, and

$$\|P_r f\|_\pi^2 = \sum_{j=0}^{r} \|D_j f\|_\pi^2$$

and

$$\|f\|_\pi^2 = \sum_{j=0}^{r} \|D_j f\|_\pi^2 + \|f - P_r f\|_\pi^2 .$$

Proof. A coarse average of a fine average is the coarse average. A fine average leaves a coarse constant function unchanged. Hence $P_i P_j = P_{\min(i,j)}$. Expanding products of the differences gives the orthogonality relations. The sum telescopes to P_r. Orthogonal Pythagorean identities give the norm formulas. □

Corollary E6.3 (Law of total variance)**.** *For a real valued f, let $P_0 f = \mathbb{E}_\pi[f]1$. Then*

$$\operatorname{Var}_\pi(f) = \operatorname{Var}_\pi(P_Q f) + \|f - P_Q f\|_\pi^2 .$$

Proof. Subtract the mean and apply the orthogonal decomposition into the range of P_Q and its orthogonal complement. □

Standard Fact E6.4 (General conditional expectations). On a probability space, conditional expectation onto a sub sigma algebra is the orthogonal projector in L^2. Nested sigma algebras satisfy the same tower law. Under the usual hypotheses, martingale convergence gives a limit in L^2 and almost surely. Standard treatments include [Wil91].

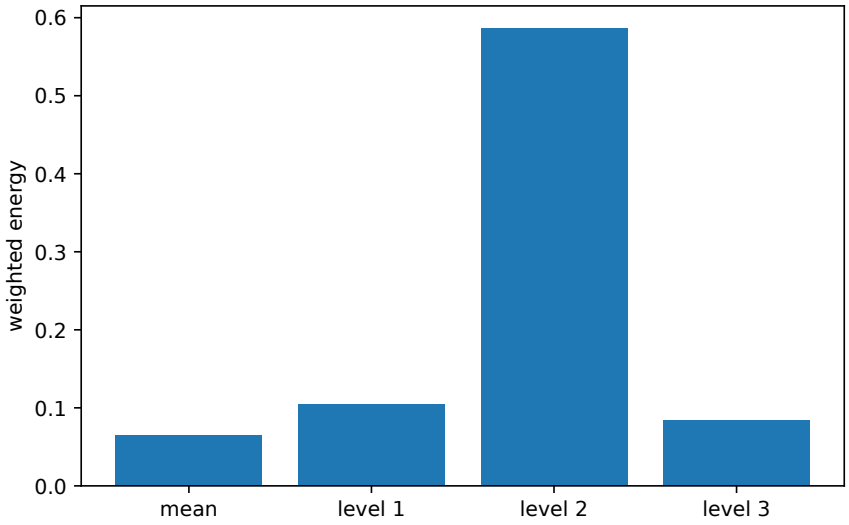


Figure E6. Energy carried by the mean packet and three successive martingale difference packets in a finite weighted example.

E7 Gaussian transfer and Hermite packets

Let γ be the standard Gaussian probability measure on $\mathbb{R}$. For $t \geq 0$, define the Ornstein Uhlenbeck transfer

$$(\mathcal{O}_t f)(x) = \int_{\mathbb{R}} f\left(\mathrm{e}^{-t}x + \sqrt{1-\mathrm{e}^{-2t}}\, y\right) \gamma(\,\mathrm{d}y).$$

Let H_n be the probabilists' Hermite polynomial defined by

$$\exp\left(ux - \frac{u^2}{2}\right) = \sum_{n=0}^{\infty} H_n(x)\frac{u^n}{n!},$$

and set $h_n = H_n/\sqrt{n!}$.

Theorem E7.1 (Ornstein Uhlenbeck semigroup). *The operators $(\mathcal{O}_t)_{t\geq 0}$ form a positive Markov contraction semigroup on $L^p(\gamma)$ for every $1 \leq p \leq \infty$. On $L^2(\gamma)$, the semigroup is self adjoint and*

$$\mathcal{O}_t h_n = \mathrm{e}^{-nt} h_n.$$

The family $(h_n)_{n\geq 0}$ is an orthonormal basis of $L^2(\gamma)$.

Proof. Positivity and preservation of constants follow from the integral formula. If X and Y are independent standard Gaussian variables, then

$$\mathrm{e}^{-s}\left(\mathrm{e}^{-t}x + \sqrt{1-\mathrm{e}^{-2t}}X\right) + \sqrt{1-\mathrm{e}^{-2s}}Y$$

has the same distribution as

$$\mathrm{e}^{-(s+t)}x + \sqrt{1-\mathrm{e}^{-2(s+t)}}Z$$

with Z standard Gaussian. This proves the semigroup law. Jensen's inequality and invariance of γ give the L^p contraction. The joint Gaussian pair

$$\left(X, \mathrm{e}^{-t}X + \sqrt{1-\mathrm{e}^{-2t}}Y\right)$$

is symmetric under interchange, which proves self adjointness.

Apply $\mathcal{O}_t$ to the Hermite generating function. Gaussian integration gives

$$\begin{aligned}\mathcal{O}_t \exp\left(ux - \frac{u^2}{2}\right) &= \exp\left(\mathrm{e}^{-t}ux - \frac{\mathrm{e}^{-2t}u^2}{2}\right)\\ &= \sum_{n=0}^{\infty} \mathrm{e}^{-nt} H_n(x)\frac{u^n}{n!}.\end{aligned}$$

Comparison of coefficients proves the eigenvalue formula. The same generating function shows

$$\int H_m H_n \,\mathrm{d}\gamma = n!\delta_{mn}.$$

The Hermite polynomials span all polynomials, and polynomials are dense in $L^2(\gamma)$. Hence the normalized family is an orthonormal basis. See [Jan97] for the general Gaussian Hilbert space setting. □

Corollary E7.2 (Hermite packet reconstruction)**.** *Let P_n be the orthogonal projection onto $\mathbb{C}h_n$. Then*

$$I = \sum_{n=0}^{\infty} P_n$$

in the strong operator topology and

$$\mathcal{O}_t = \sum_{n=0}^{\infty} \mathrm{e}^{-nt} P_n.$$

For $f \in L^2(\gamma)$,

$$\|\mathcal{O}_t f\|_2^2 = \sum_{n=0}^{\infty} \mathrm{e}^{-2nt}\,\|P_n f\|_2^2\,.$$

Proof. Apply Parseval's identity and the eigenvalue formula in theorem E7.1. □

Proposition E7.3 (Generator and Dirichlet form)**.** *On the polynomial subspace, the infinitesimal generator of $(\mathcal{O}_t)$ is the differential operator*

$$\mathcal{L}f = f'' - xf'.$$

It satisfies

$$\mathcal{L}h_n = -nh_n.$$

For every polynomial f,

$$\langle -\mathcal{L}f, f\rangle_{L^2(\gamma)} = \int_{\mathbb{R}} |f'(x)|^2\,\gamma(\,\mathrm{d}x).$$

Proof. Write

$$\rho_t = \mathrm{e}^{-t} = 1 - t + O(t^2), \qquad \sigma_t^2 = 1 - \mathrm{e}^{-2t} = 2t + O(t^2).$$

For a polynomial f, Taylor expansion at x and Gaussian integration give

$$\begin{aligned}\mathcal{O}_t f(x) &= f(x) + f'(x)(\rho_t - 1)x + \frac{1}{2}f''(x)\sigma_t^2 + o(t)\\ &= f(x) + t(f''(x) - xf'(x)) + o(t).\end{aligned}$$

Every omitted term is a polynomial in the Gaussian variable multiplied by a quantity that is $o(t)$, so the expansion may be integrated term by term. Therefore

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{O}_t f\Big|_{t=0} = f'' - xf'.$$

The Hermite eigenvalue identity also follows by differentiating $\mathcal{O}_t h_n = \mathrm{e}^{-nt}h_n$ at zero. Gaussian integration by parts gives

$$\int_{\mathbb{R}} f'(x)\overline{g(x)}\gamma(\,\mathrm{d}x) = \int_{\mathbb{R}} f(x)\overline{xg(x) - g'(x)}\gamma(\,\mathrm{d}x).$$

Set $g = f'$ and rearrange. □

Corollary E7.4 (Gaussian spectral gap)**.** *If $f \in L^2(\gamma)$ has mean zero, then*

$$\|\mathcal{O}_t f\|_2 \le \mathrm{e}^{-t}\,\|f\|_2\,.$$

For every polynomial f,

$$\mathrm{Var}_\gamma(f) \le \int_{\mathbb{R}} |f'(x)|^2\,\gamma(\,\mathrm{d}x).$$

Proof. A mean zero function has no h_0 component, so every surviving Hermite multiplier is at most e^{-t}. For a polynomial

$$f - \mathbb{E}_\gamma f = \sum_{n\ge 1} c_n h_n.$$

Then

$$\mathrm{Var}_\gamma(f) = \sum_{n\ge 1} |c_n|^2 \le \sum_{n\ge 1} n\,|c_n|^2 = \langle -\mathcal{L}f, f\rangle\,.$$

Apply proposition E7.3. □

Corollary E7.5 (Fractional Hermite transfer)**.** *Let $-\mathcal{L}$ be the nonnegative Ornstein Uhlenbeck number operator with*

$$-\mathcal{L}h_n = nh_n.$$

For every $\alpha > 0$,

$$\mathrm{e}^{-t(-\mathcal{L})^\alpha} f = \sum_{n=0}^{\infty} \mathrm{e}^{-tn^\alpha} P_n f.$$

This is a contraction semigroup on $L^2(\gamma)$.

Proof. The spectral theorem applies to the orthonormal Hermite basis. Every multiplier has modulus at most one and obeys the scalar semigroup law. □

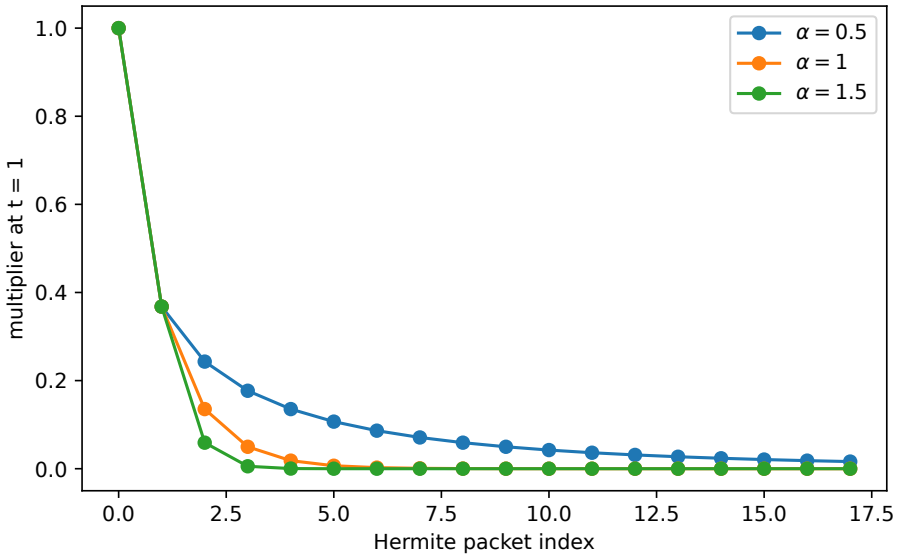


Figure E7. Hermite packet multipliers for the Ornstein Uhlenbeck semigroup and two fractional powers of its number operator.

E8 Bayesian transfer and evidence packets

Let

$$\Delta_{n-1} = \left\{ p \in \mathbb{R}^n : p_i \geq 0, \quad \sum_i p_i = 1 \right\}$$

be the probability simplex. Let $\lambda \in [0, \infty)^n$ have positive evidence

$$Z_\lambda(p) = \sum_i \lambda_i p_i.$$

The Bayesian transfer is

$$\mathcal{B}_\lambda(p) = \frac{\lambda \odot p}{Z_\lambda(p)}.$$

Proposition E8.1 (Scale invariance and composition)**.** *For every $c > 0$,*

$$\mathcal{B}_{c\lambda} = \mathcal{B}_\lambda.$$

If all relevant evidences are positive, then

$$\mathcal{B}_\mu(\mathcal{B}_\lambda(p)) = \mathcal{B}_{\lambda \odot \mu}(p).$$

Proof. A common likelihood scale cancels between the numerator and denominator. In the second identity, the first evidence appears in both the numerator and the denominator of the second update and cancels. □

Let $I = I_1 \sqcup \cdots \sqcup I_m$ be a partition. Define packet evidences

$$Z_\alpha = \sum_{i \in I_\alpha} \lambda_i p_i.$$

For $Z_\alpha > 0$, define

$$m_\alpha = \frac{Z_\alpha}{\sum_\beta Z_\beta}, \qquad q_{\alpha,i} = \frac{\lambda_i p_i}{Z_\alpha} \quad (i \in I_\alpha).$$

Packets with zero evidence receive posterior mass zero.

Theorem E8.2 (Bayesian packet reconstruction). *The posterior is the disjoint packet synthesis*

$$\mathcal{B}_\lambda(p) = \sum_{\alpha : Z_\alpha > 0} m_\alpha q_\alpha.$$

The local shape q_α depends only on the prior and likelihood inside packet I_α. Coupling between different packets occurs through the normalization of the packet evidences.

Proof. For $i \in I_\alpha$,

$$m_\alpha q_{\alpha,i} = \frac{Z_\alpha}{\sum_\beta Z_\beta} \frac{\lambda_i p_i}{Z_\alpha} = \frac{\lambda_i p_i}{Z_\lambda(p)}.$$

The packet supports are disjoint. □

Theorem E8.3 (Likelihood perturbation). *Let $\lambda, \mu \geq 0$ have positive evidences under a fixed prior p. Set*

$$d_p(\lambda, \mu) = \sum_i |\lambda_i - \mu_i| \, p_i.$$

Then

$$\left\| \mathcal{B}_\lambda(p) - \mathcal{B}_\mu(p) \right\|_1 \leq \frac{2 d_p(\lambda, \mu)}{\max\{Z_\lambda(p), Z_\mu(p)\}}.$$

Proof. Add and subtract $\mu \odot p / Z_\lambda(p)$. This gives

$$\left\| \mathcal{B}_\lambda(p) - \mathcal{B}_\mu(p) \right\|_1 \leq \frac{2 d_p(\lambda, \mu)}{Z_\lambda(p)}.$$

Interchanging λ and μ gives the same estimate with $Z_\mu(p)$. Take the smaller of the two upper bounds. □

Theorem E8.4 (Stability with respect to the prior). *Assume*

$$0 < a \leq \lambda_i \leq b$$

for every i. Then

$$\left\| \mathcal{B}_\lambda(p) - \mathcal{B}_\lambda(q) \right\|_1 \leq \frac{2b}{a} \left\| p - q \right\|_1$$

for all priors $p, q \in \Delta_{n-1}$.

Proof. The evidences are at least a. Add and subtract $\lambda \odot q / Z_\lambda(p)$. The numerator difference is at most $b \left\| p - q \right\|_1$, and the evidence difference has the same bound. The two terms give the stated constant. □

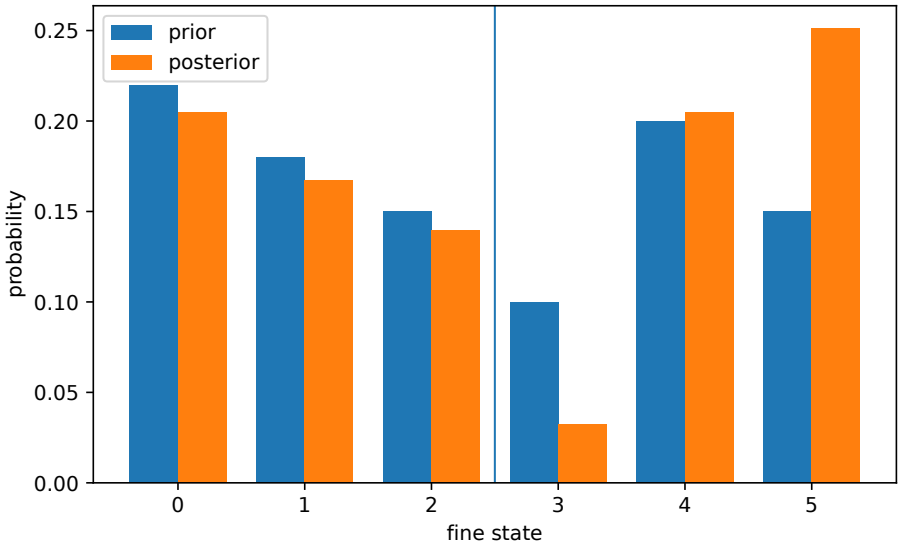


Figure E8. A prior and posterior on two coarse packets. The likelihood is constant in the first packet and varies inside the second, so only the second local conditional distribution changes.

E9 Bayesian coarsening and finite sufficiency

A fixed coarse description will be called sufficient when it reproduces the coarse posterior masses for every positive prior and leaves the conditional distribution inside every packet unchanged. This excludes a coarse rule that hides an unrecorded local update.

Let C send a fine probability vector to its block masses

$$(Cp)_\alpha = \sum_{i \in I_\alpha} p_i.$$

Assume first that every block mass is positive. Write

$$r_{\alpha,i} = \frac{p_i}{(Cp)_\alpha}$$

for the prior conditional distribution inside packet I_α. Define the prior dependent packet average likelihood

$$\overline{\lambda}_\alpha(p) = \sum_{i \in I_\alpha} \lambda_i r_{\alpha,i}.$$

Theorem E9.1 (Exact two stage Bayesian update). *For every strictly positive prior,*

$$C\mathcal{B}_\lambda(p) = \mathcal{B}_{\overline{\lambda}(p)}(Cp).$$

For every packet with $\overline{\lambda}_\alpha(p) > 0$, the posterior conditional distribution is

$$q_{\alpha,i} = \frac{\lambda_i r_{\alpha,i}}{\overline{\lambda}_\alpha(p)}.$$

A packet with zero average likelihood has zero posterior mass. Thus a fine Bayesian update consists of a coarse update by packet average likelihoods followed by independent local updates inside the packets that survive.

Proof. The posterior mass of packet I_α is

$$\frac{\sum_{i\in I_\alpha}\lambda_i p_i}{\sum_j \lambda_j p_j} = \frac{(Cp)_\alpha \overline{\lambda}_\alpha(p)}{\sum_\beta (Cp)_\beta \overline{\lambda}_\beta(p)}.$$

This is the coarse Bayesian formula. Dividing the fine posterior coordinates by the packet mass gives the local formula. □

Theorem E9.2 (Characterization of fixed coarse sufficiency)**.** *Assume $\lambda_i > 0$ for every state. The following conditions are equivalent.*

1. *There is a positive coarse likelihood $\ell \in (0,\infty)^m$ such that*

$$C\mathcal{B}_\lambda(p) = \mathcal{B}_\ell(Cp)$$

   for every strictly positive prior p, and the posterior conditional distribution inside every packet equals its prior conditional distribution.

2. *The likelihood is constant on every packet.*

When these conditions hold, ℓ_α can be chosen as the common value of λ on I_α, and it is unique up to a common positive scale.

Proof. If λ is constant on each packet, theorem E9.1 has packet averages independent of the prior and its local posterior formula gives $q_\alpha = r_\alpha$. This proves the first condition.

Conversely, fix any strictly positive prior. Equality of the local posterior and prior conditionals gives

$$\frac{\lambda_i r_{\alpha,i}}{\overline{\lambda}_\alpha(p)} = r_{\alpha,i}$$

for every $i \in I_\alpha$. Positivity of the conditional prior implies

$$\lambda_i = \overline{\lambda}_\alpha(p)$$

for every state in the packet. Thus the likelihood is constant on every packet. If two positive coarse likelihoods define the same update for every positive coarse prior, comparison of any two posterior odds shows that their coordinate ratios agree. Hence they differ only by a common positive factor. □

Theorem E9.3 (Coarsest fixed sufficient partition)**.** *Let $\mathcal{L}$ be a nonempty family of strictly positive likelihood vectors. Define*

$$i \sim_{\mathcal{L}} j \quad\Longleftrightarrow\quad \lambda_i = \lambda_j \quad \text{for every } \lambda \in \mathcal{L}.$$

Let $\mathcal{Q}_{\mathcal{L}}$ be the partition into equivalence classes. A partition $\mathcal{Q}$ gives a fixed coarse sufficient description in the sense of theorem E9.2 for every $\lambda \in \mathcal{L}$ if and only if every block of $\mathcal{Q}$ is contained in a block of $\mathcal{Q}_{\mathcal{L}}$. Hence $\mathcal{Q}_{\mathcal{L}}$ is the unique coarsest partition with this property.

Proof. Apply theorem E9.2 to every likelihood in the family. A fixed sufficient description exists exactly when each likelihood is constant on every block of $\mathcal{Q}$. This is equivalent to every block of $\mathcal{Q}$ being contained in one common level set of the family. The equivalence class partition is itself admissible, and every other admissible partition refines it. □

Definition E9.4 (Within packet information loss)**.** For a prior p, define

$$\eta_i(p) = \lambda_i - \overline{\lambda}_\alpha(p) \quad \text{when } i \in I_\alpha.$$

Then

$$\sum_{i \in I_\alpha} r_{\alpha,i}\eta_i(p) = 0.$$

Proposition E9.5 (Exact local distortion)**.** *For every packet with positive mass and evidence,*

$$q_{\alpha,i} - r_{\alpha,i} = \frac{r_{\alpha,i}\eta_i(p)}{\overline{\lambda}_\alpha(p)}.$$

Consequently,

$$\|q_\alpha - r_\alpha\|_1 = \frac{\sum_{i \in I_\alpha} r_{\alpha,i}\,|\eta_i(p)|}{\overline{\lambda}_\alpha(p)}.$$

If λ is positive on the packet, then

$$\|q_\alpha - r_\alpha\|_1 \le \frac{\operatorname{osc}_{I_\alpha}(\lambda)}{\min_{i \in I_\alpha} \lambda_i}.$$

Proof. Subtract $r_{\alpha,i}$ from the local posterior formula in theorem E9.1. The norm identity follows because the factor multiplying η_i is nonnegative. The last estimate uses

$$\sum_i r_{\alpha,i}\,|\eta_i| \le \operatorname{osc}_{I_\alpha}(\lambda)$$

and $\overline{\lambda}_\alpha(p) \ge \min_{I_\alpha} \lambda$. □

Theorem E9.6 (Approximate coarse Bayesian rule)**.** *Choose one positive number ℓ_α for every packet and define the block constant fine likelihood*

$$\tilde{\lambda}_i = \ell_\alpha \quad (i \in I_\alpha).$$

Then

$$\|C\mathcal{B}_\lambda(p) - \mathcal{B}_\ell(Cp)\|_1 \le \frac{2\sum_i p_i \left|\lambda_i - \tilde{\lambda}_i\right|}{\max\{Z_\lambda(p), Z_{\tilde{\lambda}}(p)\}}.$$

Proof. Block constancy gives

$$C\mathcal{B}_{\tilde{\lambda}}(p) = \mathcal{B}_\ell(Cp).$$

Aggregation is an ℓ^1 contraction. Apply theorem E8.3 to the two fine posteriors and then aggregate. □

Definition E9.7 (Relative information)**.** For probability vectors q and p, define

$$D(q\|p) = \sum_i q_i \log \frac{q_i}{p_i}.$$

The convention is $0 \log 0 = 0$. A term with $q_i > 0$ and $p_i = 0$ has value $+\infty$.

Theorem E9.8 (Exact information decomposition)**.** *Let p and q be probability vectors on a finite partition $(I_\alpha)_\alpha$, and assume $q_i = 0$ whenever $p_i = 0$. Let*

$$m = Cp, \qquad n = Cq,$$

and let p_α and q_α be the conditional distributions inside packet I_α whenever the corresponding mass is positive. Then

$$D(q\|p) = D(n\|m) + \sum_{\alpha : n_\alpha > 0} n_\alpha D(q_\alpha \| p_\alpha).$$

The first term is the information retained by the coarse masses. The second term is the exact information carried by changes inside the packets.

Proof. For $i \in I_\alpha$, write

$$p_i = m_\alpha p_{\alpha,i}, \qquad q_i = n_\alpha q_{\alpha,i}.$$

Then

$$\log \frac{q_i}{p_i} = \log \frac{n_\alpha}{m_\alpha} + \log \frac{q_{\alpha,i}}{p_{\alpha,i}}.$$

Multiply by q_i, sum first inside each packet, and then sum over the packets. This is the finite relative entropy chain rule. Standard information theoretic treatments include [CT06]. □

Corollary E9.9 (Coarse data processing and equality)**.** *Under the assumptions of theorem E9.8,*

$$D(Cq\|Cp) \le D(q\|p).$$

Equality holds if and only if

$$q_\alpha = p_\alpha$$

for every packet with positive posterior mass.

Proof. Every conditional relative information term in theorem E9.8 is nonnegative. It vanishes exactly when the two conditional distributions agree. □

Corollary E9.10 (Information created by a Bayesian update)**.** *Let $q = \mathcal{B}_\lambda(p)$ with positive likelihood and positive prior. Then*

$$D(q\|p) = \sum_i q_i \log \lambda_i - \log Z_\lambda(p).$$

Inside a packet,

$$D(q_\alpha \| r_\alpha) = \sum_{i \in I_\alpha} q_{\alpha,i} \log \frac{\lambda_i}{\overline{\lambda}_\alpha(p)}.$$

Hence

$$D(q_\alpha \| r_\alpha) \le \log \frac{\max_{i \in I_\alpha} \lambda_i}{\min_{i \in I_\alpha} \lambda_i}.$$

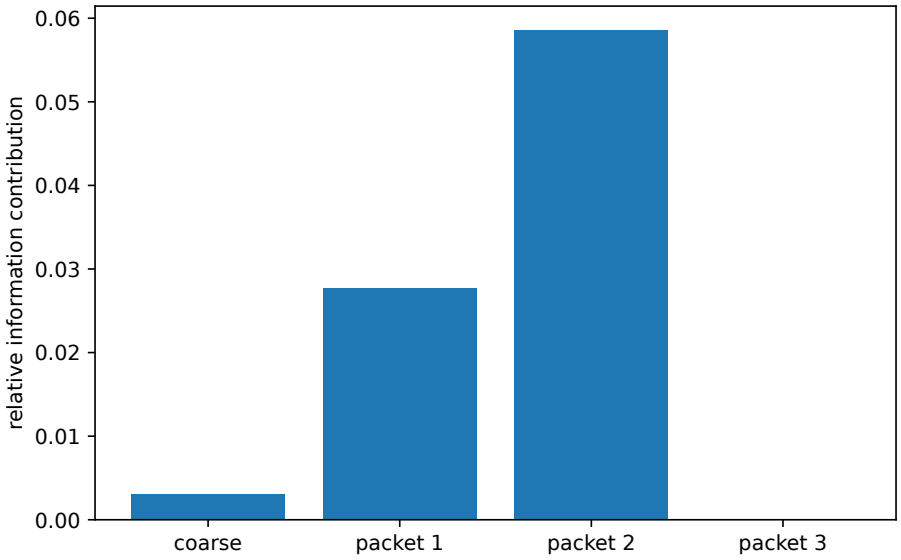


Figure E9. Relative information separated into a coarse packet contribution and conditional contributions inside the packets.

Proof. The posterior ratio is

$$\frac{q_i}{p_i} = \frac{\lambda_i}{Z_\lambda(p)}.$$

The local ratio is

$$\frac{q_{\alpha,i}}{r_{\alpha,i}} = \frac{\lambda_i}{\overline{\lambda}_\alpha(p)}.$$

Insert these identities into the definition of relative entropy. The final estimate follows because both λ_i and the packet average lie between the packet minimum and maximum. □

Corollary E9.11 (Information criterion for local sufficiency)**.** *For a fixed positive likelihood, the within packet term in theorem E9.8 vanishes for every positive prior if and only if the likelihood is constant on every packet.*

Proof. If the likelihood is block constant, every local posterior equals its local prior. Conversely, choose any positive prior. Vanishing relative entropy gives $q_\alpha = r_\alpha$. The local posterior formula then forces all likelihood values in that packet to equal their packet average. Repeat for every packet. □

Remark E9.12. The exact coarse likelihood in theorem E9.1 depends on the conditional prior inside each packet. Theorem E9.2 identifies when that dependence disappears. Theorem E9.8 measures what is lost when only the coarse posterior masses are retained.

E10 Bayesian flows and hidden Markov recursion

Fix a score vector $g \in \mathbb{R}^n$. For $t \geq 0$, define

$$\Phi_t(p)_i = \frac{\mathrm{e}^{tg_i} p_i}{\sum_j \mathrm{e}^{tg_j} p_j}.$$

Theorem E10.1 (Bayesian flow and replicator equation)**.** *The maps Φ_t form a nonlinear semigroup:*

$$\Phi_{s+t} = \Phi_s \circ \Phi_t.$$

Every coordinate satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t} p_i(t) = (g_i - \langle g, p(t)\rangle) p_i(t).$$

The support of $p(t)$ is independent of t, and for any two coordinates in the support,

$$\frac{p_i(t)}{p_j(t)} = \mathrm{e}^{t(g_i - g_j)} \frac{p_i(0)}{p_j(0)}.$$

Proof. The semigroup law follows from multiplication of the likelihoods e^{sg} and e^{tg}. Differentiating the normalized formula gives the replicator equation. A coordinate that starts at zero remains zero, while a positive coordinate remains positive. Dividing two explicit coordinate formulas gives the odds identity. □

Theorem E10.2 (Evidence cumulants and score monotonicity)**.** *Set*

$$\psi(t) = \log\left(\sum_i \mathrm{e}^{t g_i} p_i(0)\right).$$

Then

$$\psi'(t) = \langle g, p(t)\rangle, \qquad \psi''(t) = \mathrm{Var}_{p(t)}(g) \geq 0.$$

Consequently,

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle g, p(t)\rangle = \mathrm{Var}_{p(t)}(g).$$

Proof. Differentiate the log normalizing constant once and twice. The second derivative is the centered second moment under the tilted distribution. □

Corollary E10.3 (Concentration on the maximal score packet)**.** *Let*

$$G = \max\{g_i : p_i(0) > 0\}$$

and let M be the set of supported indices with score G. If M is the full support, then the conclusion below is immediate. Otherwise, suppose

$$\delta = \min\{G - g_i : p_i(0) > 0, \quad i \notin M\} > 0.$$

Then

$$1 - \sum_{i \in M} p_i(t) \leq \frac{1 - p_M(0)}{p_M(0)} \mathrm{e}^{-\delta t},$$

where $p_M(0) = \sum_{i \in M} p_i(0)$.

Proof. Factor e^{tG} from the normalizing constant. The total weight outside M is at most $(1 - p_M(0))\mathrm{e}^{-\delta t}$, while the weight on M is $p_M(0)$. □

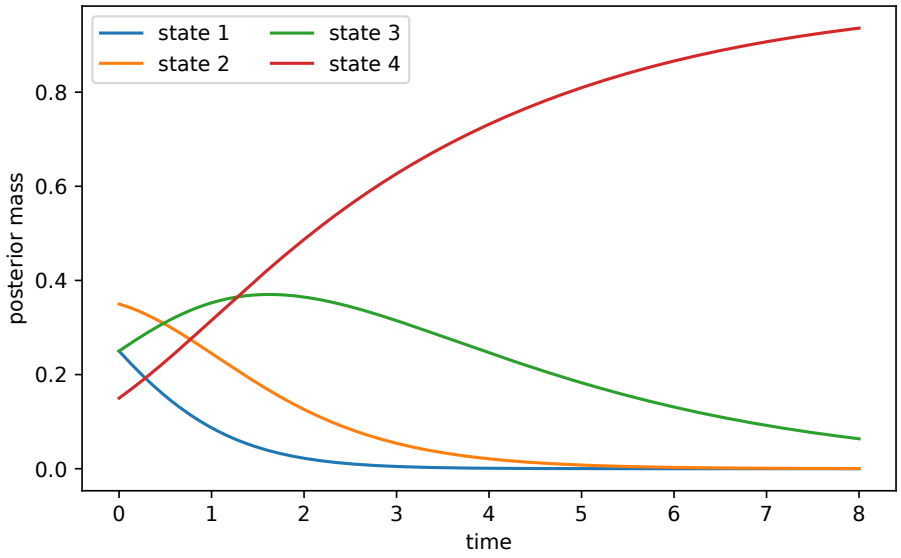


Figure E10. A finite Bayesian flow. Mass concentrates on the states with the largest score, while every relative odds ratio follows an exact exponential law.

E10.1 Prediction and update

Let K be column stochastic and let $\lambda^{(r)}$ be the likelihood of the rth observation. The hidden Markov filter is

$$\hat{p}_r = K p_{r-1}, \qquad p_r = \mathcal{B}_{\lambda^{(r)}}(\hat{p}_r).$$

A standard account of finite hidden Markov models appears in [Rab89].

Proposition E10.4 (Simplex preservation)**.** *Suppose p_{r-1} is a probability vector and the new evidence is positive. Prediction produces the probability vector $\hat{p}_r$, and normalization produces the probability vector p_r.*

Proof. Column stochasticity preserves nonnegativity and total mass. Bayesian normalization then preserves the same properties. □

Theorem E10.5 (One step filter stability)**.** *Let K and $\widetilde{K}$ be column stochastic. Assume*

$$0 < a \le \lambda_i \le b.$$

Then

$$\left\| \mathcal{B}_\lambda(Kp) - \mathcal{B}_\lambda(\widetilde{K}q) \right\|_1 \le \frac{2b}{a} \left(\|p - q\|_1 + \left\| K - \widetilde{K} \right\|_{1\to 1} \right).$$

Proof. Apply theorem E8.4 to Kp and $\widetilde{K}q$. Then use

$$\left\| Kp - \widetilde{K}q \right\|_1 \le \|K(p-q)\|_1 + \left\| (K - \widetilde{K})q \right\|_1$$

and the stochastic contraction property. □

Example E10.6 (Beta binomial transfer)**.** A beta prior with parameters (a, b) and binomial data containing s successes in n trials gives the posterior parameters

$$(a + s, b + n - s).$$

The update is additive in the two exponent coordinates. This is a finite dimensional conjugate realization of likelihood composition. Standard Bayesian references include [BS94].

E11 Fractional functional calculus

Let H be a finite dimensional Hilbert space and let $L = L^* \geq 0$. Write its spectral resolution as

$$L = \sum_j \lambda_j P_j.$$

For $\alpha > 0$, define

$$L^\alpha = \sum_j \lambda_j^\alpha P_j$$

and

$$T_\alpha(t) = \mathrm{e}^{-tL^\alpha} = \sum_j \mathrm{e}^{-t\lambda_j^\alpha} P_j.$$

Theorem E11.1 (Fractional packet calculus)**.** *For $\alpha, \beta > 0$,*

$$L^\alpha L^\beta = L^{\alpha+\beta}.$$

For $s, t \geq 0$,

$$T_\alpha(s)T_\alpha(t) = T_\alpha(s + t).$$

Every spectral packet $P_j H$ reduces L^α and $T_\alpha(t)$.

Proof. The claims hold on each spectral packet and the projectors are mutually orthogonal. □

Theorem E11.2 (Balakrishnan formula in finite dimension)**.** *For $0 < \alpha < 1$,*

$$L^\alpha = \frac{\sin(\pi\alpha)}{\pi} \int_0^\infty s^{\alpha-1} L(sI + L)^{-1} \, \mathrm{d}s.$$

The integral converges in operator norm.

Proof. It is enough to prove the formula on an eigenvector with eigenvalue $x \geq 0$. For $x > 0$, the substitution $s = xu$ gives

$$\frac{\sin(\pi\alpha)}{\pi} \int_0^\infty s^{\alpha-1} \frac{x}{s + x} \, \mathrm{d}s = x^\alpha \frac{\sin(\pi\alpha)}{\pi} \int_0^\infty \frac{u^{\alpha-1}}{1 + u} \, \mathrm{d}u = x^\alpha.$$

The value at $x = 0$ is zero. There are finitely many eigenvalues, so the operator integral converges. This is the finite spectral form of the fractional power construction in [Bal60]. □

Theorem E11.3 (Exact fractional naturality)**.** *Let L_f and L_c be finite dimensional positive self adjoint operators and let C satisfy*

$$CL_f = L_c C.$$

Then, for every $\alpha > 0$ and $t \geq 0$,

$$CL_f^\alpha = L_c^\alpha C$$

and

$$C\mathrm{e}^{-tL_f^\alpha} = \mathrm{e}^{-tL_c^\alpha} C.$$

Proof. The union of the two finite spectra is a finite set. There is a polynomial that agrees with x^α on that set, and another polynomial that agrees with e^{-tx^α}. Exact intertwining passes to every polynomial by induction. □

Theorem E11.4 (Fractional defect bound)**.** *Let $0 < \alpha < 1$. Suppose*

$$L_f \geq mI, \qquad L_c \geq mI$$

for some $m > 0$. Define

$$\Delta_L = CL_f - L_c C.$$

Then

$$\left\| CL_f^\alpha - L_c^\alpha C \right\| \leq \alpha m^{\alpha-1} \left\| \Delta_L \right\|.$$

Consequently,

$$\left\| C\mathrm{e}^{-tL_f^\alpha} - \mathrm{e}^{-tL_c^\alpha} C \right\| \leq \alpha t m^{\alpha-1} \left\| \Delta_L \right\|.$$

Proof. Set

$$R_f(s) = (sI + L_f)^{-1}, \qquad R_c(s) = (sI + L_c)^{-1}.$$

The resolvent identity gives

$$CR_f(s) - R_c(s)C = -R_c(s)\Delta_L R_f(s).$$

Since $LR = I - sR$,

$$CL_f R_f(s) - L_c R_c(s) C = sR_c(s)\Delta_L R_f(s).$$

Insert this into theorem E11.2. The spectral lower bound gives

$$\left\| R_f(s) \right\|, \left\| R_c(s) \right\| \leq \frac{1}{s+m}.$$

Therefore

$$\begin{aligned} \left\| CL_f^\alpha - L_c^\alpha C \right\| &\leq \frac{\sin(\pi\alpha)}{\pi} \left\| \Delta_L \right\| \int_0^\infty \frac{s^\alpha}{(s+m)^2} \, \mathrm{d}s \\ &= \alpha m^{\alpha-1} \left\| \Delta_L \right\|. \end{aligned}$$

The semigroup estimate follows from theorem E3.1, applied to the contraction generators $-L_f^\alpha$ and $-L_c^\alpha$. □

Remark E11.5. The factor $m^{\alpha-1}$ records the singular behavior of x^α near zero. Without a common positive lower bound, a uniform Lipschitz estimate of this form cannot be expected.

Standard Fact E11.6 (Stable subordination). Let $0 < \alpha < 1$ and $t > 0$. There is a probability measure $\eta_{\alpha,t}$ on $[0, \infty)$ such that

$$\int_0^\infty \mathrm{e}^{-s\lambda} \eta_{\alpha,t}(\,\mathrm{d}s) = \mathrm{e}^{-t\lambda^\alpha}$$

for every $\lambda \geq 0$. Consequently, for a finite dimensional positive self adjoint operator,

$$\mathrm{e}^{-tL^\alpha} = \int_0^\infty \mathrm{e}^{-sL} \eta_{\alpha,t}(\,\mathrm{d}s)$$

in operator norm. The measures satisfy the convolution law

$$\eta_{\alpha,s} * \eta_{\alpha,t} = \eta_{\alpha,s+t}.$$

The existence of $\eta_{\alpha,t}$ follows from complete monotonicity of $\mathrm{e}^{-t\lambda^\alpha}$. The operator identity is obtained by applying the scalar Laplace transform on every spectral packet. This is the finite spectral form of Bochner subordination. See [SSV12].

Theorem E11.7 (Poisson subordination). *For $t > 0$,*

$$\mathrm{e}^{-t\sqrt{L}} = \frac{t}{2\sqrt{\pi}} \int_0^\infty s^{-3/2} \mathrm{e}^{-t^2/(4s)} \mathrm{e}^{-sL}\,\mathrm{d}s.$$

The integral converges in operator norm.

Proof. On a spectral packet with eigenvalue λ, the formula reduces to

$$\mathrm{e}^{-t\sqrt{\lambda}} = \frac{t}{2\sqrt{\pi}} \int_0^\infty s^{-3/2} \mathrm{e}^{-t^2/(4s) - \lambda s}\,\mathrm{d}s.$$

This scalar identity follows from the Gaussian integral. Recombine the finitely many spectral packets. □

Remark E11.8. For sectorial and nonnormal operators, fractional powers require a resolvent or contour calculus together with domain control. The finite self adjoint theory proved here does not remove those analytic requirements. A broader account appears in [SSV12].

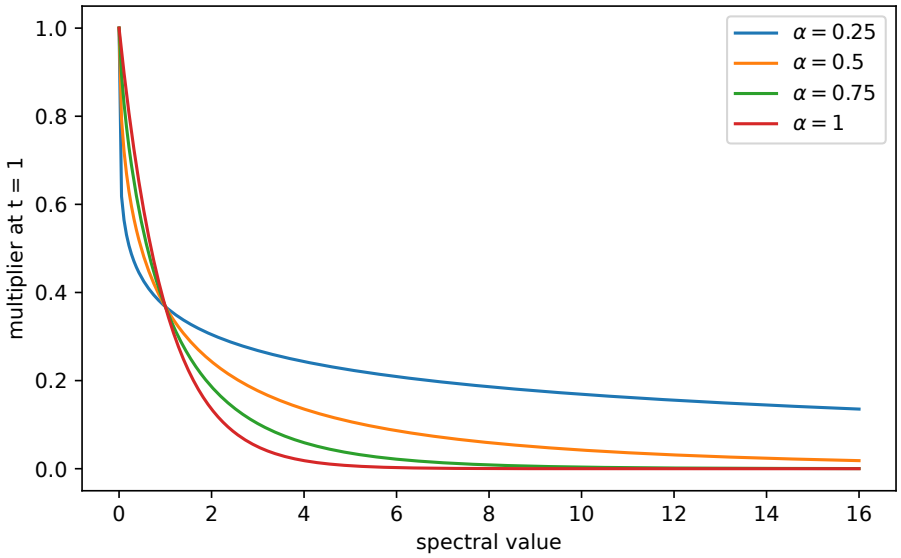


Figure E11. Fractional semigroup multipliers at time one. Smaller exponents preserve more of the high spectral range.

E12 Green operators and finite difference transfer

Consider

$$-u'' = f \quad \text{on } (0,1), \qquad u(0) = u(1) = 0.$$

The continuum Green kernel is

$$G(x,y) = \min(x,y) - xy.$$

Thus

$$u(x) = \int_0^1 G(x,y) f(y)\,\mathrm{d}y.$$

For n interior nodes, let

$$h = \frac{1}{n+1}, \qquad x_i = ih,$$

and define

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

Theorem E12.1 (Exact nodal Green reconstruction). *For $1 \le i, j \le n$,*

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

Proof. The inverse of the unscaled tridiagonal matrix has entry

$$\frac{\min(i,j)(n+1-\max(i,j))}{n+1}.$$

Multiplying by h^2 and dividing by h gives

$$h\min(i,j)\left(1-\frac{\max(i,j)}{n+1}\right)=\min(x_i,x_j)-x_ix_j.$$

□

Corollary E12.2 (Discrete Green quadrature)**.** *If u^h solves*

$$L_hu^h=f^h,$$

then every nodal value has the exact representation

$$u_i^h=h\sum_{j=1}^{n}G(x_i,x_j)f_j^h.$$

Thus the discrete Green operator is the nodal rectangle quadrature of the continuum Green kernel.

Proof. By theorem E12.1,

$$(L_h^{-1})_{ij}=hG(x_i,x_j).$$

Multiply the inverse matrix by the source vector. □

Proposition E12.3 (Discrete maximum bound)**.** *The inverse L_h^{-1} has nonnegative entries and*

$$\left\|L_h^{-1}\right\|_{\infty\to\infty}=\max_{1\le i\le n}\frac{x_i(1-x_i)}{2}\le\frac{1}{8}.$$

Proof. Nonnegativity follows from the entry formula in theorem E12.1. Let $v_i=x_i(1-x_i)/2$. Since the second difference is exact on quadratics,

$$L_hv=1.$$

Hence the ith row sum of L_h^{-1} is v_i. For a nonnegative matrix, the infinity operator norm is the largest row sum. □

Theorem E12.4 (Second order Poisson solution error)**.** *Suppose $u\in C^4[0,1]$ solves the Dirichlet problem above. Let*

$$f_i^h=f(x_i)$$

and let u^h solve

$$L_hu^h=f^h.$$

Then

$$\max_i\left|u_i^h-u(x_i)\right|\le\frac{h^2}{96}\left\|u^{(4)}\right\|_\infty.$$

Proof. Taylor expansion with remainder gives

$$\left\|L_h(u(x_i))_i-f^h\right\|_\infty\le\frac{h^2}{12}\left\|u^{(4)}\right\|_\infty.$$

Apply L_h^{-1} and use proposition E12.3. □

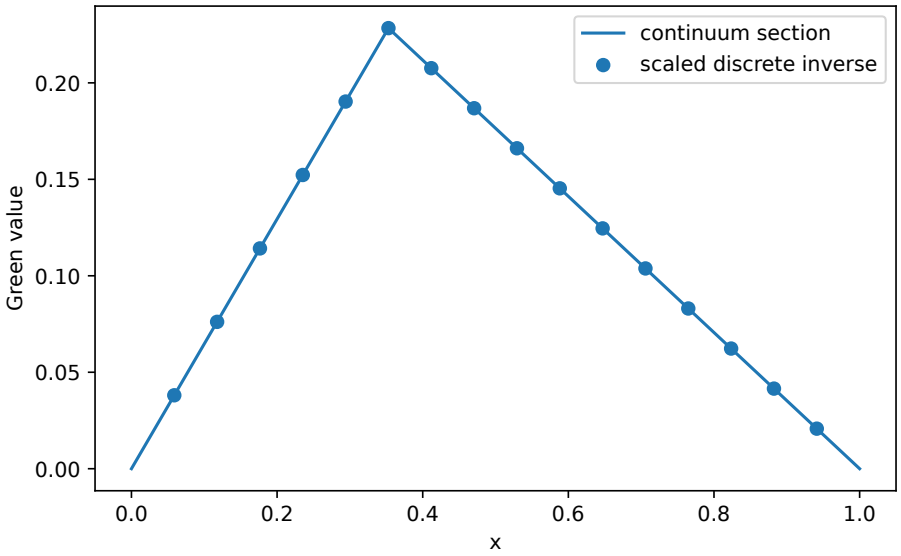


Figure E12. The continuum Dirichlet Green section and the scaled discrete inverse. The nodal values agree exactly.

The discrete eigenpairs are

$$\lambda_{k,h} = \frac{4}{h^2}\sin^2\left(\frac{k\pi h}{2}\right), \qquad v_{k,h}(i) = \sin(k\pi x_i).$$

Theorem E12.5 (Low mode eigenvalue control). *For $kh \le 1$,*

$$0 \le (k\pi)^2 - \lambda_{k,h} \le \frac{(k\pi)^4}{12}h^2.$$

Consequently, for every fixed k,

$$\lambda_{k,h} = (k\pi)^2 + O(h^2).$$

Proof. Set $x = k\pi h/2$. The condition $kh \le 1$ gives $0 \le x \le \pi/2$. On this interval,

$$\sin x \ge x - \frac{x^3}{6},$$

so

$$0 \le x^2 - \sin^2 x \le \frac{x^4}{3}.$$

Multiply by $4/h^2$. □

Corollary E12.6 (Heat and fractional mode transfer). *Let $\lambda_k = (k\pi)^2$. For $t \ge 0$,*

$$0 \le \mathrm{e}^{-t\lambda_{k,h}} - \mathrm{e}^{-t\lambda_k} \le \frac{t(k\pi)^4}{12}h^2\mathrm{e}^{-t\lambda_{k,h}}.$$

For $0 < \alpha \le 1$ and fixed k,

$$\lambda_{k,h}^{\alpha} = \lambda_k^{\alpha} + O(h^2)$$

and

$$\mathrm{e}^{-t\lambda_{k,h}^{\alpha}} = \mathrm{e}^{-t\lambda_k^{\alpha}} + O(th^2)$$

uniformly for t in bounded intervals.

Proof. The ordinary heat estimate follows from the mean value theorem and theorem E12.5. For fixed k, the eigenvalues stay uniformly away from zero once h is small. The function x^{α} is then Lipschitz on a common compact interval containing both eigenvalues. Apply the mean value theorem twice. □

Finite difference schemes and their stability theory are treated more broadly in [Str04]. The present result uses the special one dimensional Green structure to obtain an explicit constant.

E13 Finite state Feynman Kac transfer

How to read the result

The potential below is real. Its exponential is a potential weight in general, and a survival weight when the potential is nonnegative. This distinction does not change the finite transfer formula.

Let K be column stochastic. Let V be a real potential on the states and let

$$D_V = \operatorname{diag}(V).$$

For functions written as columns, the backward Markov operator is K^T.

Theorem E13.1 (Discrete Feynman Kac formula)**.** *Define*

$$M_V = \operatorname{diag}(\mathrm{e}^{-V})K^T.$$

For terminal data f,

$$(M_V^n f)_i = \mathbb{E}_i\left[\exp\left(-\sum_{r=0}^{n-1} V(X_r)\right) f(X_n)\right].$$

Proof. For one step,

$$(M_V f)_i = \mathrm{e}^{-V(i)} \sum_j K_{ji} f_j.$$

Iterating expands the matrix product over all paths. The product of transition entries is the path probability, and the diagonal factors give the exponential potential weight. □

Theorem E13.2 (Continuous time finite state formula)**.** *Let Q be a finite Markov generator and let $(X_t)_{t\geq 0}$ be the corresponding chain. Then*

$$\left(\mathrm{e}^{t(Q^T - D_V)} f\right)_i = \mathbb{E}_i\left[\exp\left(-\int_0^t V(X_s)\,\mathrm{d}s\right) f(X_t)\right].$$

If $V \geq v_0$, then

$$\left\|\mathrm{e}^{t(Q^T - D_V)}\right\|_{\infty\to\infty} \leq \mathrm{e}^{-v_0 t}.$$

Proof. Conditioning over the first small time interval gives the backward equation

$$\dot{u} = (Q^T - D_V)u, \qquad u(0) = f.$$

The finite system has a unique solution, namely the displayed matrix exponential. The expectation bound follows from

$$\exp\left(-\int_0^t V(X_s)\,\mathrm{d}s\right) \le \mathrm{e}^{-v_0 t}.$$

This finite argument is the matrix form of the Feynman Kac principle originating in [Kac49]. □

Theorem E13.3 (Alternating propagation and killing)**.** *For every $t \ge 0$,*

$$\mathrm{e}^{t(Q^T - D_V)} = \lim_{n\to\infty} \left(\mathrm{e}^{-tD_V/n}\mathrm{e}^{tQ^T/n}\right)^n$$

in operator norm. Each finite product alternates one Markov propagation step with one diagonal survival weight.

Proof. Set $B = -D_V$ and $C = Q^T$. For bounded matrices,

$$\mathrm{e}^{tB/n}\mathrm{e}^{tC/n} = I + \frac{t}{n}(B + C) + O(n^{-2})$$

uniformly in operator norm. The standard telescoping comparison with $(I + t(B + C)/n)^n$ shows that the product converges to $\mathrm{e}^{t(B+C)}$. This is the displayed factor order, so the limit follows. □

Corollary E13.4 (Positivity and composition)**.** *The operators*

$$F_V(t) = \mathrm{e}^{t(Q^T - D_V)}$$

form a positive semigroup. If $f \ge 0$, then $F_V(t)f \ge 0$.

Proof. Both claims follow from the expectation formula. □

E14 Schur Feshbach elimination

Let $X = X_P \oplus X_Q$ and write

$$A = \begin{pmatrix} A_{PP} & A_{PQ} \\ A_{QP} & A_{QQ} \end{pmatrix}.$$

For $zI - A_{QQ}$ invertible, define

$$\Sigma_P(z) = A_{PQ}(zI - A_{QQ})^{-1}A_{QP}$$

and

$$F_P(z) = zI - A_{PP} - \Sigma_P(z).$$

Theorem E14.1 (Schur Feshbach resolvent). *Assume $zI - A_{QQ}$ is invertible. Then $zI - A$ is invertible if and only if $F_P(z)$ is invertible. In that case,*

$$P(zI - A)^{-1}P = F_P(z)^{-1}.$$

In finite dimension,

$$\det(zI - A) = \det(zI - A_{QQ}) \det F_P(z).$$

Proof. Set

$$L_z = \begin{pmatrix} I & -A_{PQ}(zI - A_{QQ})^{-1} \\ 0 & I \end{pmatrix}$$

and

$$S_z = \begin{pmatrix} F_P(z) & 0 \\ -A_{QP} & zI - A_{QQ} \end{pmatrix}.$$

Direct multiplication gives $zI - A = L_z S_z$. The first factor is invertible. The claims follow from triangular inversion and the determinant of the second factor. This is the finite block form of the Feshbach reduction [Fes58]. □

E14.1 Time domain memory

For the homogeneous equation $\dot{x} = Ax$, variation of constants gives

$$x_Q(t) = \mathrm{e}^{tA_{QQ}} x_Q(0) + \int_0^t \mathrm{e}^{(t-s)A_{QQ}} A_{QP} x_P(s) \, \mathrm{d}s.$$

Substitution into the retained equation gives

$$\begin{aligned} \dot{x}_P(t) = A_{PP} x_P(t) + A_{PQ} \mathrm{e}^{tA_{QQ}} x_Q(0) \\ + \int_0^t \mathcal{K}(t-s) x_P(s) \, \mathrm{d}s, \end{aligned}$$

where

$$\mathcal{K}(t) = A_{PQ} \mathrm{e}^{tA_{QQ}} A_{QP}.$$

Theorem E14.2 (Memory and self energy). *If z lies to the right of the spectrum of A_{QQ}, then*

$$\int_0^\infty \mathrm{e}^{-zt} \mathcal{K}(t) \, \mathrm{d}t = \Sigma_P(z).$$

If, for constants $M \geq 0$ and $\gamma > 0$,

$$\left\| \mathrm{e}^{tA_{QQ}} \right\| \leq M \mathrm{e}^{-\gamma t},$$

then

$$\left\| \mathcal{K}(t) \right\| \leq M \left\| A_{PQ} \right\| \left\| A_{QP} \right\| \mathrm{e}^{-\gamma t}$$

and

$$\left\| \mathcal{K} \right\|_{L^1(0,\infty)} \leq \frac{M}{\gamma} \left\| A_{PQ} \right\| \left\| A_{QP} \right\|.$$

Proof. If $X_Q = \{0\}$, the memory kernel and self-energy are both zero, so the identity is immediate. Otherwise write $B = A_{QQ}$, which is a bounded operator on the hidden Banach space, and put $s(B) = \max \operatorname{Re} \operatorname{Spec}(B)$. Choose a real β with $s(B) < \beta < \operatorname{Re} z$. Spectral mapping and the spectral-radius formula give $r(\mathrm{e}^B) = \mathrm{e}^{s(B)} < \mathrm{e}^\beta$, hence $\|\mathrm{e}^{nB}\| \le M_\beta \mathrm{e}^{\beta n}$ for every integer $n \ge 0$, after increasing M_β to cover the finitely many initial values. Write $t = n + r$, where $0 \le r < 1$. Then

$$\|\mathrm{e}^{tB}\| \le M_\beta \exp(\|B\| + \max(0, -\beta))\mathrm{e}^{\beta t}.$$

Thus $\mathrm{e}^{-zt}\mathrm{e}^{tB}$ is Bochner integrable and tends to zero. Integrating its derivative gives both inverse identities for $R = \int_0^\infty \mathrm{e}^{-zt}\mathrm{e}^{tB}\,\mathrm{d}t$, so $R = (zI - B)^{-1}$. This is the bounded-generator growth principle in [EN00, Chapter IV, Corollary 3.12]. Multiplication by the coupling blocks gives the first identity. The remaining estimates follow directly from the assumed positive decay rate. □

Theorem E14.3 (Criterion for exact local closure)**.** *Assume $x_Q(0) = 0$ and $d = \dim X_Q < \infty$. The following conditions are equivalent.*

1. *For every retained initial state,*
$$P\mathrm{e}^{tA}P = \mathrm{e}^{tA_{PP}}P$$
for all $t \ge 0$.
2. *$\Sigma_P(z) = 0$ for every sufficiently large z.*
3. *$\mathcal{K}(t) = 0$ for every $t \ge 0$.*
4. *If $d = \dim X_Q$, then*
$$A_{PQ}A_{QQ}^k A_{QP} = 0$$
for $0 \le k < d$.

Proof. By theorem E2.5, the Laplace transform of the projected propagator is the projected resolvent. By theorem E14.1, the first condition is equivalent to

$$(zI - A_{PP} - \Sigma_P(z))^{-1} = (zI - A_{PP})^{-1}$$

for all sufficiently large z, which is equivalent to the second condition. The second and third conditions are equivalent by theorem E14.2 and uniqueness of the Laplace transform.

For large z,

$$\Sigma_P(z) = \sum_{k=0}^{\infty} \frac{A_{PQ}A_{QQ}^k A_{QP}}{z^{k+1}}.$$

Thus the second condition implies the moment identities. Conversely, the first d identities imply all later ones by the Cayley Hamilton theorem, so the series vanishes. □

Remark E14.4. A finite hidden block produces a strictly proper rational self energy. It tends to zero as z tends to infinity. Hence a nonzero constant self energy cannot arise from hidden dynamics alone. A constant correction must already be included in the retained generator.

E15 Memory modes and controlled compression

How to read the result

Stable hidden modes may carry polynomial factors when a Jordan block is present. A spectral gap controls eventual decay, while nonnormal coupling can produce substantial intermediate memory. The compression certificate below is optimal for its declared primary-block selection problem.

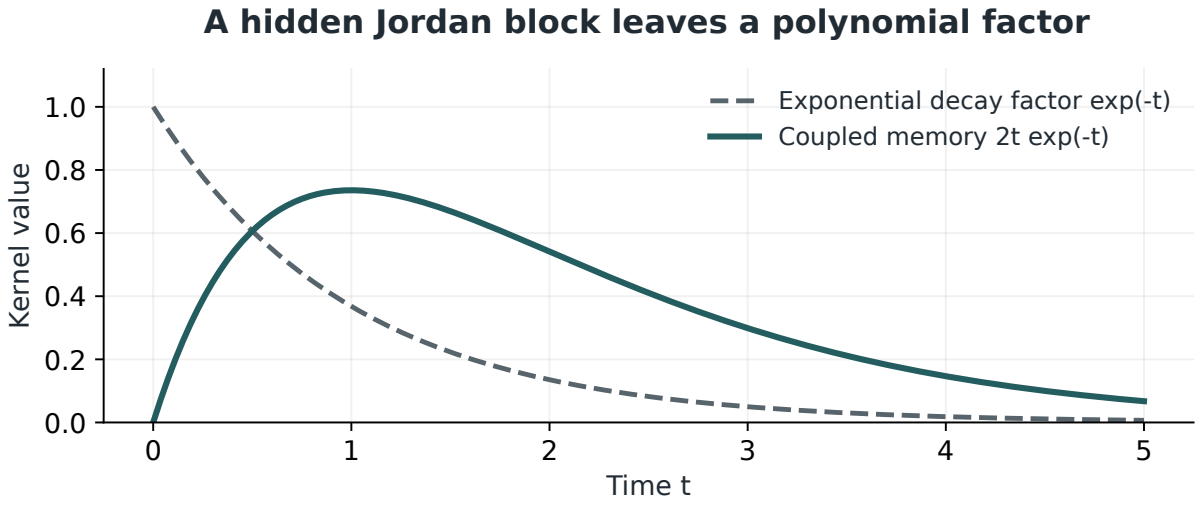


Figure E13. A two-coordinate Jordan hidden block has an exponential decay factor and a polynomial factor. The nilpotent polynomial is checked exactly at the displayed rational times. The exponential curve is a plotted illustration of the formula.

Finite hidden systems produce rational self energies and finite modal memory. This is the state space viewpoint familiar from linear systems theory [Kai80]. The same formulas also underlie the Mori Zwanzig representation of reduced dynamics [Zwa61, CHK00].

Theorem E15.1 (Polynomial exponential memory of a finite hidden system)**.** *Let X_Q be a finite dimensional complex space. Let $\mu_1, \dots, \mu_r$ be the distinct eigenvalues of A_{QQ}, let Q_j be the corresponding primary spectral projections, and set*

$$N_j = (A_{QQ} - \mu_j I)Q_j.$$

Let ν_j be the least positive integer for which $N_j^{\nu_j} = 0$, and define

$$B_{j,k} = A_{PQ} N_j^k Q_j A_{QP} \qquad (0 \le k < \nu_j).$$

Then

$$\mathcal{K}(t) = \sum_{j=1}^{r} \mathrm{e}^{\mu_j t} \sum_{k=0}^{\nu_j - 1} \frac{t^k}{k!} B_{j,k}$$

and

$$\Sigma_P(z) = \sum_{j=1}^{r} \sum_{k=0}^{\nu_j - 1} \frac{B_{j,k}}{(z - \mu_j)^{k+1}}.$$

Suppose $\operatorname{Re}\mu_j < 0$ *for every* j*. If* $J \subseteq \{1,\dots,r\}$ *and* $\mathcal{K}_J$ *retains the complete primary blocks indexed by* J*, then*

$$\|\mathcal{K} - \mathcal{K}_J\|_{L^1(0,\infty)} \le \sum_{j\notin J} \sum_{k=0}^{\nu_j-1} \frac{\|B_{j,k}\|}{(-\operatorname{Re}\mu_j)^{k+1}}.$$

Proof. The primary decomposition gives

$$A_{QQ}Q_j = \mu_j Q_j + N_j$$

with commuting scalar and nilpotent parts on $Q_j X_Q$. Therefore

$$\mathrm{e}^{tA_{QQ}}Q_j = \mathrm{e}^{\mu_j t}\sum_{k=0}^{\nu_j-1}\frac{t^k}{k!}N_j^k Q_j.$$

Multiplication by A_{PQ} and A_{QP} gives the memory formula. On the jth primary space, the nilpotent resolvent expansion is

$$(zI - A_{QQ})^{-1}Q_j = \sum_{k=0}^{\nu_j-1}\frac{N_j^k Q_j}{(z-\mu_j)^{k+1}}.$$

Multiplication by the coupling blocks gives the self energy formula for every z outside the hidden spectrum. Equivalently, on a right half plane it is the Laplace transform identity

$$\int_0^\infty \mathrm{e}^{-zt}\mathrm{e}^{\mu_j t}\frac{t^k}{k!}\,\mathrm{d}t = \frac{1}{(z-\mu_j)^{k+1}}.$$

Under stability,

$$\int_0^\infty \mathrm{e}^{\operatorname{Re}\mu_j t}\frac{t^k}{k!}\,\mathrm{d}t = \frac{1}{(-\operatorname{Re}\mu_j)^{k+1}}.$$

Sum the omitted terms. □

Corollary E15.2 (Optimal selection for the certified primary block bound)**.** *Under the stable assumptions of theorem E15.1, define the primary block weights*

$$w_j = \sum_{k=0}^{\nu_j-1}\frac{\|B_{j,k}\|}{(-\operatorname{Re}\mu_j)^{k+1}}.$$

Among all sets J *containing* k *primary blocks, the certified upper bound*

$$\sum_{j\notin J} w_j$$

is minimized by retaining the k *largest weights. If*

$$w_{(1)} \ge w_{(2)} \ge \cdots \ge w_{(r)},$$

then the least number of retained primary blocks certified to give L^1 *error at most* ε *is the least* k *for which*

$$\sum_{j>k} w_{(j)} \le \varepsilon.$$

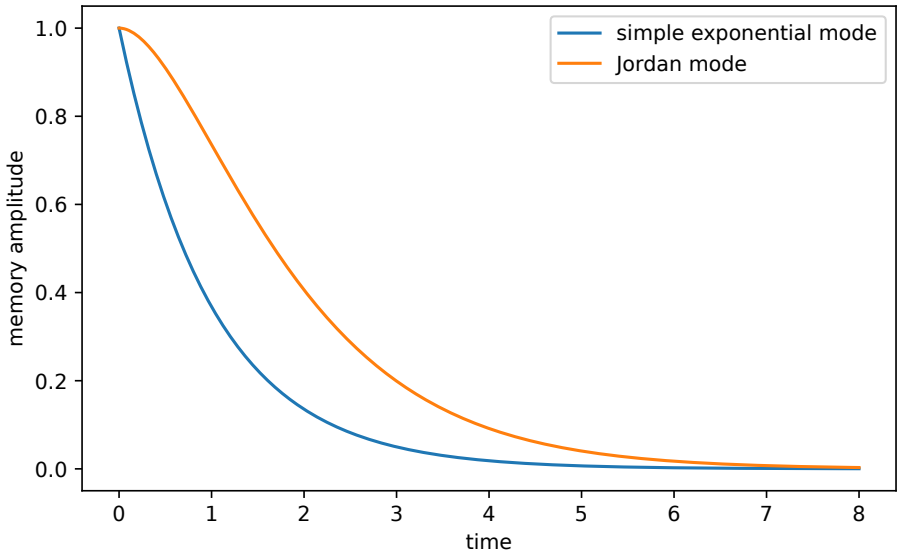


Figure E14. A stable Jordan block produces a polynomial exponential memory mode. A diagonal hidden block produces only pure exponentials.

Proof. If a retained set contains a block with smaller weight than an omitted block, exchanging them reduces the omitted weight sum. Repeating the exchange gives the set of the k largest weights. The tolerance statement is the corresponding sorted tail condition. □

Corollary E15.3 (Normal hidden dynamics)**.** *Suppose X_Q is a finite dimensional Hilbert space. If A_{QQ} is normal with spectral resolution*

$$A_{QQ} = \sum_{j=1}^{r} \mu_j Q_j,$$

then

$$\mathcal{K}(t) = \sum_{j=1}^{r} \mathrm{e}^{\mu_j t} B_j, \qquad \Sigma_P(z) = \sum_{j=1}^{r} \frac{B_j}{z - \mu_j},$$

where

$$B_j = A_{PQ} Q_j A_{QP}.$$

If $\operatorname{Re} \mu_j < 0$, then

$$\|\mathcal{K} - \mathcal{K}_J\|_{L^1(0,\infty)} \leq \sum_{j \notin J} \frac{\|B_j\|}{-\operatorname{Re} \mu_j}.$$

Proof. Normality removes every nilpotent part, so $\nu_j = 1$. Apply theorem E15.1. □

Corollary E15.4 (Pole order and hidden Jordan depth)**.** *Every finite hidden block produces a strictly proper rational self energy. At an eigenvalue μ_j, its pole order is at most the size of the largest Jordan block of A_{QQ} at μ_j. Equality holds when the corresponding highest coefficient $B_{j,\nu_j - 1}$ is nonzero.*

Proof. Read the pole order from the partial fraction expansion in theorem E15.1. □

Corollary E15.5 (Finite auxiliary state realization)**.** *Every kernel of the form*

$$\mathcal{K}(t) = B \mathrm{e}^{tD} C$$

can be represented exactly by adjoining the finite auxiliary state

$$\dot{y} = Dy + Cx_P, \qquad \dot{x}_P = A_{PP}x_P + By, \qquad y(0) = 0.$$

A nonzero auxiliary initial state adds the explicit forcing $Be^{tD}y(0)$. The self energy of the augmented system is

$$B(zI - D)^{-1}C.$$

Proof. Eliminate y by variation of constants or apply theorem E14.1 to the augmented block matrix. □

Theorem E15.6 (Volterra stability under kernel compression)**.** *Let $S(t)$ be a strongly continuous semigroup on a Banach space X, fix $T > 0$, and assume*

$$\|S(t)\| \le M_T \quad (0 \le t \le T).$$

Let $K, \widetilde{K} \in L^1((0,T), \mathcal{B}(X))$, let $g \in L^1((0,T), X)$, and let $x, \tilde{x} \in C([0,T], X)$. Suppose these are mild solutions with the same initial state and forcing, except that K is replaced by $\widetilde{K}$:

$$x(t) = S(t)x_0 + \int_0^t S(t-s)g(s)\,\mathrm{d}s + \int_0^t S(t-s)\int_0^s K(s-r)x(r)\,\mathrm{d}r\,\mathrm{d}s.$$

Assume

$$\sup_{0\le t\le T}\|\tilde{x}(t)\| \le B.$$

Set

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

Then

$$\sup_{0\le t\le T}\|x(t) - \tilde{x}(t)\| \le M_T BT\varepsilon \exp(M_T LT).$$

Proof. Let

$$E(t) = \sup_{0\le r\le t}\|x(r) - \tilde{x}(r)\|.$$

Subtract the two mild equations. Fubini's theorem and the triangle inequality give

$$E(t) \le M_T L\int_0^t E(r)\,\mathrm{d}r + M_T Bt\varepsilon.$$

For $t \le T$, replace the last factor t by T and apply Gronwall's inequality. □

Corollary E15.7 (Trajectory bound for modal truncation)**.** *Under the assumptions of theorems E15.1 and E15.6, replacing $\mathcal{K}$ by $\mathcal{K}_J$ changes the retained trajectory by at most*

$$M_T BT\exp(M_T LT)\sum_{j\notin J}\sum_{k=0}^{\nu_j-1}\frac{\|B_{j,k}\|}{(-\operatorname{Re}\mu_j)^{k+1}}.$$

Proof. Insert the L^1 error from theorem E15.1 into theorem E15.6. □

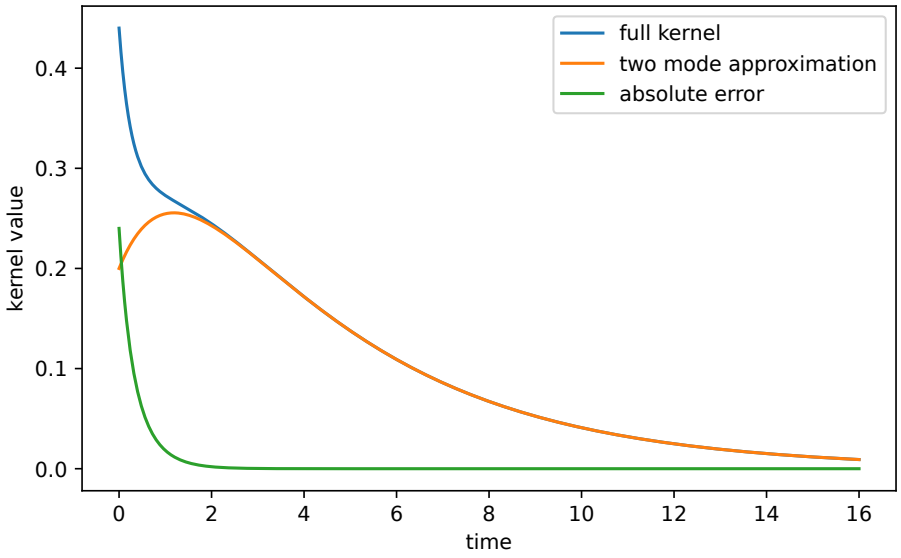


Figure E15. Full and truncated memory kernels with their pointwise difference. The bound uses integrated error.

Remark E15.8. A finite hidden realization is exact for rational self energy. A general memory kernel may require infinitely many modes or a continuum of modes. The compression theorem gives a stability estimate once an L^1 approximation has been supplied. The largest weight rule is optimal for the certified blockwise bound. It need not minimize the true kernel error when different modal terms cancel.

E16 Survival, moments, and competing risks

Let A be a transient state subgenerator. Its off diagonal entries are nonnegative and its column sums are nonpositive. Let p_0 be a transient probability vector. Define

$$p(t) = \mathrm{e}^{tA} p_0, \qquad S(t) = 1^T p(t), \qquad a^T = -1^T A.$$

The row vector a^T contains the exit rates.

Theorem E16.1 (State dependent killing as survival). *Let Q be a finite Markov generator, let $V \geq 0$, and set*

$$A = Q - D_V.$$

For an initial distribution p_0, define

$$S_V(t) = 1^T \mathrm{e}^{tA} p_0.$$

Then

$$S_V(t) = \mathbb{E}_{p_0}\left[\exp\left(-\int_0^t V(X_s)\,\mathrm{d}s\right)\right].$$

Also,

$$-S_V'(t) = 1^T D_V \mathrm{e}^{tA} p_0.$$

The matrix $A = Q - D_V$ has nonnegative off diagonal entries and nonpositive column sums, so it is the subgenerator of the killed, possibly nontransient chain. Thus the potential in the finite Feynman Kac formula

is a state dependent killing rate in the forward sub-Markov semigroup. Almost-sure absorption follows under the additional spectral stability hypothesis used below.

Proof. Average theorem E13.2 with terminal function $f = 1$ against the initial distribution. Transposition gives

$$p_0^T \mathrm{e}^{t(Q^T - D_V)} 1 = 1^T \mathrm{e}^{t(Q - D_V)} p_0.$$

Since $1^T Q = 0$, differentiation gives

$$S_V'(t) = -1^T D_V \mathrm{e}^{tA} p_0.$$

□

Theorem E16.2 (Survival and hazard identities)**.** *The event density and hazard are*

$$f(t) = a^T \mathrm{e}^{tA} p_0 = -S'(t), \qquad h(t) = \frac{f(t)}{S(t)}$$

whenever $S(t) > 0$. The survival function is nonincreasing and

$$S(t) = \exp\left(- \int_0^t h(s)\, \mathrm{d}s \right).$$

Proof. Because A is a Metzler matrix, e^{tA} is nonnegative. The column sum assumption makes a^T nonnegative. Differentiate $S(t)$:

$$S'(t) = 1^T A \mathrm{e}^{tA} p_0 = -a^T \mathrm{e}^{tA} p_0.$$

The remaining identity follows by integrating $(\log S)' = -h$. □

Theorem E16.3 (Laplace transform and moments)**.** *Assume the spectrum of A lies in the open left half plane. Then absorption occurs with probability one. For $s \geq 0$,*

$$\mathbb{E}[\mathrm{e}^{-s\tau}] = a^T (sI - A)^{-1} p_0.$$

For every integer $m \geq 1$,

$$\mathbb{E}[\tau^m] = m! 1^T (-A)^{-m} p_0.$$

Proof. The Laplace transform follows by integrating the density in theorem E16.2. Stability of A gives

$$\int_0^\infty \mathrm{e}^{-st} \mathrm{e}^{tA}\, \mathrm{d}t = (sI - A)^{-1}.$$

At $s = 0$,

$$a^T (-A)^{-1} p_0 = 1^T p_0 = 1.$$

For the moments,

$$\int_0^\infty t^m \mathrm{e}^{tA}\, \mathrm{d}t = m! (-A)^{-m-1}.$$

Multiplying by $a^T = 1^T(-A)$ gives the result. These are the standard matrix formulas for phase type laws [Neu81, ANO96]. □

Proposition E16.4 (Residual life as a restarted semigroup)**.** *For $S(t) > 0$, define the conditional transient distribution*

$$p_t = \frac{\mathrm{e}^{tA} p_0}{S(t)}.$$

Then

$$\frac{S(t+s)}{S(t)} = 1^T \mathrm{e}^{sA} p_t.$$

Proof. Use the semigroup law in the numerator $S(t+s) = 1^T \mathrm{e}^{sA} \mathrm{e}^{tA} p_0$. □

Theorem E16.5 (Competing risk packets)**.** *Suppose the exit row splits into nonnegative cause rows*

$$a^T = \sum_{r=1}^{q} a_r^T.$$

Define

$$f_r(t) = a_r^T \mathrm{e}^{tA} p_0.$$

Then

$$f(t) = \sum_{r=1}^{q} f_r(t).$$

The cause specific hazard is

$$h_r(t) = \frac{f_r(t)}{S(t)},$$

and

$$h(t) = \sum_{r=1}^{q} h_r(t).$$

If the spectrum of A lies in the open left half-plane, the eventual probability of cause r is

$$\int_0^\infty f_r(t)\,\mathrm{d}t = a_r^T (-A)^{-1} p_0.$$

Proof. Every identity follows by linearity and the stable resolvent integral. □

Remark E16.6. A finite phase type hazard need not be monotone. The semigroup formulas guarantee nonnegative density and decreasing survival, not monotonicity of the hazard.

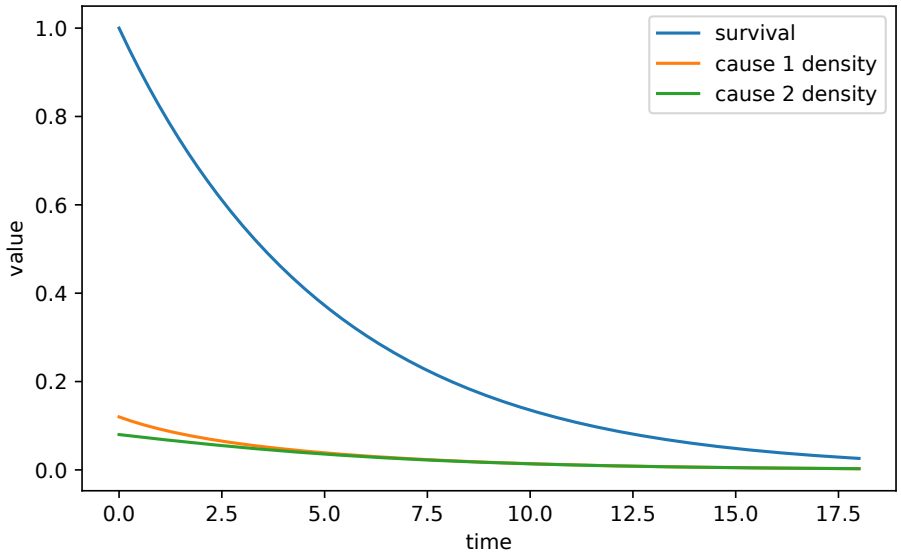


Figure E16. Survival and two cause specific event densities from one transient semigroup.

E17 Finite illustrations and computational checks

The figures in the article use small matrices chosen to make each mechanism visible. They are not evidence for claims outside the proved finite setting. The accompanying verification script recomputes the following identities with independently generated data.

1. Packet block convolution, the block differential system, and the Laplace resolvent formula.
2. The Duhamel and resolvent defect identities.
3. Exact and perturbed Markov lumping.
4. Conditional expectation, tower laws, martingale packet energy, and Hermite multipliers.
5. Bayesian reconstruction, sufficiency, local distortion, and information decomposition.
6. The evidence variance identity along a Bayesian flow.
7. The Balakrishnan formula and the fractional defect estimate.
8. Nodal Green reconstruction and the Poisson error bound.
9. Discrete Feynman Kac path enumeration.
10. The Schur Feshbach resolvent and memory transform.
11. Jordan memory modes, modal truncation, and the Volterra stability estimate.
12. Feynman Kac killing, survival transforms, moments, and competing risk masses.

Identity	Typical residual
Packet block convolution	below 3×10^{-14}
Duhamel defect	below 3×10^{-12}
Markov lumping	below 3×10^{-14}
Bayesian reconstruction	below 3×10^{-15}
Fractional spectral law	below 3×10^{-12}
Nodal Green identity	below 3×10^{-13}
Feynman Kac enumeration	below 3×10^{-14}
Schur Feshbach resolvent	below 3×10^{-12}
Memory Laplace identity	below 3×10^{-11}
Survival moment formulas	below 3×10^{-11}

Table E2. Numerical tolerances in the finite verification suite. The proofs do not depend on these computations.

E18 Limits and further questions

The bounded inverse limit theorem isolates one clean route to continuum evolution. Unbounded generators require graph norms, domain compatibility, and a generation theorem. The range leakage theorem converts the closure quantity into an off diagonal generator estimate for bounded orthogonal refinement models. Concrete discretizations still require a uniform estimate for that off diagonal block.

The finite Bayesian sufficiency theorem is exact, but it concerns one likelihood and one fixed partition. The relative entropy chain rule measures both retained and discarded information for that update. Statistical sufficiency for a parametric family requires a common packet structure that preserves the full parameter dependence. In continuous models the corresponding object is a sufficient sigma algebra or statistic.

The fractional defect theorem uses a common spectral gap. Removing that gap requires weaker norms or a Hölder estimate near the null space. For sectorial operators, the same calculation should be carried out through resolvent bounds on a sector, with domains tracked explicitly. The finite spectral statement does not settle those issues.

The Green and Poisson results exploit a one dimensional second difference matrix. Higher dimensions, irregular meshes, and variable coefficients require a separate stability theory. The packet language remains useful only after coercivity, consistency, and boundary approximation have been established.

Memory compression is exact once a finite hidden realization is known and quantitative once an integrable kernel approximation is supplied. The primary decomposition shows that a finite hidden block always produces polynomial exponential modes, including the effects of Jordan chains. Choosing a short approximation remains a different problem. Strongly nonnormal systems can exhibit transient growth and may require balanced or rational approximation methods. General Volterra kernels need not admit a short finite realization.

Finally, Markov lumping, Bayesian sufficiency, and memory closure answer different questions. Lumpability asks whether coarse probabilities evolve autonomously. Bayesian sufficiency asks whether a fixed coarse likelihood reproduces the posterior packet masses without hiding a change in the conditional distribution inside a packet. Memory closure asks whether discarded dynamics can be represented without history. Similar notation should not obscure these distinctions.

E19 Conclusion

A bounded semigroup becomes a matrix of packet transfers once a finite packet system is fixed. The matrix obeys convolution in time, a closed differential system, a Laplace resolvent formula, and a directed path expansion. A generator defect propagates through Duhamel's formula and through the resolvent calculus. Exact compatibility gives a semigroup on the bounded inverse limit. Summable adjacent defects give convergent reconstructed evolutions. Range leakage, and in particular an off diagonal generator block, determines whether the limit retains the semigroup law.

In probability, the same structure appears in several forms. Markov transfer is linear and exact lumpability is the vanishing of a canonical kernel obstruction. Conditional expectation is an orthogonal stochastic projector whose differences form martingale packets. The Ornstein Uhlenbeck semigroup resolves exactly into Hermite packets. Bayesian transfer is nonlinear because of evidence normalization. Its exact two stage decomposition separates coarse packet evidence from local conditional change. Relative entropy then records exactly how much information lies at each level. A fixed sufficient coarse description exists for every prior precisely when the likelihood is constant inside each packet. The common level sets of a likelihood family give its unique coarsest fixed sufficient partition.

Positive fractional powers preserve exact intertwining. Under a common spectral gap, the Balakrishnan formula turns an ordinary generator defect into an explicit fractional defect. Green operators and finite differences show how source reconstruction fits the same transfer language. Feynman Kac weights Markov paths by a potential without leaving the semigroup framework.

Hidden packets provide the final structural lesson. Exact elimination does not erase them. It returns them as a rational self energy in the resolvent and a causal memory kernel in time. The primary decomposition of the hidden generator produces every finite memory kernel as a finite sum of polynomial exponential modes. Stable blocks can be truncated with an integrated error certificate, and the Volterra estimate transfers that certificate to the retained trajectory. State dependent killing identifies the Feynman Kac potential with a survival mechanism. Survival and competing risks are then scalar reconstructions of a transient semigroup.

The article therefore separates four operations that are often compressed into one informal idea of coarse dynamics:

$$\begin{aligned} \text{intertwining} &\longrightarrow \text{refinement control} \\ &\longrightarrow \text{statistical sufficiency} \\ &\longrightarrow \text{memory closure.} \end{aligned}$$

Each arrow has its own hypotheses and its own obstruction.

Part III Retained results and operator calculus

R1 Transfer paths and product-topology limits

Let $T(t)$ be an arbitrary strongly continuous semigroup on a complex Banach space X, with generator A, and suppose $\|T(t)\| \le Me^{\omega t}$ for $M \ge 1$ and $\omega \in \mathbb{R}$. Let $(P_\alpha)_{\alpha\in I_0}$ be a finite family of bounded pairwise orthogonal idempotents summing to I, and put $T_{\beta\alpha}(t) = P_\beta T(t) P_\alpha$. The generator need not be bounded. In the following resolvent statement the integral is strong: apply the integrand to each $x \in X$ and take its Bochner integral in X. It is not asserted to be a Bochner integral in the operator-norm space $\mathcal{B}(X)$. The general Laplace formula is [EN00, Chapter II, Theorem 1.10].

Theorem R1.1 (Packet convolution). *For all $s, t \ge 0$ and all packets α, β,*

$$\boxed{T_{\beta\alpha}(s+t) = \sum_{\gamma\in I_0} T_{\beta\gamma}(s) T_{\gamma\alpha}(t).}$$

The original semigroup is reconstructed from its blocks by

$$T(t) = \sum_{\alpha,\beta\in I_0} T_{\beta\alpha}(t).$$

Proof. Insert the identity between the two semigroup factors:

$$\begin{aligned} P_\beta T(s+t) P_\alpha &= P_\beta T(s) \left(\sum_\gamma P_\gamma\right) T(t) P_\alpha \\ &= \sum_\gamma P_\beta T(s) P_\gamma T(t) P_\alpha. \end{aligned}$$

The reconstruction formula follows by inserting the packet identity on both sides of $T(t)$. □

Proposition R1.2 (Packet resolvent convolution). *Let z belong to the resolvent set of A. If* Re z *is larger than a growth bound for the semigroup, then*

$$P_\beta (zI - A)^{-1} P_\alpha = \int_0^\infty \mathrm{e}^{-zt} T_{\beta\alpha}(t)\, \mathrm{d}t.$$

Proof. The standard Laplace representation of the resolvent is

$$(zI - A)^{-1} = \int_0^\infty \mathrm{e}^{-zt} T(t)\, \mathrm{d}t.$$

Multiplication by the bounded packet projections may be passed through the Bochner integral. □

For the path-order statement, now assume A is bounded as in Part II. A directed edge from α to β means that $P_\beta A P_\alpha$ is nonzero, with $\alpha \ne \beta$; a packet always has the length-zero path to itself.

Corollary R1.3 (Minimum transfer order)**.** *Suppose every directed packet path from α to β uses at least m_0 off diagonal transfers. Then*

$$P_\beta \mathrm{e}^{tA} P_\alpha = O(t^{m_0}) \quad \text{as } t \downarrow 0.$$

If there is no directed path from α to β, the block vanishes for all t.

Proof. Every nonzero term in theorem E2.8 contains at least m_0 time integrations. The simplex $0 < r_1 < \cdots < r_m < t$ has volume $t^m/m!$, so the first possible order is t^{m_0}. If there is no path, every packet term vanishes. □

Let $C_j : X_{j+1} \to X_j$ be bounded linear maps, and put $C_{\ell,m} = C_\ell \cdots C_{m-1}$ for $\ell < m$, with $C_{j,j} = I$. No common ambient space is assumed.

Theorem R1.4 (Several level telescoping identity)**.** *Let T_j be bounded operators on X_j, and set*

$$E_j = C_j T_{j+1} - T_j C_j.$$

Then

$$\boxed{C_{\ell,m} T_m - T_\ell C_{\ell,m} = \sum_{j=\ell}^{m-1} C_{\ell,j} E_j C_{j+1,m}.}$$

Proof. Expand each summand as

$$C_{\ell,j+1} T_{j+1} C_{j+1,m} - C_{\ell,j} T_j C_{j,m}.$$

The intermediate terms cancel. The first surviving term is $C_{\ell,m} T_m$, and the last is $-T_\ell C_{\ell,m}$. □

Corollary R1.5 (Summable generator defects)**.** *Suppose C_j are contractions and $(T_j(t))$ are contraction semigroups with bounded generators A_j. Set*

$$\delta_j = \left\| C_j A_{j+1} - A_j C_j \right\|.$$

Then

$$\left\| C_{\ell,m} T_m(t) - T_\ell(t) C_{\ell,m} \right\| \le t \sum_{j=\ell}^{m-1} \delta_j.$$

If $\sum_{j \ge 0} \delta_j < \infty$, the right side tends to zero as $\ell \to \infty$, uniformly in $m > \ell$ and uniformly for t in a fixed bounded interval.

Proof. By theorem E3.1, the interlevel semigroup defect has norm at most $t\delta_j$. Apply theorem R1.4 and use the contraction bounds on all coarsening maps. □

Definition R1.6 (Projective limit space)**.** Let (X_j, C_j) be a sequence of Banach spaces and bounded linear maps $C_j : X_{j+1} \to X_j$. Its projective limit is

$$X_\infty = \left\{ (x_j)_{j \ge 0} \in \prod_{j \ge 0} X_j : C_j x_{j+1} = x_j \text{ for every } j \right\}.$$

We equip X_∞ with the subspace topology inherited from the product topology.

Theorem R1.7 (Exact projective limit semigroup). *Suppose each X_j carries a strongly continuous semigroup $(T_j(t))_{t\geq 0}$, and*

$$C_j T_{j+1}(t) = T_j(t) C_j$$

for every j and every $t \geq 0$. Then

$$T_\infty(t)(x_j)_{j\geq 0} = (T_j(t)x_j)_{j\geq 0}$$

defines a strongly continuous semigroup on X_∞.

Proof. Exact intertwining gives

$$C_j T_{j+1}(t) x_{j+1} = T_j(t) C_j x_{j+1} = T_j(t) x_j.$$

Thus the coordinatewise evolution remains in the projective limit. The semigroup law holds in every coordinate. Strong continuity in the product topology follows because every coordinate map

$$t \longmapsto T_j(t)x_j$$

is continuous. □

Corollary R1.8 (Coordinate generators). *Assume the setting of theorem R1.7. If $x = (x_j)$ satisfies $x_j \in \mathrm{Dom}(A_j)$ and the sequence $(A_j x_j)$ belongs to X_∞, then x lies in the generator domain of T_∞ and*

$$A_\infty x = (A_j x_j)_{j\geq 0}.$$

Proof. The difference quotients converge in every coordinate to $A_j x_j$. The assumed compatibility places the coordinatewise limit in X_∞, which is exactly convergence in the product topology. □

The product topology used here differs from the bounded inverse-limit norm of Part II. No uniform boundedness of coordinate generators is required by the coordinatewise statement.

R2 Reversible coarse dynamics

Let Q be a finite Markov generator acting on column probability vectors, let (B_a) be a partition into nonempty blocks, and let $C_{ai} = 1_{i\in B_a}$. Strong lumpability means $CQ = \overline{Q}C$ for a coarse Markov generator. For observables, the backward generator is Q^T; weighted L^2 inner products use the stated stationary probability.

Theorem R2.1 (Reversibility descends). *Assume Q is strongly lumpable and reversible with respect to a positive stationary distribution π, so*

$$Q_{ij}\pi_j = Q_{ji}\pi_i.$$

Define

$$\overline{\pi}_a = \sum_{i\in B_a} \pi_i.$$

Then $\overline{\pi}$ is stationary for $\overline{Q}$, and

$$\overline{Q}_{ab}\overline{\pi}_b = \overline{Q}_{ba}\overline{\pi}_a.$$

Thus the coarse chain is reversible.

Proof. Stationarity follows from

$$\overline{Q}\overline{\pi} = \overline{Q}C\pi = CQ\pi = 0.$$

For reversibility, use the aggregate rate formula:

$$\begin{aligned}\overline{Q}_{ab}\overline{\pi}_b &= \sum_{j\in B_b} \pi_j \sum_{i\in B_a} Q_{ij} \\ &= \sum_{i\in B_a}\sum_{j\in B_b} \pi_i Q_{ji} \\ &= \overline{Q}_{ba}\overline{\pi}_a.\end{aligned}$$

□

Corollary R2.2 (Coarse Dirichlet form)**.** *Under the hypotheses of theorem R2.1, let Jg be the fine function that is constant with value g_a on B_a. Then*

$$-\langle Jg, Q^T Jg\rangle_{L^2(\pi)} = -\left\langle g, \overline{Q}^T g\right\rangle_{L^2(\overline{\pi})}.$$

Proof. Expand the left side and group the fine indices by their coarse blocks. The aggregate rate criterion converts the resulting sums into the coarse generator entries. □

R3 Prior-restricted sufficiency and filtering

Let L be a strictly positive finite likelihood and let $\mathcal{B}_L(p) = L \odot p / \langle L, p\rangle$ for a probability vector p. Fix probability vectors $r^{(a)}$ supported on B_a, and restrict the priors to $p_i = q_a r_i^{(a)}$ for $i \in B_a$, with q a coarse probability vector.

Proposition R3.1 (Sufficiency relative to a prior disintegration)**.** *Define*

$$\overline{L}_a = \sum_{i\in B_a} L_i r_i^{(a)}.$$

For every prior in the declared family,

$$C\mathcal{B}_L(p) = \mathcal{B}_{\overline{L}}(q).$$

Proof. The evidence in block B_a is

$$Z_a = q_a \sum_{i\in B_a} L_i r_i^{(a)} = q_a \overline{L}_a.$$

Apply theorem E8.2. □

Lemma R3.2 (Normalization bound)**.** *Let $a, b \in [0,\infty)^m$ be nonzero, and set $A = \|a\|_1$ and $B = \|b\|_1$. Then*

$$\boxed{\|a/A - b/B\|_1 \le \frac{2\|a-b\|_1}{\max\{A,B\}}.}$$

Proof. Using A in the denominator gives

$$\|a/A - b/B\|_1 \le \frac{\|a-b\|_1}{A} + \|b\|_1 \, |1/A - 1/B| \, .$$

The second term equals $|B-A|/A$, which is at most $\|a-b\|_1/A$. Interchanging a and b gives the analogous bound with B. Taking the smaller of the two bounds gives the result. □

For an arbitrary prior p, put $q = Cp$ and $Z(p, L) = \langle L, p\rangle$. Define $\ell_a(p) = \sum_{i \in B_a} L_i p_i / q_a$ when $q_a > 0$; choose any positive value when $q_a = 0$. Only the product $q_a \ell_a(p)$ enters the formulas.

Theorem R3.3 (Exact information loss formula)**.** *For every prior,*

$$\boxed{C\mathcal{B}_L(p) = \mathcal{B}_{\ell(p)}(q).}$$

Let $\overline{L}$ be any fixed positive coarse likelihood and set

$$d(p \mid L, \overline{L}) = \sum_{a=1}^{m} q_a \left| \ell_a(p) - \overline{L}_a \right| .$$

Then

$$\|C\mathcal{B}_L(p) - \mathcal{B}_{\overline{L}}(q)\|_1 \le \frac{2d(p \mid L, \overline{L})}{\max\{Z(p, L), \langle \overline{L}, q\rangle\}}.$$

Proof. The packet evidence in block a is $q_a \ell_a(p)$, which proves the exact identity. Apply lemma R3.2 to

$$a = q \odot \ell(p), \qquad b = q \odot \overline{L}.$$

Their ℓ^1 norms are the two evidences, and $\|a-b\|_1 = d(p \mid L, \overline{L})$. □

Corollary R3.4 (Oscillation bound)**.** *Suppose $\overline{L}_a$ is any representative likelihood for block B_a, and set*

$$\omega_a = \max_{i \in B_a} \left| L_i - \overline{L}_a \right| .$$

Then

$$d(p \mid L, \overline{L}) \le \sum_a q_a \omega_a.$$

Thus small within block likelihood oscillation gives a small coarse posterior defect whenever the evidence is bounded away from zero.

Proof. The conditional average $\ell_a(p)$ lies in the convex hull of the likelihood values inside B_a. Hence

$$\left| \ell_a(p) - \overline{L}_a \right| \le \omega_a.$$

Multiply by q_a and sum. □

Along the Bayesian flow $p_i(t) = e^{t g_i} p_i(0) / \sum_j e^{t g_j} p_j(0)$, the following statement holds for every real observable h.

Corollary R3.5 (Covariance identity). *For any observable $h \in \mathbb{R}^n$,*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_{p(t)}[h] = \mathrm{Cov}_{p(t)}(h, g).$$

In particular,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_{p(t)}[g] = \mathrm{Var}_{p(t)}(g) \geq 0.$$

Proof. Differentiate $\sum_i h_i p_i(t)$ and insert the replicator equation. □

For stochastic matrices $K, \overline{K}$ and positive likelihoods $L_y, \overline{L}_y$, put $\Phi_y = \mathcal{B}_{L_y} \circ K$ and $\overline{\Phi}_y = \mathcal{B}_{\overline{L}_y} \circ \overline{K}$.

Theorem R3.6 (Exact coarse hidden Markov filter). *Assume*

$$CK = \overline{K}C$$

and every likelihood L_y is constant on each coarse block with block values $\overline{L}_y$. Then

$$C\Phi_y(p) = \overline{\Phi}_y(Cp)$$

for every prior. For any observation sequence $y_1, \dots, y_r$, the complete filtering recursion commutes with aggregation:

$$C\Phi_{y_r} \circ \cdots \circ \Phi_{y_1} = \overline{\Phi}_{y_r} \circ \cdots \circ \overline{\Phi}_{y_1} C.$$

Proof. Exact Markov lumping gives $CKp = \overline{K}Cp$. The Bayesian sufficiency criterion gives exact commutation of the update. Their composition proves the one step identity. Iterate it over the observation sequence. □

Theorem R3.7 (One step filter defect). *Let $r = Kp$, $q = Cr$, and let $\ell_y(r)$ be the exact coarse likelihood obtained by averaging L_y inside each block with respect to r. Assume*

$$0 < \overline{m} \leq (\overline{L}_y)_a \leq \overline{M}.$$

Then

$$\left\|C\Phi_y(p) - \overline{\Phi}_y(Cp)\right\|_1 \leq \frac{2d(r \mid L_y, \overline{L}_y)}{\max\{\langle L_y, r\rangle, \langle \overline{L}_y, q\rangle\}} + \frac{2\overline{M}}{\overline{m}} \left\|CKp - \overline{K}Cp\right\|_1.$$

Proof. Insert the intermediate posterior $\mathcal{B}_{\overline{L}_y}(CKp)$. The first difference is bounded by theorem R3.3. The second is bounded by theorem E8.4 applied to the coarse likelihood. □

R4 Fractional modes and killed transfer

For the one-dimensional Dirichlet difference operator, let $h = 1/M$ with an integer $M \geq 2$, let $1 \leq k \leq M-1$, and set $\lambda_{k,h} = 4h^{-2}\sin^2(k\pi h/2)$. The retained explicit coefficient below is valid for every discrete mode.

Corollary R4.1 (Fractional low mode dispersion). *Let $0 < \alpha \leq 1$. Then*

$$0 \leq (k\pi)^{2\alpha} - \lambda_{k,h}^{\alpha} \leq \frac{\alpha(4k^2)^{\alpha-1}(k\pi)^4}{12}h^2.$$

For every $t \geq 0$,

$$\left|\mathrm{e}^{-t\lambda_{k,h}^{\alpha}} - \mathrm{e}^{-t(k\pi)^{2\alpha}}\right| \leq \frac{t\alpha(4k^2)^{\alpha-1}(k\pi)^4}{12}h^2.$$

Proof. The chord bound $\sin x \geq 2x/\pi$ on $[0, \pi/2]$ gives $\lambda_{k,h} \geq 4k^2$. Since $x \mapsto x^{\alpha}$ is increasing and has decreasing derivative,

$$(k\pi)^{2\alpha} - \lambda_{k,h}^{\alpha} \leq \alpha\lambda_{k,h}^{\alpha-1}((k\pi)^2 - \lambda_{k,h}).$$

Use theorem E12.5 and the lower bound for $\lambda_{k,h}$. The exponential estimate follows from

$$\left|\mathrm{e}^{-ta} - \mathrm{e}^{-tb}\right| \leq t\,|a-b|$$

for nonnegative a and b. □

Let Q and $\overline{Q}$ be finite Markov generators in the column convention. All times in this section satisfy $t \geq 0$.

Theorem R4.2 (Potential perturbation bound). *Let V and W be nonnegative diagonal matrices. In the supremum norm,*

$$\boxed{\left\|\mathrm{e}^{t(Q^T-V)} - \mathrm{e}^{t(Q^T-W)}\right\|_{\infty\to\infty} \leq t\,\|V-W\|_{\infty\to\infty}.}$$

Proof. Both killed Markov semigroups are contractions in the supremum norm. Apply the Duhamel identity with generator defect $W - V$. □

Theorem R4.3 (Coarse Feynman Kac transfer). *Assume $CQ = \overline{Q}C$, and let $J = C^T$ lift a coarse function to the corresponding block constant fine function. Suppose the potential is block constant, so*

$$VJ = J\overline{V}.$$

Then

$$\boxed{\mathrm{e}^{t(Q^T-V)}J = J\mathrm{e}^{t(\overline{Q}^T-\overline{V})}}$$

for every $t \geq 0$.

Proof. Transpose the Markov intertwining relation to obtain

$$Q^T J = J\overline{Q}^T.$$

Together with $VJ = J\overline{V}$, this gives

$$(Q^T - V)J = J(\overline{Q}^T - \overline{V}).$$

The exponential intertwining follows from the power series. □

R5 Memory compression, rational realization, and phase-type consequences

Fix $T > 0$. For the next estimate let A be a bounded operator on a Banach space X, let $\mathcal{K}, \widetilde{\mathcal{K}} \in L^1((0,T), \mathcal{B}(X))$, and let f be continuous. Take continuously differentiable solutions of the displayed equations. This direct bounded-generator estimate retains a sharper exponential constant than the general mild-semigroup estimate in Part II.

Theorem R5.1 (Stability of memory compression). *Let x and $\tilde{x}$ solve*

$$\dot{x}(t) = Ax(t) + \int_0^t \mathcal{K}(t-s)x(s)\,\mathrm{d}s + f(t)$$

and

$$\dot{\tilde{x}}(t) = A\tilde{x}(t) + \int_0^t \widetilde{\mathcal{K}}(t-s)\tilde{x}(s)\,\mathrm{d}s + f(t)$$

with the same initial value. Fix $T > 0$, and assume

$$\sup_{0\le t\le T} \|\tilde{x}(t)\| \le M.$$

Set

$$a = \|A\|, \qquad k = \|\mathcal{K}\|_{L^1(0,T)}, \qquad \delta = \left\|\mathcal{K} - \widetilde{\mathcal{K}}\right\|_{L^1(0,T)}.$$

Then

$$\boxed{\sup_{0\le t\le T} \|x(t) - \tilde{x}(t)\| \le MT\delta \mathrm{e}^{(a+k)T}.}$$

Proof. Let $e = x - \tilde{x}$. Integrating the difference equation and changing the order of integration gives

$$\|e(t)\| \le (a+k)\int_0^t \|e(r)\|\,\mathrm{d}r + Mt\delta.$$

Since $t \le T$, Gronwall's inequality gives

$$\|e(t)\| \le MT\delta \mathrm{e}^{(a+k)t}.$$

Take the supremum over $[0,T]$. □

For the exact closure criterion of Part II, allowing an arbitrary bounded local generator B does not enlarge the class: differentiating $Pe^{tA}P = e^{tB}P$ at $t = 0$ forces $B = A_{PP}$ on the retained space. Thus its zero-memory criterion excludes every alternative bounded local generator.

A finite auxiliary realization of a matrix kernel means $\mathcal{K}(t) = Be^{tD}C$, with finite matrices of compatible sizes. Laplace transforms are taken in a right half-plane of absolute convergence. The rational charac-

terization concerns continuous kernels of exponential order, so uniqueness of the ordinary Laplace transform applies.

Theorem R5.2 (Rational memory realization)**.** *A matrix valued memory transform has a finite auxiliary realization if and only if it is a strictly proper rational matrix function. More precisely,*

$$\boxed{\widehat{\mathcal{K}}(z) = B(zI - D)^{-1}C}$$

if and only if

$$\boxed{\mathcal{K}(t) = B\mathrm{e}^{tD}C.}$$

Proof. The forward implication follows by taking the Laplace transform. The resulting matrix is rational and tends to zero as z tends to infinity.

Conversely, decompose every entry of a strictly proper rational matrix into partial fractions. A term

$$\frac{c}{(z-\lambda)^r}$$

is an entry of the resolvent of an r by r Jordan block with eigenvalue λ. Taking a direct sum of these blocks and choosing the input and output maps entry by entry gives matrices B, D, C with the required resolvent. Inverse Laplace transformation gives the time domain formula. □

Corollary R5.3 (Exponential sum memory)**.** *If*

$$\mathcal{K}(t) = \sum_{j=1}^{m} B_j \mathrm{e}^{\lambda_j t} C_j,$$

then $\mathcal{K}$ has a finite auxiliary realization with one auxiliary block for each exponential. Repeated poles produce polynomial factors in t and are realized by Jordan blocks.

Proof. If B_j and C_j have an intermediate space of dimension d_j, use a block diagonal matrix D with blocks $\lambda_j I_{d_j}$, or the required Jordan blocks when poles repeat. Assemble the matrices B_j and C_j into the output and input maps. □

For the survival consequences, A is a finite Metzler matrix with nonpositive column sums and spectrum in the open left half-plane, p_0 is a probability vector, $S(t) = 1^T e^{tA} p_0$, and $f(t) = -S'(t)$. The one-phase model has $\lambda > 0$. The survival Laplace identity of Part II also holds for every complex s with $\operatorname{Re} s \geq 0$: stability gives absolute convergence of $\int_0^\infty e^{-st} e^{tA}\, dt = (sI - A)^{-1}$, and multiplying by the exit row and p_0 gives the same transform.

Corollary R5.4 (Phase type rationality)**.** *The Laplace transform of every finite phase type distribution is rational. If A is diagonalizable, then $S(t)$ and $f(t)$ are finite linear combinations of exponential functions. Jordan blocks add polynomial factors in t.*

Proof. The resolvent $(sI - A)^{-1}$ is rational. The time domain statement follows from the Jordan form of A. □

Corollary R5.5 (One phase model)**.** *If $A = (-\lambda)$ and $p_0 = 1$, then*

$$S(t) = \mathrm{e}^{-\lambda t}, \qquad f(t) = \lambda \mathrm{e}^{-\lambda t}, \qquad h(t) = \lambda.$$

Proof. Insert the scalar generator into theorem E16.2. □

R6 Earlier exact examples

These five short examples are retained from the earlier human edition. Their numerical parameters are independent of the later visual guide's worked models.

Example R6.1 (One transfer between two packets)**.** Let

$$A = \begin{pmatrix} a & 0 \\ c & b \end{pmatrix}$$

and let the two coordinate axes be the packet spaces. If $a \neq b$, then

$$\mathrm{e}^{tA} = \begin{pmatrix} \mathrm{e}^{at} & 0 \\ c\dfrac{\mathrm{e}^{at} - \mathrm{e}^{bt}}{a-b} & \mathrm{e}^{bt} \end{pmatrix}.$$

The lower left block is produced by one transfer from the first packet to the second. Its expansion begins with ct, in agreement with corollary R1.3. When $a = b$, the same block is cte^{at}.

Example R6.2 (A four state lumpable chain)**.** Partition four states into $B_1 = \{1, 2\}$ and $B_2 = \{3, 4\}$. For positive rates a, b, u, v, set

$$Q = \begin{pmatrix} -u-a & u & b/2 & b/2 \\ u & -u-a & b/2 & b/2 \\ a/2 & a/2 & -v-b & v \\ a/2 & a/2 & v & -v-b \end{pmatrix}.$$

The rates u and v move inside the coarse blocks. Every state in B_1 leaves for B_2 at total rate a, and every state in B_2 leaves for B_1 at total rate b. Hence

$$\overline{Q} = \begin{pmatrix} -a & b \\ a & -b \end{pmatrix}$$

and $CQ = \overline{Q}C$. The internal rates disappear from the coarse generator even though they remain present in the fine evolution.

Example R6.3 (A visible likelihood loss)**.** Let $B_1 = \{1, 2\}$, let $B_2 = \{3\}$, and take

$$p = (3/10, 2/10, 5/10), \qquad L = (1, 3, 2).$$

The coarse prior is $(1/2, 1/2)$. The exact block likelihoods are

$$\ell_1(p) = \frac{(3/10) + 3(2/10)}{1/2} = \frac{9}{5}, \qquad \ell_2(p) = 2.$$

Thus the exact coarse posterior is $(9/19, 10/19)$. If the first block is replaced by the fixed representative value 2, the approximate coarse posterior is $(1/2, 1/2)$. The ℓ^1 defect is $1/19$, while theorem R3.3 gives the valid upper bound $1/10$.

Example R6.4 (One hidden mode)**.** Consider

$$A = \begin{pmatrix} -a & c \\ d & -b \end{pmatrix}, \qquad a, b > 0,$$

and retain the first coordinate. The memory kernel and self energy are

$$\mathcal{K}(t) = cd\mathrm{e}^{-bt}, \qquad \Sigma(z) = \frac{cd}{z+b}.$$

The retained coordinate has an exact time local closure for every retained initial value precisely when $cd = 0$. When $cd \neq 0$, one auxiliary scalar state realizes the memory exactly.

Example R6.5 (Two sequential phases)**.** Let

$$A = \begin{pmatrix} -\lambda & 0 \\ \lambda & -\mu \end{pmatrix}, \qquad p_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

where $\lambda \neq \mu$ are positive. The first phase moves to the second at rate λ, and the second is absorbed at rate μ. Direct calculation gives

$$S(t) = \frac{\mu\mathrm{e}^{-\lambda t} - \lambda\mathrm{e}^{-\mu t}}{\mu - \lambda},$$

$$f(t) = \frac{\lambda\mu}{\mu - \lambda}\left(\mathrm{e}^{-\lambda t} - \mathrm{e}^{-\mu t}\right).$$

The ratio $f(t)/S(t)$ is generally not constant. A time local Markov model on two transient phases can therefore produce a nonconstant hazard after the phase coordinate is hidden.

R7 Holomorphic operator substitution

> **A route through this section**
>
> The scalar function, zero-free domain, and logarithm branch are fixed before substituting an operator. The bounded Banach-space functional calculus here is separate from the finite canonical-jet contract of Part I. Keep those spaces and hypotheses visible when using the reciprocal family.

This section receives the operator-substitution result deliberately moved out of Paper IV. Fix an integer $N \geq 1$ and the entire scalar function

$$\mathsf{D}_N(z) = \sum_{m=0}^{\infty} \frac{(-1)^m z^{Nm}}{(Nm)!}.$$

More generally, the result applies to any specified holomorphic scalar function that is nonzero on the declared domain. On a simply connected zero-free domain U, choose $g = \log \mathsf{D}_N$ once and define $\mathsf{S}_N^{(r)}(z) = \exp(-rg(z))$ for complex r. The bounded Banach-space result below uses classical holomorphic functional calculus; it is separate from the finite canonical-jet calculus formalized in Part I.

Standard Fact R7.1 (Holomorphic functional calculus). Let A be a bounded operator on a complex Banach space. If f is holomorphic on a neighborhood of $\mathrm{Spec}(A)$, then

$$f(A) = \frac{1}{2\pi i}\int_{\Gamma} f(z)(zI - A)^{-1}\, dz$$

for an admissible positively oriented contour Γ lying in the holomorphy domain and the resolvent set, with winding number one about the spectrum and zero outside the domain. The calculus preserves sums, products, and composition. We use the standard form from [DS58]; an accessible proof is given in [Tay, Appendix A].

Proposition R7.2 (Operator reciprocal family)**.** *Let A be bounded and suppose $\mathrm{Spec}(A)$ lies in a simply connected region on which D_N has no zeros. Fix a holomorphic branch $g = \log \mathsf{D}_N$ there. Then*

$$\mathsf{S}_N^{(r)}(A) = \exp(-r g(A)) = (\mathsf{D}_N^{-r})(A)$$

is well defined and

$$\mathsf{S}_N^{(r)}(A)\mathsf{S}_N^{(s)}(A) = \mathsf{S}_N^{(r+s)}(A).$$

If (P_i) is a finite complete family of bounded idempotents, then

$$P_i f(A) P_j = \frac{1}{2\pi i} \int_\Gamma f(z) P_i (zI - A)^{-1} P_j \, dz$$

for every holomorphic f on a neighborhood of the spectrum.

Proof. The spectral mapping theorem makes $\mathsf{D}_N(A)$ invertible. The product identity follows from multiplicativity of the functional calculus and the fixed logarithm. Applying P_i and P_j to the Cauchy integral formula gives the packet formula. □

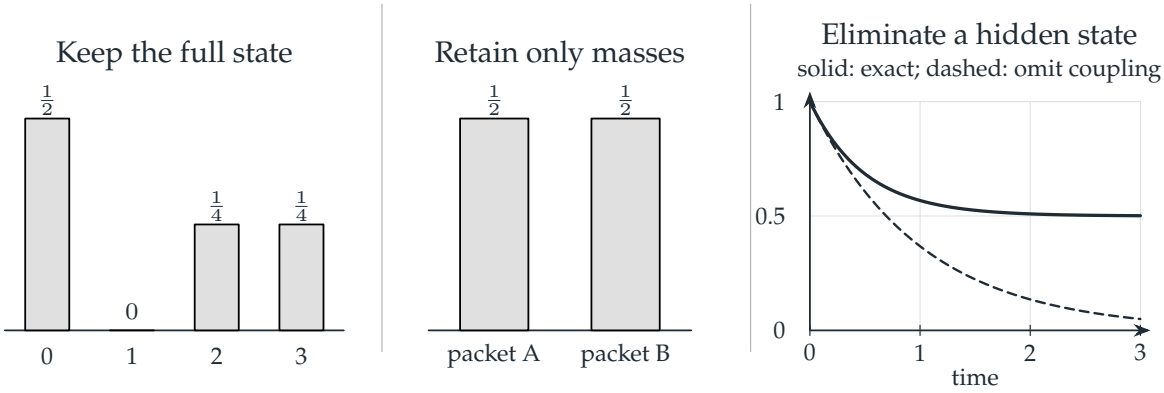
Part IV Worked visual atlas

V1 / PAPER VI / A WORKED VISUAL ATLAS

Three ways to reduce a model

Traditional 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 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_π	Conditional expectation of observables.	A weighted average, self-adjoint for the π 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

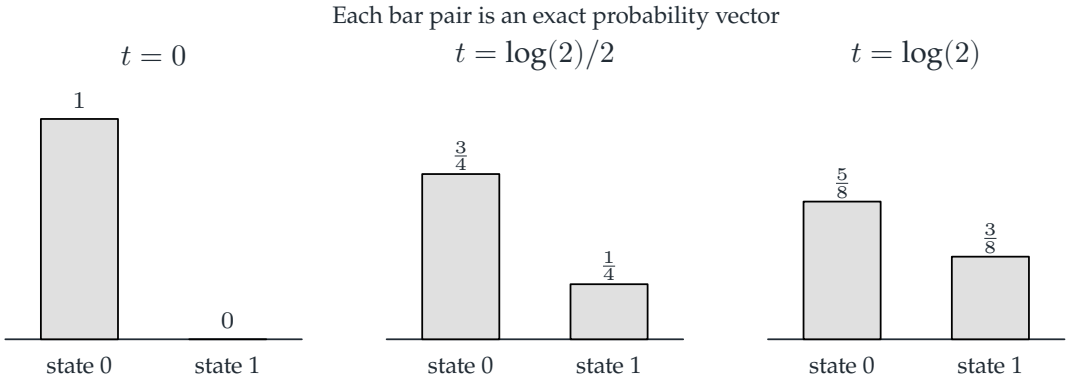
Traditional Part I: Theorems 1.2 and 4.1. Finite matrix exponentials and Markov positivity.

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

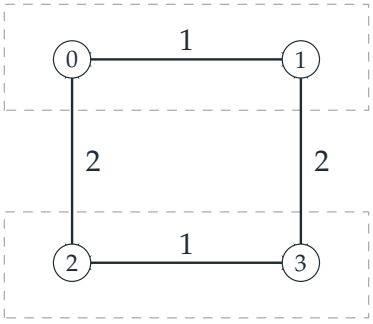
Traditional Part I: Theorem 2.1, Corollary 2.2, and the packet path expansion.

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 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.

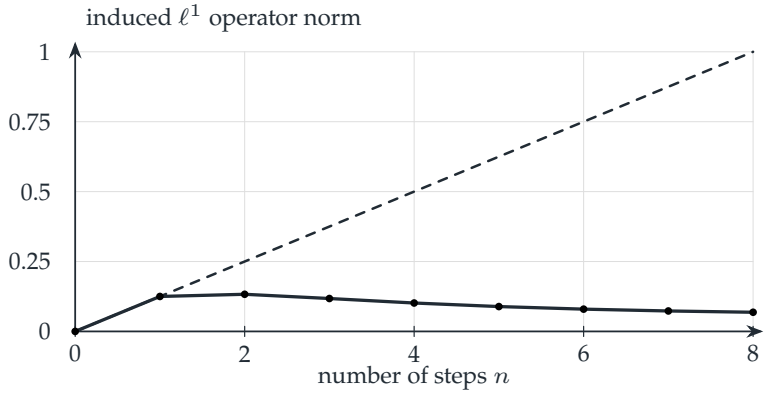
V6 / DEFECTS / A GUARANTEE WITHOUT A FIT

A broken lump and its error budget

Traditional 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_ε 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 ℓ^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

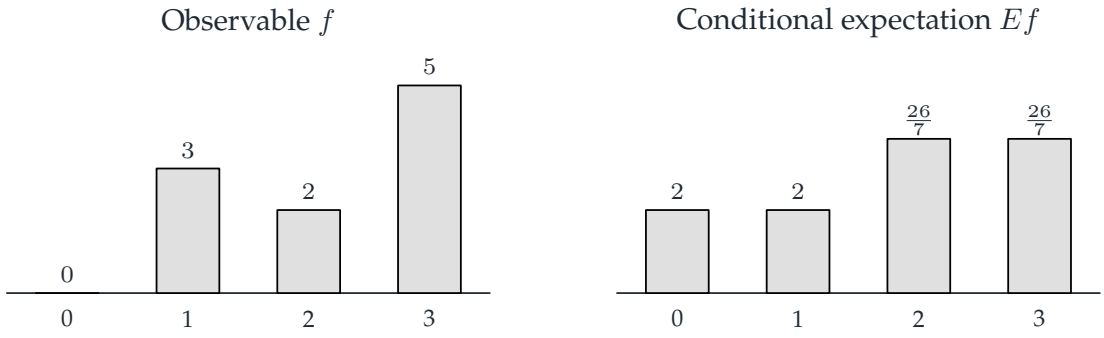
Traditional Part I: Theorem 6.1. Strictly positive weights are part of the setting.

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

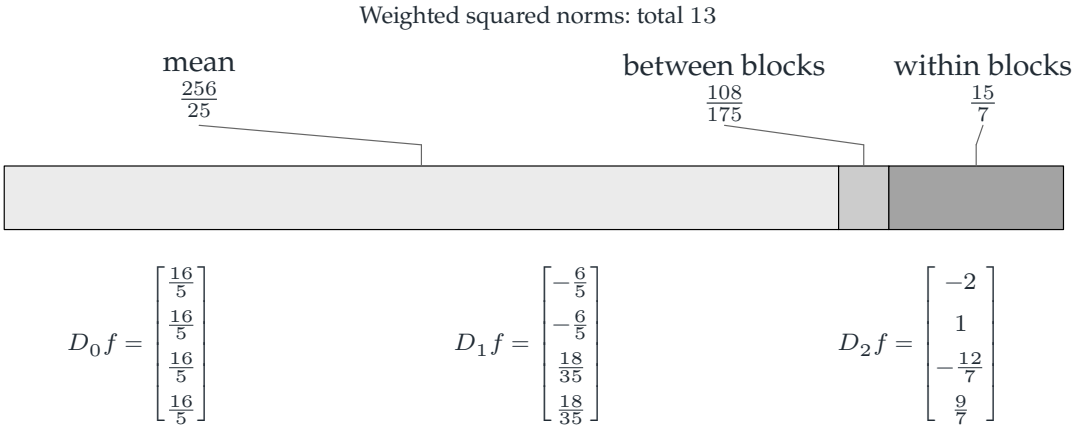
Where the variance goes

Traditional 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

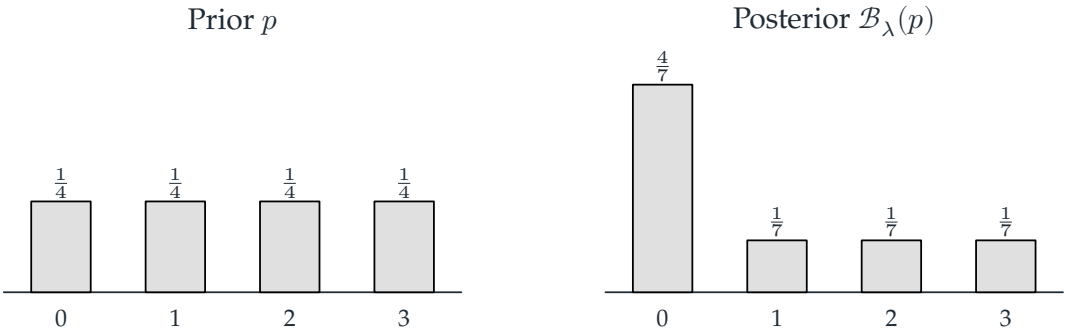
Traditional Part I: Theorems 7.1–7.3. Normalization, packet reconstruction, and positive evidence.

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

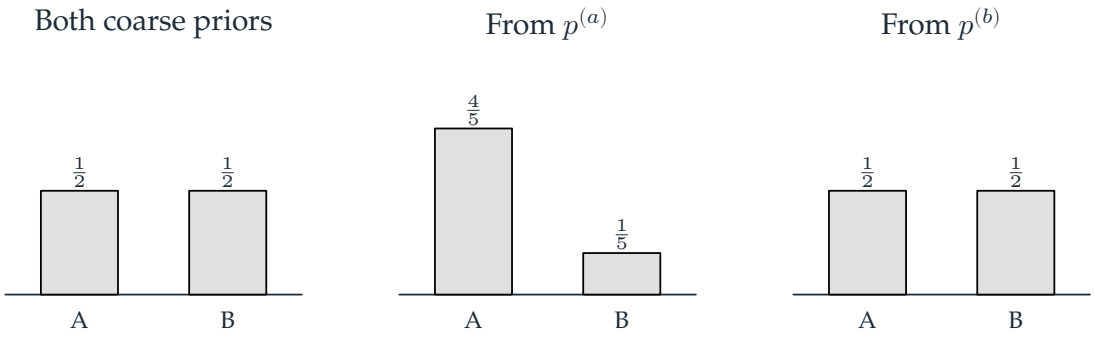
Traditional Part I: Theorem 8.2 and the one-packet exception.

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

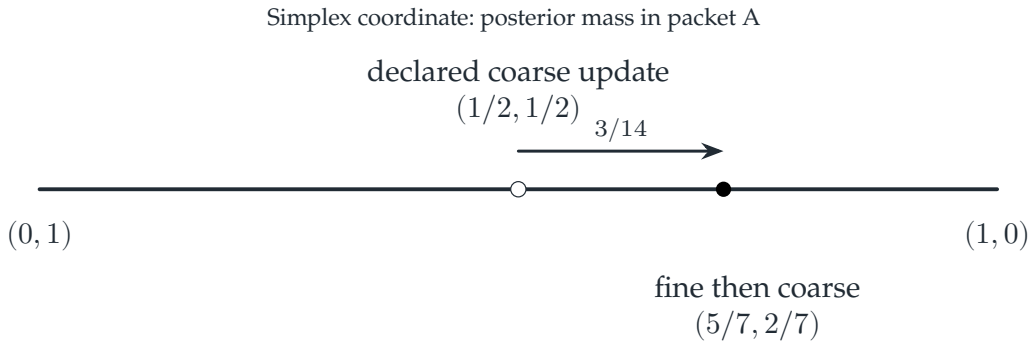
The exact information-loss vector

Traditional 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 ℓ^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

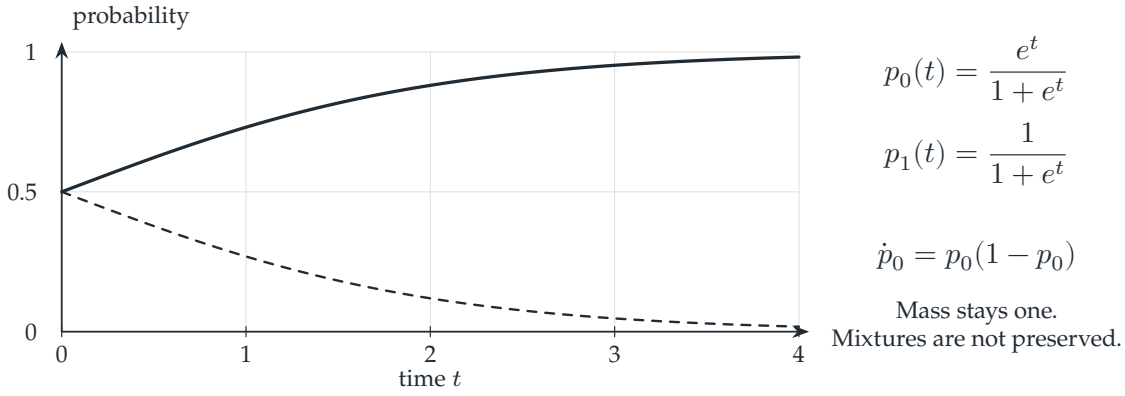
A nonlinear semigroup of odds

Traditional 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

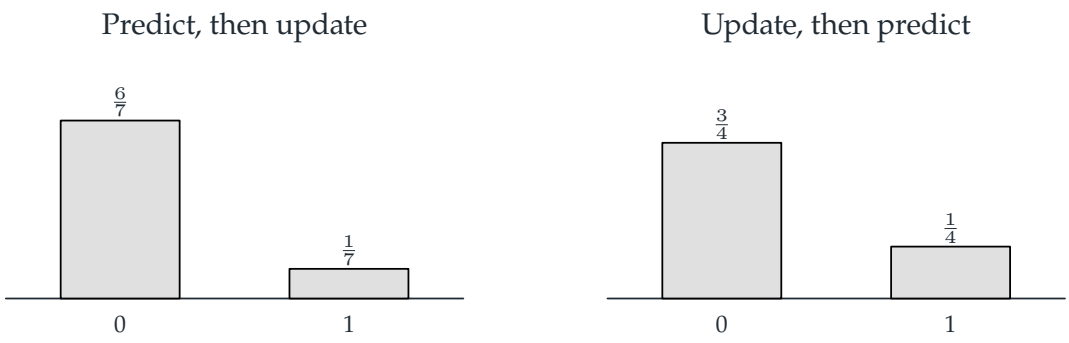
Prediction and observation do not commute

Traditional 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

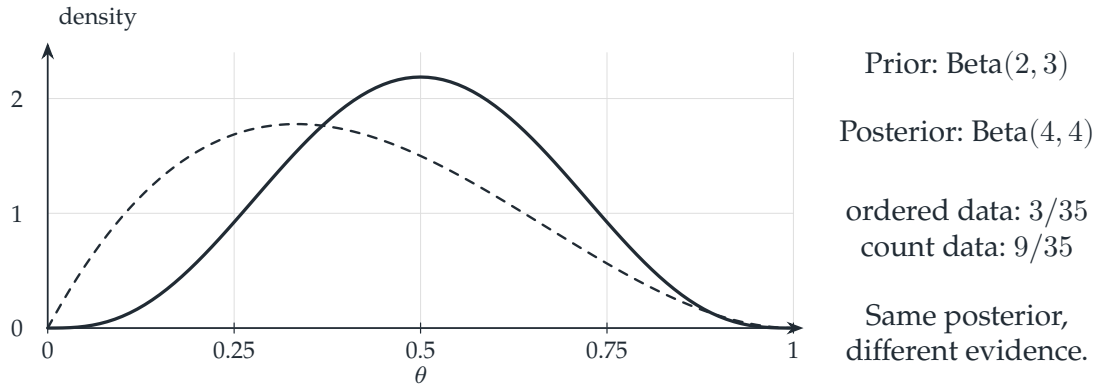
Conjugacy tracks exponents

Traditional 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 $140\theta^3(1-\theta)^3$.



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 θ, 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 θ 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

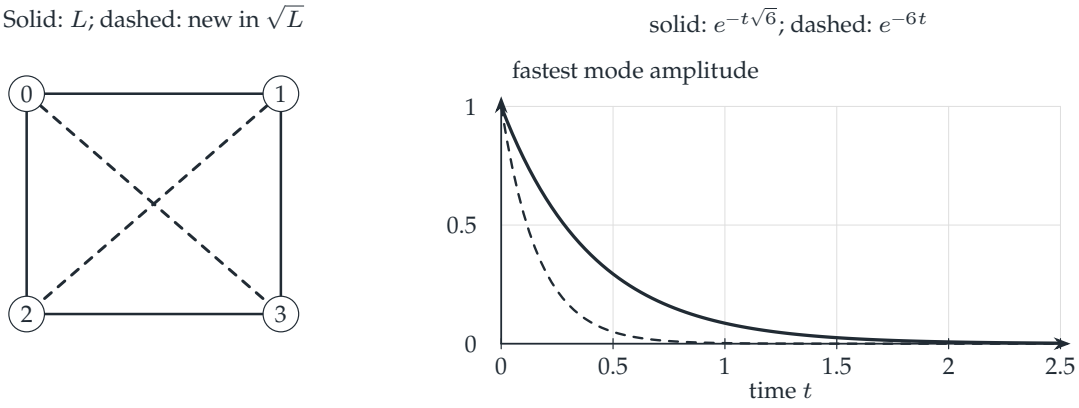
The modes stay; their clocks change

Traditional 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.



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

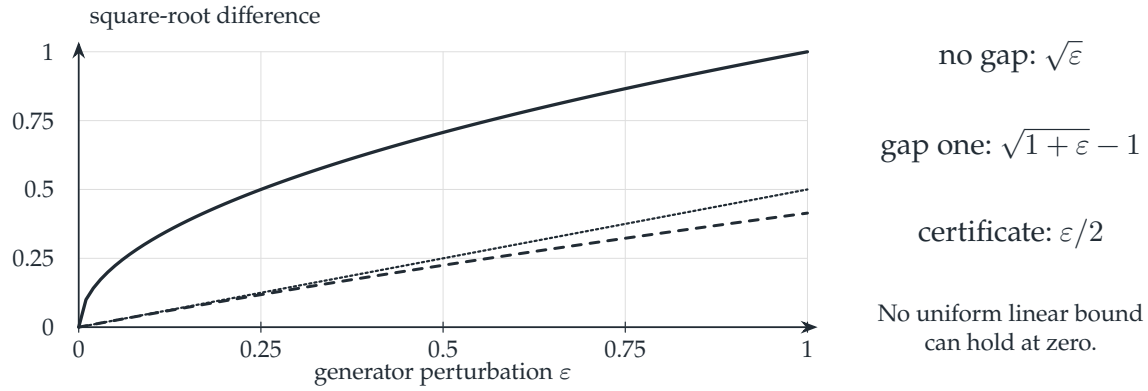
A gap changes the error scale

Traditional 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 ε. 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

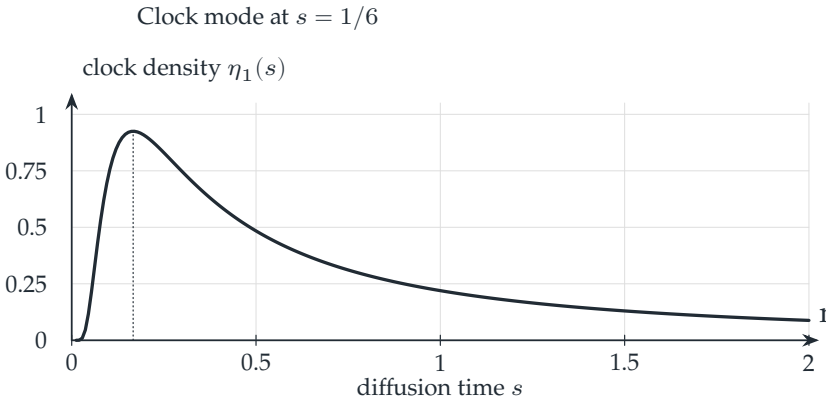
A random diffusion clock

Traditional 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 η_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 η_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

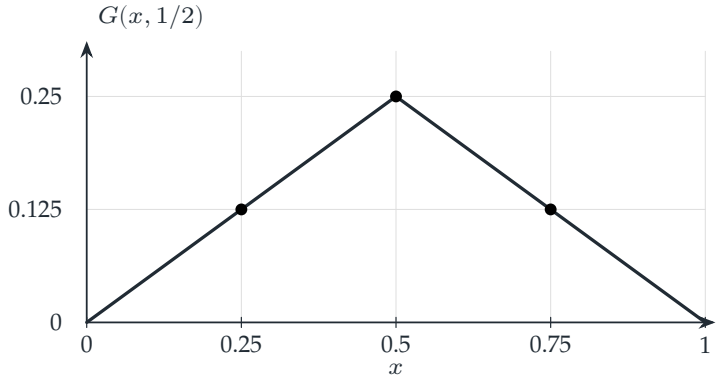
Traditional Part I: Theorem 11.1. Exact nodal Green reconstruction for this one-dimensional stencil.

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

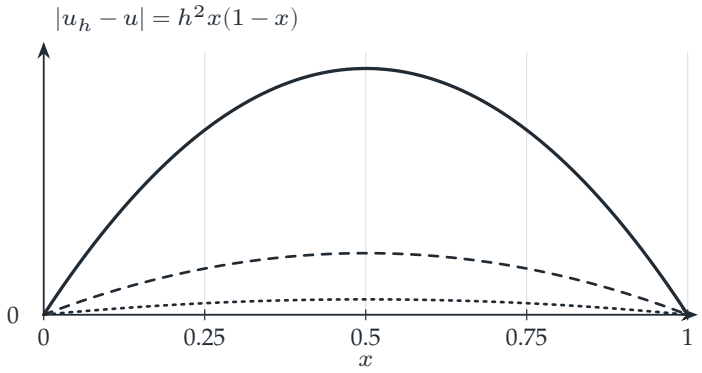
Use $u(x) = x - x^4$, which vanishes at both endpoints and solves $-u'' = 12x^2$. Its centered second difference is not exact:

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

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

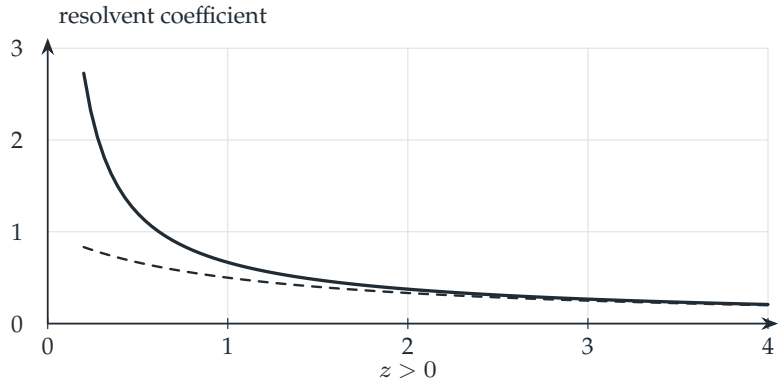
The two-state evolution from V3 has generator

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

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

Write the same system as

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

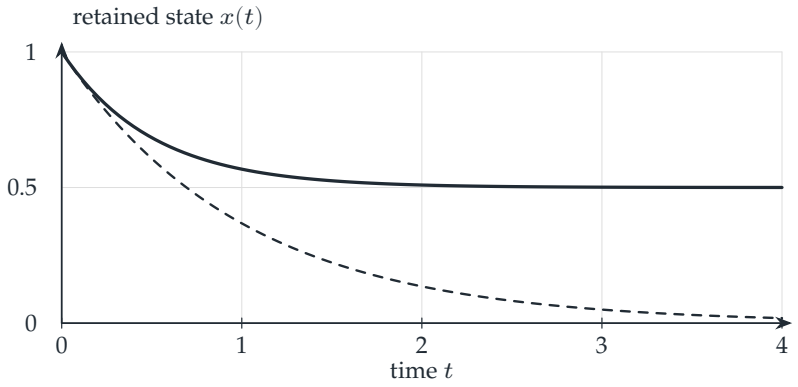
Solving the second equation first gives

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

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.



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 Part I: Theorem 13.2 and Section 15. Functional calculus must retain nilpotent data.

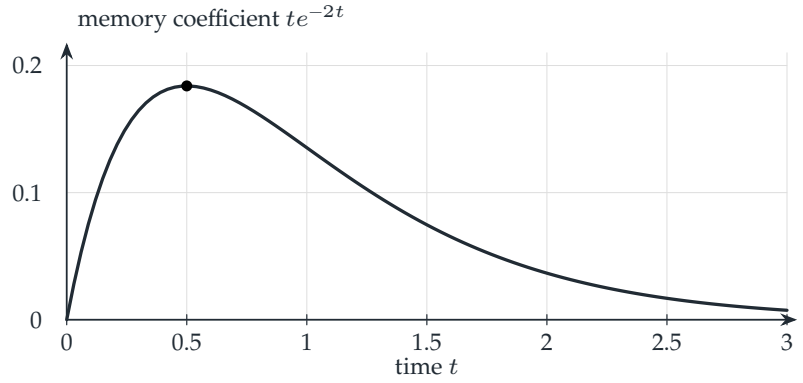
Now use a two-dimensional hidden block

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

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 Part I: Theorem 14.1. Survival, density, hazard, and moments have distinct meanings.

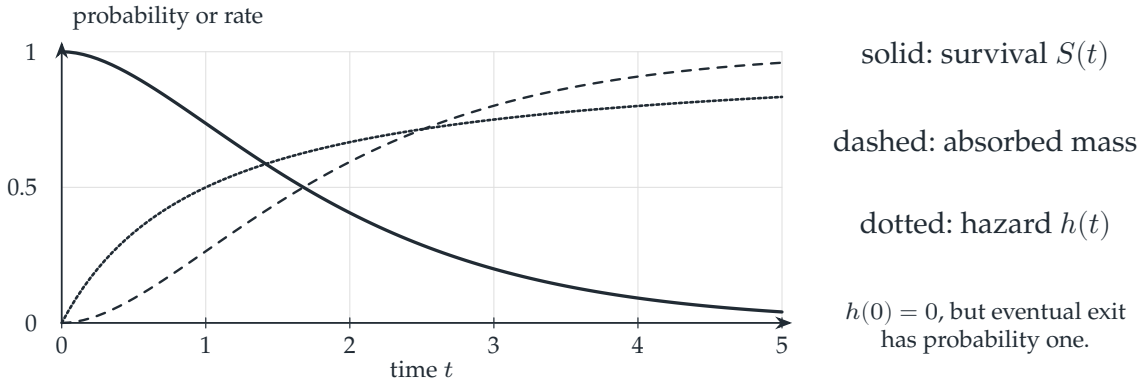
Let a transient process move through two successive rate-one stages, then leave the system. With column probabilities, the transient generator and starting vector are

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

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.



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 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 Part I: Proposition 9.3 with Theorems 5.1 and 8.2. Markov closure and likelihood sufficiency are separate.

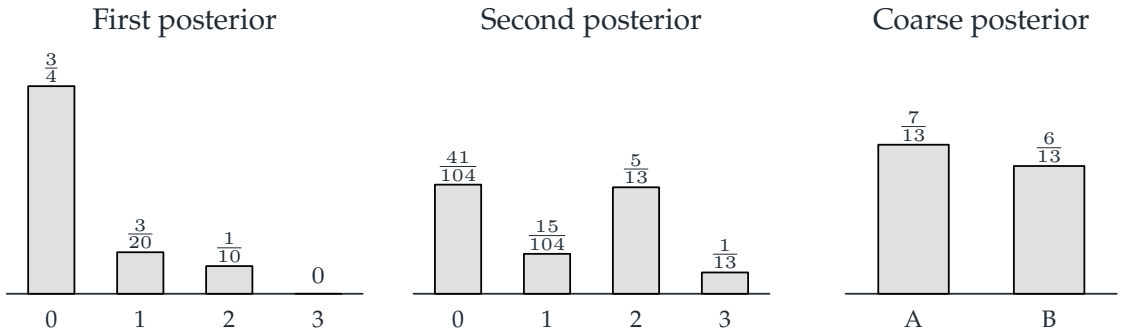
Use the four-state kernel K and aggregation C from V5, starting at $p_0 = e_0$. Observe two likelihood vectors constant inside both packets:

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

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

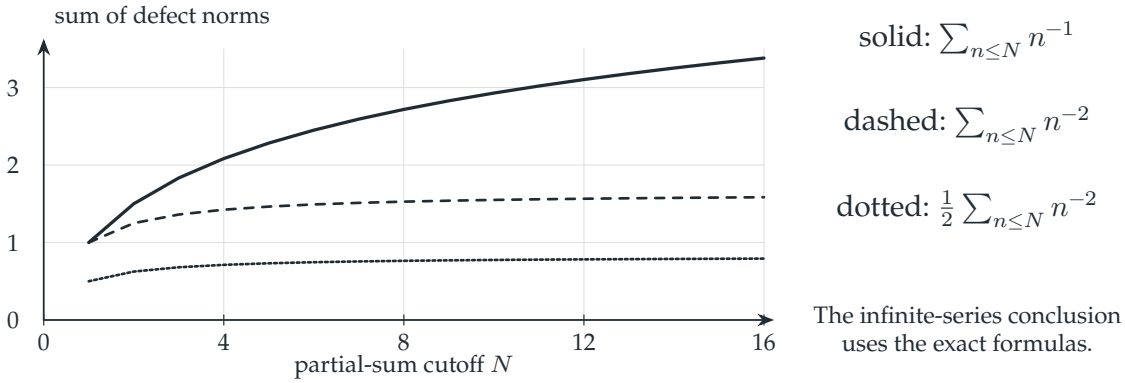
Small at each level is not summable across levels

Traditional 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

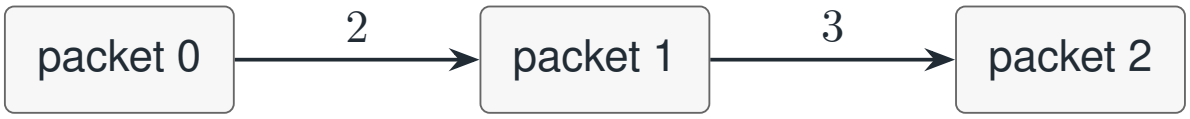
Traditional Part II: Theorem E2.8 and Corollary E2.9. Acyclicity concerns the transfer graph.

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

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



$$1 \qquad 2t \qquad 3t^2$$

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

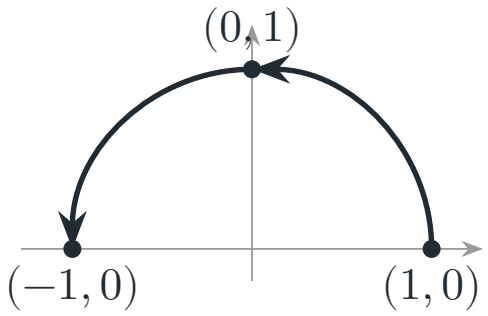
Convergence does not close the dynamics

Traditional 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

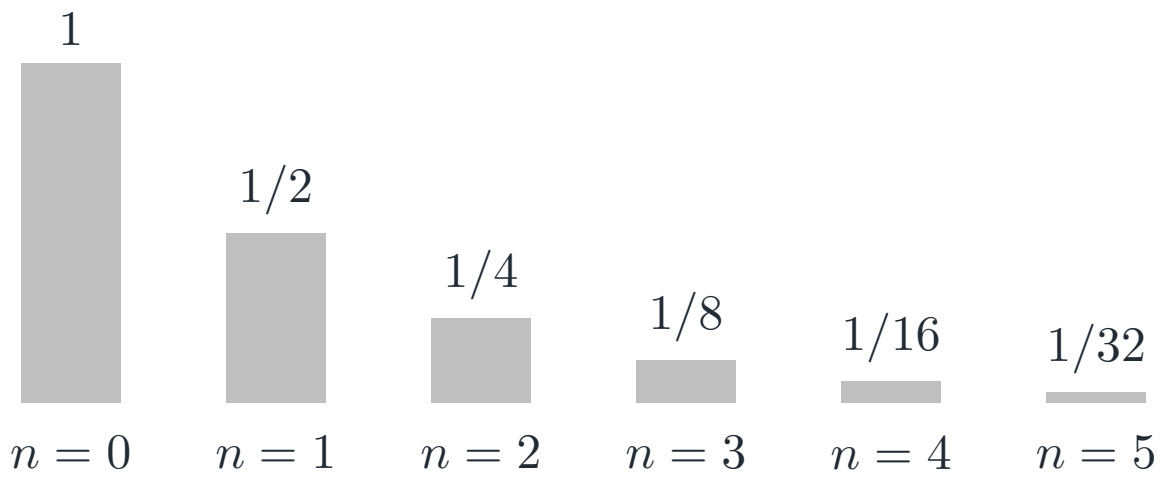
Traditional Part II: Theorem E4.5 and Corollary E4.7. This example has identity connecting maps.

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

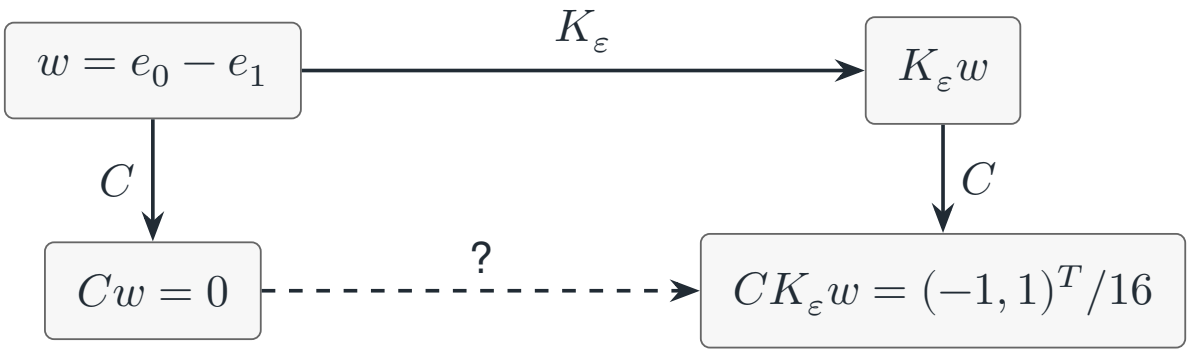
Traditional Part II: Theorem E5.3 and Corollary E5.4. A quotient criterion with a concrete witness.

Return to the four-state model of V–4 and the perturbed stochastic matrix K_ε 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

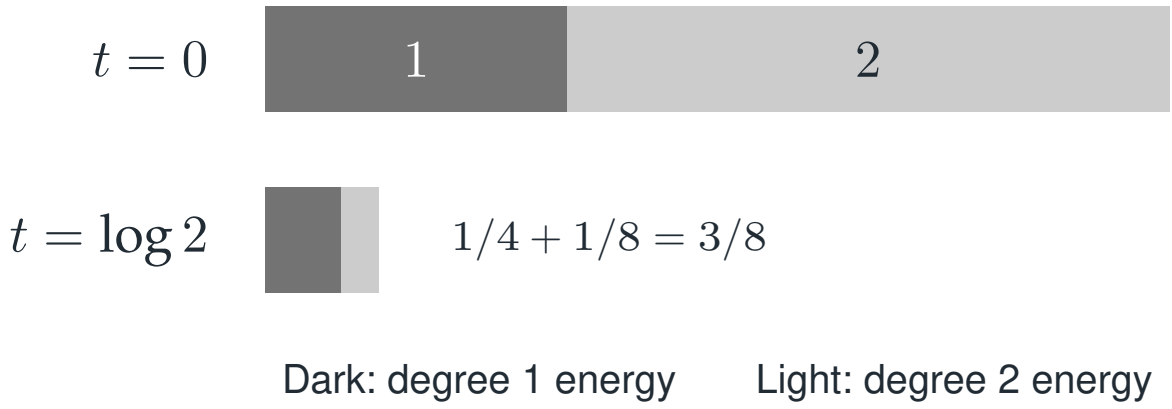
Gaussian packets have separate clocks

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

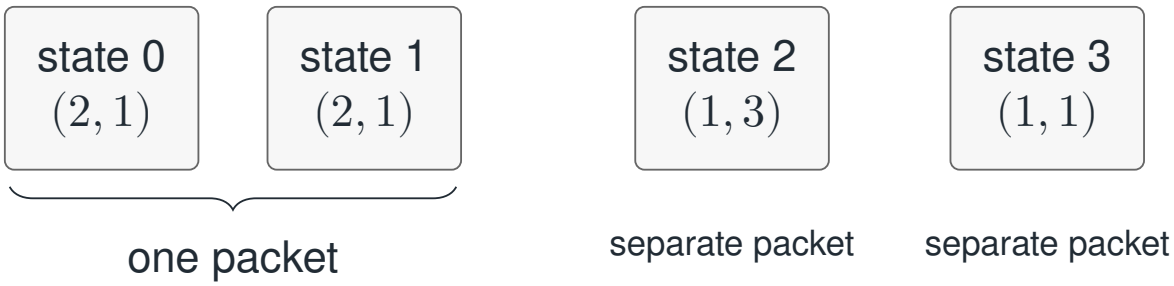
On four states, consider two likelihoods

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

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 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

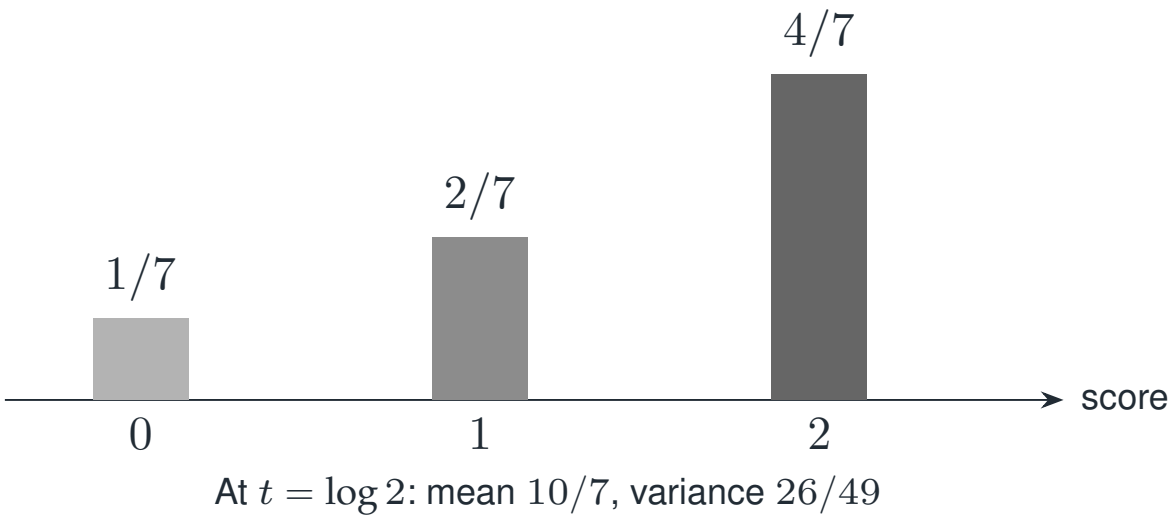
The score rises because it has variance

Traditional 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

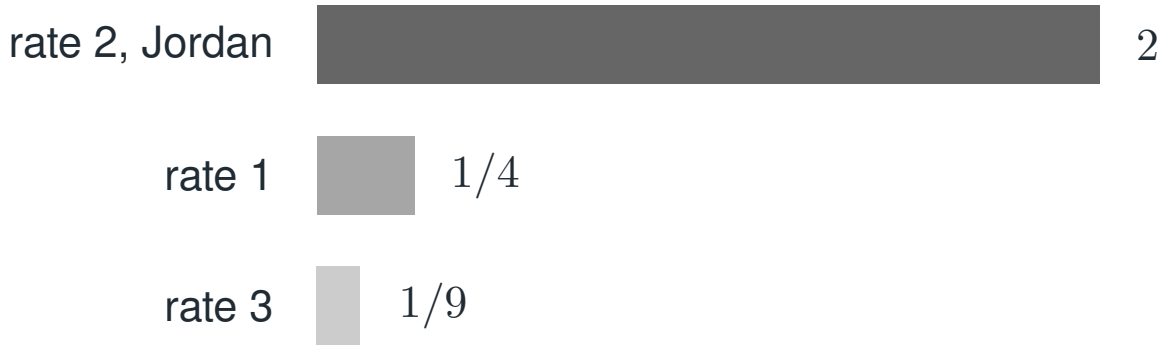
Traditional Part II: Theorem E15.1 and Corollary E15.2. Stable primary blocks; block count is the budget.

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

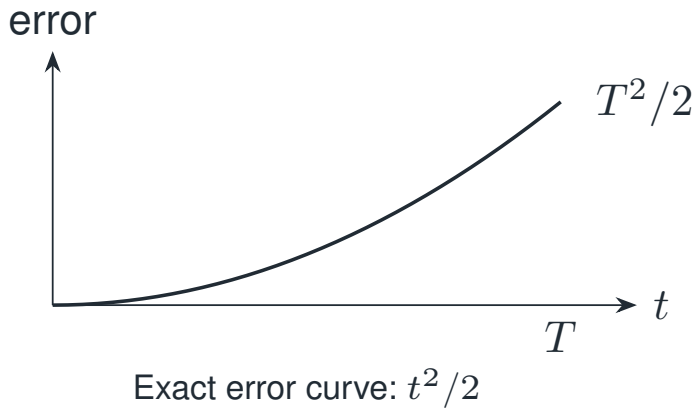
A kernel certificate becomes a trajectory certificate

Traditional 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

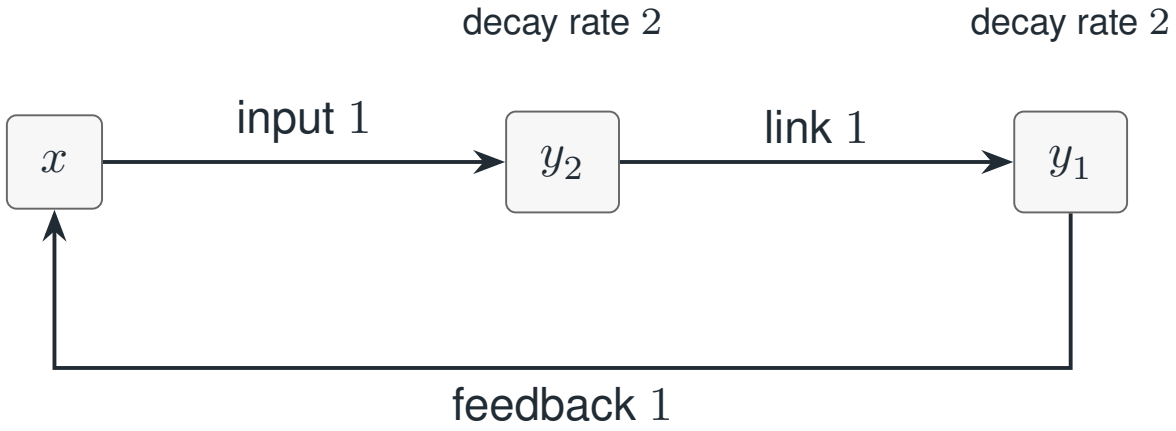
Two auxiliary states realize a repeated pole

Traditional 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

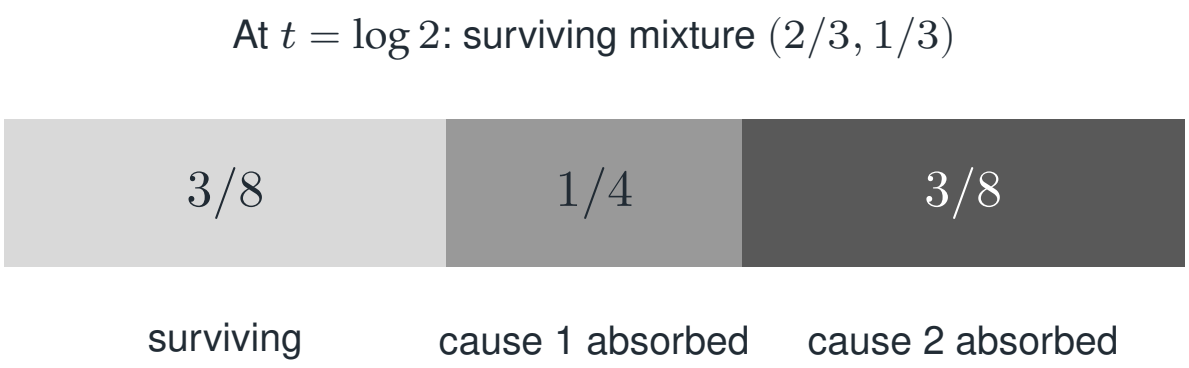
Survivors have a different mixture of causes

Traditional 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

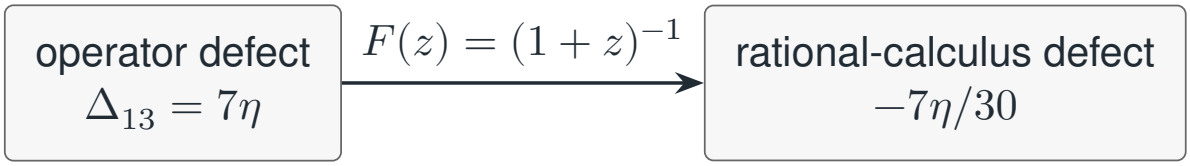
An operator substitution transports the defect too

Traditional 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 η.

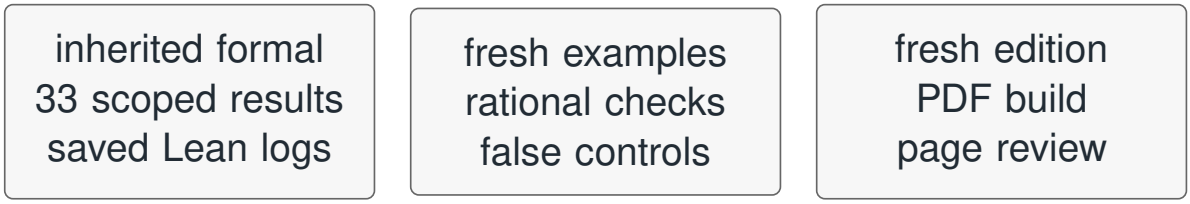
The general holomorphic defect formula integrates the corresponding resolvent defect under its contour and spectral assumptions. This rational example verifies an actual matrix inverse identity. It does not remove those analytic hypotheses or establish a calculus for arbitrary unbounded operators.

> **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.



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.

References

[ANO96] Søren Asmussen, Olle Nerman, and Marita Olsson. Fitting phase type distributions via the expectation maximization algorithm. *Scandinavian Journal of Statistics*, 23(4):419–441, 1996.

[Bal60] A. V. Balakrishnan. Fractional powers of closed operators and the semigroups generated by them. *Pacific Journal of Mathematics*, 10(2):419–437, 1960.

[BS94] José M. Bernardo and Adrian F. M. Smith. *Bayesian Theory*. John Wiley and Sons, Chichester, 1994.

[CHK00] Alexandre J. Chorin, Ole H. Hald, and Raz Kupferman. Optimal prediction and the mori zwanzig representation of irreversible processes. *Proceedings of the National Academy of Sciences of the United States of America*, 97(7):2968–2973, 2000.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley and Sons, Hoboken, New Jersey, second edition, 2006.

[DS58] Nelson Dunford and Jacob T. Schwartz. *Linear Operators. Part I: General Theory*, volume 7 of *Pure and Applied Mathematics*. Interscience Publishers, New York, 1958.

[EN00] Klaus Joachim Engel and Rainer Nagel. *One Parameter Semigroups for Linear Evolution Equations*, volume 194 of *Graduate Texts in Mathematics*. Springer, New York, 2000.

[Fes58] Herman Feshbach. Unified theory of nuclear reactions. *Annals of Physics*, 5(4):357–390, 1958.

[Jan97] Svante Janson. *Gaussian Hilbert Spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997.

[Kac49] Mark Kac. On distributions of certain wiener functionals. *Transactions of the American Mathematical Society*, 65(1):1–13, 1949.

[Kai80] Thomas Kailath. *Linear Systems*. Prentice Hall, Englewood Cliffs, New Jersey, 1980.

[KS60] John G. Kemeny and J. Laurie Snell. *Finite Markov Chains*. D. Van Nostrand, Princeton, New Jersey, 1960.

[Neu81] Marcel F. Neuts. *Matrix Geometric Solutions in Stochastic Models*. The Johns Hopkins University Press, Baltimore, 1981.

[Nor97] J. R. Norris. *Markov Chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1997.

[Paz83] Amnon Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*, volume 44 of *Applied Mathematical Sciences*. Springer, New York, 1983.

[Rab89] Lawrence R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[RP81] L. C. G. Rogers and J. W. Pitman. Markov functions. *The Annals of Probability*, 9(4):573–582, 1981.

[SSV12] René L. Schilling, Renming Song, and Zoran Vondraček. *Bernstein Functions*, volume 37 of *De Gruyter Studies in Mathematics*. De Gruyter, Berlin, second edition, 2012.

[Str04] John C. Strikwerda. *Finite Difference Schemes and Partial Differential Equations*. Society for Industrial and Applied Mathematics, Philadelphia, second edition, 2004.

[Tay] Michael E. Taylor. Lectures on Banach algebras. Author-hosted lecture notes. Undated; Appendix A. Accessed 6 October 2026.

[Wil91] David Williams. *Probability with Martingales*. Cambridge University Press, Cambridge, 1991.

[Zwa61] Robert Zwanzig. Memory effects in irreversible thermodynamics. *Physical Review*, 124(4):983–992, 1961.