REALIZATION CALCULUS

# Large Signal Model

Foundations of realization, refinement and predictive closure

Komorebi P.

# Contents

**Preface** v

Scope and evidence . . . . . v

Route through the volume . . . . . vi

**1 Realization and predictive closure** 1

1. Three objects that must remain distinct . . . . . 1

2. Exact realization and atlas consistency . . . . . 1

3. Coordinate transport and one underlying operator . . . . . 2

4. Packets, refinement and transfer blocks . . . . . 3

5. Predictive closure is the central contract . . . . . 5

6. Constructing the smallest closed linear refinement . . . . . 7

7. Joint closure of prediction and Bayesian inference . . . . . 8

8. Memory is an exact alternative to adding state . . . . . 9

9. Statistical refinement needs its own theorem . . . . . 10

10. Bayesian coarsening and the information-loss packet . . . . . 10

11. What a typed operator language must guarantee . . . . . 12

12. A bounded infinite-resolution bridge . . . . . 12

13. A complete small signal example . . . . . 13

14. Connection to observable learning . . . . . 14

**2 Learning, stability and coherent expansion** 15

15. From hidden state to an accessible predictive state . . . . . 15

16. What finite data identify . . . . . 16

17. Why finite agreement is not unrestricted discovery . . . . . 18

18. A complete output-only identification example . . . . . 19

19. A single propagation law for typed defects . . . . . 20

20. Truncating memory without losing the prediction bound . . . . . 22

21. Nonlinear packet truncation with an explicit remainder . . . . . 24

22. Learning several charts without inventing several worlds . . . . . 27

23. Refinement cannot manufacture missing information . . . . . 28

24. A growing operator grammar with coherent evidence . . . . . 29

25. The finite learned-realization theorem . . . . . 31

26. Connection to controlled prediction . . . . . 32

**3 Controlled stochastic realization** 33

27. Actions and observations are different inputs . . . . . 33

28. The state is a function of possible future experiments . . . 35
29. Recovering the operators from an observable probability table . . . 36
30. Probability validity is not a rank condition . . . 37
31. Packets can reconstruct a belief without being beliefs . . . 38
32. Action coverage is a mathematical requirement . . . 39
33. Learning the coordinate space under bounded noise . . . 40
34. Noisy operator reconstruction in a fixed core . . . 41
35. Conditioning a prediction can amplify its error . . . 41
36. A finite experimental route to the probability table . . . 43
37. A filtering defect with a packet interpretation . . . 44
38. A complete two-action rational example . . . 45
39. Integrated controlled stochastic realization . . . 46
40. From realization to admissible domains . . . 47

**4 Certified probability domains and robust learning** 48

41. The object to be certified . . . 48
42. Finite tests give an outer hierarchy, not automatic closure . . . 49
43. A finite certificate with a positive lift . . . 50
44. Robustness is a quantified invariance statement . . . 51
45. Fitting inside a certified domain . . . 52
46. A norm that measures probability mass before conditioning . . . 53
47. Whole observation laws can be stable without rare-event margins . . . 54
48. Conditioning still has an unavoidable cost . . . 55
49. Searching for a domain from observable operators . . . 56
50. Selection and noise do not invalidate an exact witness . . . 58
51. Why finite predictive rank need not mean finitely many positive states . . . 58
52. A finite approximation with an explicit price . . . 60
53. Repeated proposal and verification with a confidence budget . . . 61
54. Integrated certificate-carrying realization . . . 62
55. From polyhedral to curved certificates . . . 63

**5 Curved domains and continuous observations** 64

56. A matrix certificate for a curved domain . . . 64
57. The earlier disk obstruction has an exact finite matrix witness . . . 66
58. An observation density is not a branch probability . . . 67
59. A finite certificate over an entire observation interval . . . 68
60. Learning from integral observations rather than point probabilities . . . 70
61. Finite moments cannot validate an unrestricted density class . . . 71
62. Exact observation bins preserve their own event law . . . 72
63. Averaging transformations has a covariance defect . . . 73
64. A valid density approximation with a resolution bound . . . 73

65. Full continuous record laws inherit the stability argument . . . . . 74
66. A complete continuous-observation example . . . . . 76
67. A declared template can be recognized without being rediscovered . . . . . 79
68. The combined continuous polynomial sector . . . . . 79
69. From continuous semantics to discovery . . . . . 80

**6 Certificate discovery, approximation and sequential evidence** 81
70. An ellipsoidal certificate with an orientation test . . . . . 81
71. Geometry can be proposed from the operators themselves . . . . . 83
72. Approximate an unknown density without assuming its exact degree . . . . . 86
73. Keep approximation bias separate from coefficient error . . . . . 88
74. Evidence from dependent observations . . . . . 89
75. Actions need coverage, not only more observations . . . . . 90
76. What a residual can prove about misspecification . . . . . 91
77. Comparing a model with a process outside its representation . . . . . 92
78. An integrated theory of construction and correction . . . . . 93
79. From exact geometry to robustness . . . . . 94

**7 Robust curved certificates and noisy geometry** 95
80. A margin that has probability meaning . . . . . 95
81. Why exact boundary preservation can have zero noise radius . . . . . 96
82. An explicit repair and the price of its slack . . . . . 97
83. Noisy invariant forms are proposals, not certificates . . . . . 98
84. Finite exact checking on a continuum . . . . . 100
85. A finite uncertainty box and the role of convexity . . . . . 101
86. What a search can guarantee . . . . . 102
87. Selection from data does not remove the uncertainty quantifier . . . . . 103
88. A complete perturbed-operator construction . . . . . 103
89. The prediction cost needs a common comparison domain . . . . . 105
90. From deterministic robustness to data confidence . . . . . 106

**8 Data-supported domains and smaller repairs** 107
91. A smaller repair can be bracketed rather than guessed . . . . . 107
92. Confidence from dependent observations . . . . . 110
93. From observable moment regression to operator sets . . . . . 111
94. A nonlinear observable state with a curved moment domain . . . . . 112
95. A fully rational dependent-data reference experiment . . . . . 115
96. Structured uncertainty has an observable error bound . . . . . 118
97. Long-horizon marginal control is not full-record control . . . . . 119
98. A conditional data-to-domain theorem . . . . . 120
99. Two distinct completed reference sectors . . . . . 121
100. Scope and research boundary . . . . . 121

**A Proof and evidence register** 122
Realization and predictive closure . . . 122
Learning and stability . . . 125
Controlled stochastic realization . . . 128
Probability-domain certificates . . . 131
Curved domains and continuous observations . . . 135
Discovery, approximation and evidence . . . 138
Robust curved certificates . . . 140
Data-supported domains . . . 143

**B Formal companion** 145
What the archive checks . . . 145
Coverage at a glance . . . 145
Finite-table spectral stability . . . 156
Reproducing the formal evidence . . . 158

**Glossary** 160

**References and source lineage** 163

# Preface

A large signal model joins a representation to the operations that act on it. The central question is whether that representation retains the information required for prediction and inference. Exact storage, predictive closure, probability validity and accuracy each impose a different contract.

The deterministic contract is

$$\gamma F_u = \bar{F}_u \gamma.$$

When this identity fails, the model records a defect. It may enlarge the state, retain memory or accept a quantified approximation. The probabilistic contract adds an admissible domain, a mass functional and observation operators. Their products must represent nonnegative, normalized event laws, and their positive-evidence updates must remain admissible.

The word *large* describes a structured hypothesis family of coordinate systems, packet resolutions, typed operators, compositions, memory mechanisms and statistical alternatives. It does not specify a parameter count or assert universal discovery.

## Scope and evidence

The volume contains 87 numbered propositions with written proofs. Their classical antecedents are attributed in the references. A proposition count is not a novelty count.

The accompanying Lean project records complete formal coverage for 87 propositions in their stated mapped scopes, no partial mappings, and 0 propositions without a formalization. These counts describe manuscript correspondence, not proof difficulty. The kernel checks formal declarations. Matching those declarations to the English propositions remains a separate reviewed step.

Finite symbolic and numerical checks exercise particular examples and certificate replays. They do not replace general proofs. The formal companion reports the current clean Lean compilation, axiom audit and replay evidence with their exact scopes. No independent external referee review is claimed.

Four distinctions guide the reading. Lossless storage does not imply predictive sufficiency. A valid probability model need not be accurate. An exact certificate for a fitted model does not establish coverage of an unknown generating process. A uniform single-time marginal bound is not a uniform bound on complete observation records.

## Route through the volume

The original section and proposition numbers remain fixed. The eight chapters provide a thematic structure around them.

| Chapter | Sections | Main question |
|---|---|---|
| 1. Realization and predictive closure | 1–14 | What survives representation and compression? |
| 2. Learning and stability | 15–26 | What can finite observable data identify? |
| 3. Controlled stochastic realization | 27–40 | What does a predictive probability state represent? |
| 4. Probability domains | 41–55 | What gives fitted operators probability meaning? |
| 5. Curved and continuous models | 56–69 | How do finite certificates describe continuous observations? |
| 6. Discovery and evidence | 70–79 | What separates proposals, approximation and rejection? |
| 7. Robust curved certificates | 80–90 | What survives operator uncertainty, and at what cost? |
| 8. Data-supported domains | 91–100 | When does a covered parameter set certify a domain? |

The proof and evidence register records each proposition's assumptions, principal dependencies and code counterparts. The formal companion describes the current kernel checks, reproduces two complete endpoint modules and summarizes the additional formal results. The glossary and references close the volume. The source package retains the original manuscripts, scientific figures, reference laboratory, formal sources and verification records.

Read the surrounding definitions before using a proposition. Some assumptions precede its bold statement. In particular, Proposition 9 uses Proposition 12, a forward reference in the presentation. The principal dependency register is acyclic, but is not a complete formal proof graph.

CHAPTER 1

# Realization and predictive closure

## 1. Three objects that must remain distinct

An **observation** is an element $x$ of a declared space $X$. Examples include a finite real signal, a complex field on a finite grid, or an array of categorical labels. Its type includes shape, scalar field or palette, and boundary convention. A zero-extended finite sequence and a periodic sequence are different objects even when their displayed samples agree.

An **archive** is a lossless representation of the accepted observation. In a mathematical model this means an injective analysis map with an explicit left inverse. In an implementation, conversion and floating-point error must be accounted for separately. An archive of a measured signal does not recover physical variables that were never observed.

A **predictive state** is a summary $z = \gamma(x)$, or more generally a summary of an observation history. It is sufficient only relative to a specified family of future quantities, actions and dynamics. There is no implication from exact storage of $x$ to sufficiency of $\gamma(x)$.

The historical model's macro, meso and micro vectors are nested projections of a feature vector. They are not, merely by being nested, state variables on which the signal dynamics close. The distinction is retained in the present theory.

The term **large** denotes the structured hypothesis family: representations, resolutions, typed operators, compositions, memory mechanisms and their parameters. It is not a theorem about capacity, sample complexity or scale. Such a theorem needs a declared grammar, a complexity measure and a learning protocol.

## 2. Exact realization and atlas consistency

### 2.1 A left inverse is the reconstruction contract

Let $X$ and $C_a$ be finite-dimensional Hilbert spaces. A linear atlas consists of analysis and synthesis maps

$$A_a : X \longrightarrow C_a, \qquad S_a : C_a \longrightarrow X, \qquad S_a A_a = I_X.$$

The coefficients representing actual signals form $\mathcal{C}_a = \operatorname{im} A_a$. This distinction matters for redundant representations, where $\mathcal{C}_a$ is a proper subspace of the ambient coefficient space.

**Proposition 1. Exact realization and consistency.** The analysis map is injective, $\Pi_a = A_a S_a$ is an idempotent with image $\mathcal{C}_a$, and $\Pi_a c = c$ exactly when $c$ is a realizable coefficient vector. Moreover,

$$\|A_a x\| \geq \frac{\|x\|}{\|S_a\|}, \qquad \|S_a(c+e) - S_a c\| \leq \|S_a\|\,\|e\|.$$

The first inequality is understood for nonzero $X$. No assertion of orthogonality follows from a left inverse alone.

**Proof.** If $A_a x = 0$, then $x = S_a A_a x = 0$. Also $\Pi_a^2 = A_a(S_a A_a)S_a = \Pi_a$. Its image is contained in $\mathcal{C}_a$ and it fixes every $A_a x$, so its image is exactly $\mathcal{C}_a$. Both norm inequalities follow from the operator-norm inequality applied to $S_a$. □

For a unitary atlas $U_a$, the synthesis is $U_a^*$ and the condition number is one. For a learned invertible atlas, exact invertibility is not enough: the synthesis norm controls amplification of coefficient errors. An arbitrarily ill-conditioned exact transform is not a stable representation.

### 2.2 Several exact atlases do not create new observations

For unitary $U_a$ and positive weights $w_a$, put

$$\mathcal{A}x = (w_a U_a x)_a, \qquad \mathcal{S}(c_a)_a = \frac{1}{c}\sum_a w_a U_a^* c_a, \qquad c = \sum_a w_a^2.$$

Then $\mathcal{S}\mathcal{A} = I$ and $\|\mathcal{A}x\|^2 = c\|x\|^2$, as proved in [K1]. Here $\mathcal{A}\mathcal{S} = c^{-1}\mathcal{A}\mathcal{A}^*$ is the orthogonal projection onto the consistent coefficient subspace.

Each complete atlas contains the same observed information because it has a left inverse. Multiple atlases offer different sparsity, locality, conditioning and operator descriptions. They do not provide independent observations of the world. In particular, multiplying likelihoods from deterministic recodings as though they were independent measurements generally counts the same evidence more than once.

Categorical one-hot encoding is a separately typed injective map of sets. It is not a unitary map on the scalar-valued signal space. Its exact decoder is valid on one-hot arrays, not on arbitrary coefficient vectors. The linear tight-frame theorem does not apply to it without a separately specified linear embedding and inner product. This separation is already explicit in the capstone's categorical section.

## 3. Coordinate transport and one underlying operator

For a nonempty atlas family, define transport on realizable coefficients by

$$T_{ba} = A_b S_a : \mathcal{C}_a \longrightarrow \mathcal{C}_b.$$

For an endomorphism $F : X \to X$, define its coordinate realization by $F_a = A_a F S_a$ on $\mathcal{C}_a$. In this paragraph $F$ may be nonlinear. For a partial map, every formula is restricted to the transported admissible domain. Maps between different signal types use separate input and output atlases.

**Proposition 2. Coherent transport and operator realization.** On the consistent coefficient spaces,

$$T_{aa} = I, \qquad T_{cb}T_{ba} = T_{ca}, \qquad F_b T_{ba} = T_{ba} F_a.$$

Conversely, suppose maps $G_a : \mathcal{C}_a \to \mathcal{C}_a$ satisfy $G_b T_{ba} = T_{ba} G_a$ for every pair. Then there is a unique map $F : X \to X$ with $G_a = A_a F S_a$ for every $a$.

**Proof.** The transport identities follow by inserting $S_b A_b = I$. The operator identity follows because either path is $A_b F S_a$. For the converse, choose a reference atlas $r$ and put $F = S_r G_r A_r$. Compatibility gives $G_a A_a = A_a F$. Every consistent coefficient is $A_a x$, which proves the asserted realization. Applying $S_a$ proves uniqueness. □

An independently fitted family of atlas predictors need not satisfy this proposition. Its compatibility residual is a well-defined quantity,

$$d_{ba}(x) = G_b(A_b x) - T_{ba} G_a(A_a x).$$

After synthesis it becomes an object-space discrepancy:

$$S_b d_{ba}(x) = S_b G_b A_b x - S_a G_a A_a x.$$

Thus $\|S_b d_{ba}(x)\| \le \|S_b\| \, \|d_{ba}(x)\|$. Zero discrepancy establishes that the chart descriptions agree, not that their common prediction is correct. All charts can agree on the same wrong operator.

Redundant coefficient outputs must also be tested for consistency. Synthesizing an arbitrary coefficient vector always produces some signal, but reanalysis returns $\Pi_a c$, not necessarily the original $c$. Reconstruction and coefficient consistency are different checks.

## 4. Packets, refinement and transfer blocks

### 4.1 Packets belong to a declared resolution

Within one orthogonal atlas, let $E_i$ be pairwise orthogonal coordinate selectors with $\sum_i E_i = I$. Pull them back by a unitary map:

$$P_i = U^* E_i U, \qquad P_i P_j = \delta_{ij} P_i, \qquad P_i^* = P_i, \qquad \sum_i P_i = I_X.$$

A packet is the component $P_i x$ together with its projector, domain, atlas and resolution. A bare coefficient block without this information is not a fully specified packet.

Projectors from different atlases need not commute and must not be treated as one larger mutually orthogonal family. Weighted frame packets can sum to the identity while failing to be idempotent. These are different structures.

**Proposition 3. Complete refinement preserves reconstruction.** Suppose a finite rooted tree has root projector $I$, and every internal projector is the sum of pairwise orthogonal child projectors. Every complete antichain $\Lambda$, meeting each root-to-leaf path exactly once, satisfies

$$\sum_{v\in\Lambda} P_v = I, \qquad x = \sum_{v\in\Lambda} P_v x, \qquad \|x\|^2 = \sum_{v\in\Lambda} \|P_v x\|^2.$$

**Proof.** Replace an internal node by the sum of its children. Repeated replacement reaches any complete antichain and preserves the total projector. Different selected nodes have disjoint descendant branches, hence orthogonal ranges. The norm identity is Pythagoras. □

This is a conservation law for information under complete refinement. Deleting a child, storing only its energy, or imposing a coefficient threshold changes the contract. For a retained subset $J$ of an orthogonal decomposition, the lost signal is exactly $\sum_{i\notin J} P_i x$, and its squared norm is the sum of the omitted packet energies.

The finite best-basis recursion

$$V(v) = \min\left(c(v), \sum_{u\in \mathrm{ch}(v)} V(u)\right)$$

minimizes an additive cost over the declared finite tree [K1, R1]. The recurrence itself does not make a nonadditive predictive objective additive. If a prediction cost couples two leaves, a scalar cost at each node is generally insufficient for global optimization. Packet construction and cost evaluation must also be included in an implementation's complexity account.

### 4.2 Operations, not just signals, decompose

Let $L : X \to Y$ be linear, with orthogonal input projectors $P_i$ and output projectors $Q_j$. Its transfer blocks are $L_{ji} = Q_j L P_i$.

**Proposition 4. Transfer reconstruction and refinement.** One has

$$L = \sum_{j,i} Q_j L P_i.$$

If $P_i = \sum_r P_{i,r}$ and $Q_j = \sum_s Q_{j,s}$ are complete refinements, then

$$Q_j L P_i = \sum_{s,r} Q_{j,s} L P_{i,r}.$$

For another linear map $M : Y \to Z$ and output projectors $R_k$,

$$R_k M L P_i = \sum_j (R_k M Q_j)(Q_j L P_i).$$

**Proof.** Insert the appropriate identity decomposition on the input, output or intermediate space. Use $Q_j^2 = Q_j$ in the composition formula. □

This is the algebraic reason to make transfer blocks first-class objects. The same operation can be expressed at several resolutions, while the coarse and fine expressions remain linked by exact identities. Approximations acquire explicit omitted-block terms rather than an unnamed representation error.

### 4.3 Nonlinear operations need interactions

For a nonlinear map, $F(\sum_i P_i x)$ generally differs from $\sum_i F(P_i x)$. Applying a linear transfer formula to such a map silently loses interactions.

**Proposition 5. Polynomial packet interactions.** Let

$$F(x) = b + \sum_{m=1}^{p} B_m(x, \ldots, x),$$

where each $B_m : X^m \to Y$ is multilinear. Then

$$Q_j F(x) = Q_j b + \sum_{m=1}^{p} \sum_{i_1, \ldots, i_m} Q_j B_m(P_{i_1} x, \ldots, P_{i_m} x).$$

**Proof.** Substitute $x = \sum_i P_i x$ in every argument and distribute by multilinearity. □

For example, $(u + v)^2 = u^2 + 2uv + v^2$. A grammar containing only independent packet squares misses $2uv$. For nonpolynomial maps, one needs direct nonlinear semantics or a local expansion with a stated remainder bound. The word algebraic does not remove that obligation.

## 5. Predictive closure is the central contract

Let $\gamma : X \to Z = \gamma(X)$ be a summary and let $F_u : X \to X$ describe a deterministic transition for action $u$. The reduced description closes when there is $\bar{F}_u : Z \to Z$ such that

$$\gamma F_u = \bar{F}_u \gamma.$$

This is an equality of functions on the declared domain, not a fit statistic.

**Proposition 6. Exact quotient criterion.** Such a reduced transition exists if and only if

$$\gamma(x) = \gamma(x') \implies \gamma(F_u(x)) = \gamma(F_u(x'))$$

for every admissible $x, x'$ and every action. When it exists it is unique on $\gamma(X)$.

**Proof.** Necessity follows by applying $\bar{F}_u$ to equal summaries. For sufficiency, define $\bar{F}_u(z) = \gamma(F_u(x))$ for any representative with $\gamma(x) = z$. The stated condition makes this well-defined and also proves uniqueness. □

This criterion also applies when $X$ is a latent physical state and $\gamma$ is an observation map. Two physically different states can produce identical images and different futures. Exact image reconstruction does not resolve that hidden-state ambiguity. A history, a belief state or additional observations may be needed.

For a stochastic kernel, the corresponding condition is equality of the distributions of the next summary for states with the same current summary. Section 7 makes this finite criterion explicit.

### 5.1 The obstruction in the linear case

Suppose $X$ is a Hilbert space, $P$ is an orthogonal projection, $Q = I - P$ and $x_{t+1} = Lx_t$. The retained state is $z_t = Px_t$. Then

$$z_{t+1} = PLPz_t + PLQx_t.$$

**Proposition 7. Closure defect and compositional defect.** The retained state closes for every initial signal if and only if $PLQ = 0$. If it does not close, the error in using only $PLP$ is

$$e_t = PLQx_t, \qquad \|e_t\| \le \|PLQ\| \, \|Qx_t\|.$$

For two linear operators $L$ and $M$,

$$\boxed{PLMP - (PLP)(PMP) = PLQMP.}$$

Consequently, compressing two operations separately can discard a path that leaves the retained subspace and later returns.

**Proof.** If $PLQ = 0$, the displayed recurrence depends only on $z_t$. Conversely, apply Proposition 6 to $x$ and $Px$, which have the same retained state, to obtain $PLQx = 0$ for every $x$. The norm bound is submultiplicativity. The compositional identity follows by inserting $I = P + Q$ between $L$ and $M$. □

The relevant quantity is not omitted energy alone. A very small omitted packet can have a large effect if its transfer into retained packets is strongly amplified. Conversely, an energetic packet can be irrelevant to a specified prediction if every corresponding return block is zero.

A residual is therefore typed. Reconstruction residual, coefficient-consistency residual, predictive residual, closure residual and statistical uncertainty should not be collapsed into one scalar without a specified objective.

## 6. Constructing the smallest closed linear refinement

There is a constructive finite solution when the dynamics are a known finite family of linear maps $L_u$ on a $d$-dimensional space. It identifies which additional observables are required, rather than merely requesting a larger state vector.

Let the initial summary be $C_0 x$. Regard its rows as linear observables. Starting from $W_0 = \mathrm{row}(C_0)$, define

$$W_{n+1} = W_n + \sum_u W_n L_u, \qquad W_n L_u = \{wL_u : w \in W_n\}.$$

**Proposition 8. Minimal linear predictive completion.** The sequence stabilizes after at most $d - \dim W_0$ strict enlargements. Its limit $W_*$ is the smallest row space containing $W_0$ and invariant under right multiplication by every $L_u$. If $R$ is any full-row-rank matrix whose rows form a basis of $W_*$, then

$$C_0 = HR, \qquad RL_u = B_u R$$

for suitable matrices $H$ and $B_u$. Therefore $z = Rx$ retains the original summary and has exact reduced dynamics $z_{t+1} = B_u z_t$. Every linear summary with those two properties has row space containing $W_*$.

**Proof.** The spaces are nested and have dimension at most $d$. Equality at one step means invariance under every $L_u$, so it is a fixed point. Containment in any invariant row space containing $W_0$ follows by induction. The matrix identities express inclusion of the relevant rows in the basis $R$. Conversely, if $C_0 = HR$ and $RL_u = B_u R$, the row space of $R$ is invariant and contains $W_0$, so it contains every $W_n$. □

For one transition matrix this reduces to the span of the rows of

$$C_0,\ C_0 L,\ \ldots,\ C_0 L^{d-1}.$$

Its annihilator is the largest $L$-invariant subspace inside $\ker C_0$, the standard unobservable-subspace construction [R2]. With several actions, observables pulled back through action words are required, but the linear span still stabilizes in finite dimension.

This gives a precise role for obstruction-driven refinement: add observable directions created by the dynamics until the chosen prediction variables close. It does not guarantee that the completion is small. In some systems $W_* = X^*$ and exact memoryless closure needs the full state.

The theorem assumes the operators and relevant state are available. Applying it to estimated operators constructs closure for those estimates. It is not a proof that the real system has the

same closure. If only $C_0 x$ is measured and the full $x$ is unavailable, the augmented state $Rx$ is not automatically accessible.

## 7. Joint closure of prediction and Bayesian inference

The finite categorical case admits an equally explicit construction that joins dynamics with Paper VI's inference refinement. This is useful because a state can be closed for prediction but too coarse for a new observation, or sufficient for an update but not for the next transition.

Let $\Omega$ be a finite state set. Let $K_u(i,j)$ be known row-stochastic transition matrices. Let $\ell_y(i) > 0$ be a finite family of observation likelihoods. A partition $\pi$ has aggregation matrix

$$C_\pi(a,i) = \mathbf{1}_{i \in B_a}.$$

Distributions are column vectors. Prediction is $p \mapsto K_u^\top p$, and a Bayesian update is

$$\mathcal{B}_{\ell_y}(p) = \frac{p \odot \ell_y}{\ell_y^\top p}.$$

The coarse state must commute with both operations:

$$C_\pi K_u^\top = \bar{K}_u^\top C_\pi, \qquad C_\pi \mathcal{B}_{\ell_y}(p) = \mathcal{B}_{\bar{\ell}_y}(C_\pi p).$$

Fix a mandatory initial partition $\pi_0$ with at least two blocks. This preserves distinctions that the intended application requires and excludes the trivial one-state quotient.

**Proposition 9. Coarsest joint finite closure.** First split $\pi_0$ by the likelihood vectors $(\ell_y(i))_y$. Then repeatedly split each current block by the signatures

$$\left( \sum_{j \in B} K_u(i,j) \right)_{u,B},$$

where $B$ ranges over the current partition. The process terminates after at most $|\Omega| - |\pi_{\text{init}}|$ strict refinement rounds. The final partition is the unique coarsest refinement of $\pi_0$ on which all declared predictions and Bayesian updates close for every prior.

**Proof.** At a stable partition, states in a common block have the same transition probability into every coarse block. Those probabilities define $\bar{K}_u$ and give the first commutation identity, by checking it on point-mass priors. Likelihood constancy on each block defines $\bar{\ell}_y$ and gives the second identity by direct summation.

For minimality, every admissible partition refining $\pi_0$ has at least two blocks. By Proposition 12 below, exact updating for all priors requires each fine likelihood to be constant within each block. Thus it refines the first likelihood split. Suppose it refines a current partition. Each current block is a union of its blocks. Exact prediction makes the probability of reaching each such union constant

within any admissible block, so it also refines the next split. Induction proves it refines the final partition. Each strict refinement increases the number of blocks, which proves termination. □

The transition condition is the finite lumpability criterion [R4], expressed directly as a commutation identity. The joint construction adds the explicitly declared likelihood distinctions. Its output is a quotient of a known finite model, not an automatically learned model of an unknown world.

The accompanying rational example starts with six states. Likelihoods create three blocks, transition distinctions force two further rounds of splitting, and the final model has five states. Two original states remain genuinely equivalent for every action and observation in the declared family. The example verifies both commuting identities exactly.

## 8. Memory is an exact alternative to adding state

Refinement and memory are two ways to represent the influence of unresolved components. This can be derived rather than merely proposed as an architectural option.

For fixed linear dynamics and an orthogonal split, write

$$z_t = Px_t, \qquad q_t = Qx_t,$$

and define the restricted blocks

$$A = PLP, \quad B = PLQ, \quad C = QLP, \quad D = QLQ.$$

Thus $z_{t+1} = Az_t + Bq_t$ and $q_{t+1} = Cz_t + Dq_t$.

**Proposition 10. Exact projected memory and a conditional tail bound.** For every $t \geq 0$,

$$\boxed{z_{t+1} = Az_t + \sum_{k=0}^{t-1} BD^k Cz_{t-1-k} + BD^t q_0.}$$

If $\|D\| \leq \rho < 1$ and $\|z_s\| \leq Z$ over the relevant history, omitting terms with $k \geq H$ changes the one-step right-hand side by at most

$$\|B\|\,\|C\| Z \frac{\rho^H}{1-\rho}.$$

Dropping the initial unresolved-state term adds at most $\|B\|\rho^t\|q_0\|$.

**Proof.** Iteration of the unresolved recurrence gives

$$q_t = D^t q_0 + \sum_{s=0}^{t-1} D^{t-1-s} Cz_s.$$

Substitute this into the retained recurrence and change indices. Bound the omitted terms by the geometric series. □

The memory kernel is $BD^kC$. It records a path from retained variables into unresolved variables, evolution there, and return. The first kernel $BC$ is precisely the return path appearing in the two-step composition defect. The term $BD^t q_0$ is deterministic when $q_0$ is specified. It becomes uncertain only relative to incomplete information or a probability model for $q_0$.

This is the elementary linear instance of projection-memory reduction, associated with Mori-Zwanzig theory [R3]. A finite hand-chosen power-law or exponential memory feature is an approximation class for such effects, not a general derivation of the exact kernel.

The tail bound compares right-hand sides using the same resolved history. Replacing that history by recursively generated predictions introduces additional rollout error. Stability of that recursion needs its own bound. Also, a spectral-radius condition alone does not imply $\|D\| \le \rho$ in the norm already being used, especially for nonnormal operators.

## 9. Statistical refinement needs its own theorem

Nested feature vectors are algebraically nested when the smaller vector can be recovered from the larger by a fixed map. Predictive risk is a separate matter.

**Proposition 11. Ideal statistical benefit of refinement.** Let a target $Y$ take values in a finite-dimensional Hilbert space with $\mathbb{E}\|Y\|^2 < \infty$. Let $\mathcal{F} \subseteq \mathcal{G}$ be information sets and define $m_{\mathcal{F}} = \mathbb{E}[Y \mid \mathcal{F}]$ and $m_{\mathcal{G}} = \mathbb{E}[Y \mid \mathcal{G}]$. Then

$$\mathbb{E}\|Y - m_{\mathcal{F}}\|^2 = \mathbb{E}\|Y - m_{\mathcal{G}}\|^2 + \mathbb{E}\|m_{\mathcal{G}} - m_{\mathcal{F}}\|^2.$$

**Proof.** Expand $Y - m_{\mathcal{F}} = (Y - m_{\mathcal{G}}) + (m_{\mathcal{G}} - m_{\mathcal{F}})$. The cross term has zero conditional expectation given $\mathcal{G}$. □

Thus finer information cannot worsen the optimal squared-error predictor under a fixed distribution. It can worsen an estimated predictor with finite data, restricted function classes, changed regularization or shifted distributions. The theorem supplies the ideal target for refinement, not a guarantee for every fitting procedure.

An exact archive makes later reanalysis possible without reconstructing discarded observations from a lossy state. It does not eliminate estimation error or distribution shift.

## 10. Bayesian coarsening and the information-loss packet

This section imports a specific existing Komorebi foundation rather than inventing another inference heuristic. Let $p_i$ be a prior on a finite fine state set and let $B_a$ be coarse blocks. Define

$$P_a = \sum_{i \in B_a} p_i, \qquad r_a(p) = \frac{\sum_{i \in B_a} p_i \ell_i}{P_a} \quad (P_a > 0).$$

On zero-mass blocks the value of $r_a$ is immaterial. Updating finely and then aggregating is exactly updating $P$ with the effective likelihood $r(p)$. The effective likelihood generally depends on the fine prior inside each coarse block, information that $P$ alone does not retain.

**Proposition 12. Bayesian closure and its exact defect.** Suppose there are at least two coarse blocks and the fine and declared coarse likelihoods are strictly positive. Updating and coarsening commute for every prior if and only if

$$\ell_i = c\bar{\ell}_a \quad \text{for every } i \in B_a$$

with one common $c > 0$. For a fixed prior, put $\delta = r(p) - \bar{\ell}$, $\bar{b} = \mathcal{B}_{\bar{\ell}}(P)$ and $Z_r = r(p)^\top P$. Then

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

The fixed-prior identity also holds for nonnegative likelihoods when both normalizing evidences are positive.

**Proof.** Constancy up to one global factor gives commutation by summing the fine posterior. Conversely, compare priors supported on one state in each of two distinct coarse blocks. Equality of coarse posterior odds forces the ratio of those states' likelihoods to equal the ratio of the coarse likelihoods. Varying the representatives proves constancy within each block and the common factor across blocks. For the defect, use $Z_r = \bar{\ell}^\top P + P^\top \delta$ and subtract the two normalized vectors. □

With only one coarse block, both posteriors aggregate to one regardless of the likelihood, so the necessity criterion is false there. This boundary is explicitly retained from the corrected Paper VI statement [K2].

The matrix in the defect formula annihilates $\bar{\ell}$, recording the irrelevant common-scale direction. It does not make every likelihood discrepancy irrelevant. The construction in Proposition 9 refines until both the likelihood and transition defects vanish for a specified finite family.

### 10.1 Parameter inference is not program truth

Within one declared regression model, a normal-inverse-gamma prior gives a standard conditional Bayesian parameter posterior. The software's program weights instead use penalized held-out predictive scores and then may refit coefficients on all available training records. These are finite-candidate predictive weights. Without a corresponding generative derivation, they should not be conflated with the marginal-likelihood posterior over programs under one fixed prior.

Every candidate must be scored on the same target event and base measure. Densities on differently sized coordinate spaces are not directly comparable. An invertible nonlinear coordinate change requires the appropriate Jacobian. A redundant deterministic embedding induces a measure on its consistent subspace, not an independent full-dimensional density on all coefficient entries.

The law of total variance separates within-candidate uncertainty from disagreement of candidate means. It does not supply a probability that the truth lies outside the grammar. An explicit outside-grammar component needs its own prior and predictive distribution. Assigning an arbitrary leftover mass does not solve open-world inference.

## 11. What a typed operator language must guarantee

A program has a source type, target type, admissible domain, evaluation rule and declared approximation status. A nonlinear partial operation is not a total matrix merely because its inputs are stored in a vector. The domain of a composite is

$$\operatorname{dom}(g \circ f) = \{x \in \operatorname{dom} f : f(x) \in \operatorname{dom} g\}.$$

Composition must preserve those types and domains. Coordinate changes must preserve the represented operation as in Proposition 2. Packet refinement must preserve exact transfer semantics as in Proposition 4, or carry an explicit approximation residual. Nonlinear interactions must retain their cross-packet terms as in Proposition 5.

A grammar can be closed under a specified set of constructors without containing every possible physical law. It can also be expressive while computationally intractable. Universal discovery, practical search and semantic closure are distinct claims.

For a known operator, Proposition 8 can build its minimal linear summary. For an unknown operator, learning must address identifiability: which different candidates make identical predictions under the available observations and actions? More coordinate systems alone do not distinguish candidates that agree on the entire observed signal. Additional interventions or data may be needed.

The foundation should therefore treat a learned hypothesis as an operator with an evidence record, not just a matrix with a high normalized weight. Its record includes fitting assumptions, admissible inputs, compared alternatives, unresolved equivalences and the representation in which its residual is measured.

## 12. A bounded infinite-resolution bridge

Finite reconstruction identities do not alone prove convergence of an unbounded or infinite-dimensional model. A useful bounded extension can nevertheless be established directly.

**Proposition 13. Strong fixed-horizon refinement limit.** Let $X$ be a Hilbert space, $L : X \to X$ bounded, and $P_n$ orthogonal projections converging strongly to $I$. Put $L_n = P_n L P_n$. For every fixed $x \in X$ and every fixed nonnegative integer $k$,

$$L_n x \longrightarrow Lx, \qquad L_n^k x \longrightarrow L^k x.$$

**Proof.** Since $\|P_n\| \leq 1$,

$$\|(L_n - L)x\| \le \|L\| \, \|(P_n - I)x\| + \|(P_n - I)Lx\| \longrightarrow 0.$$

Also $\|L_n\| \le \|L\|$. Induct on $k$, using

$$(L_n^{k+1} - L^{k+1})x = L_n(L_n^k - L^k)x + (L_n - L)L^k x.$$

The first term tends to zero by the uniform norm bound and induction, and the second by strong convergence on the fixed vector $L^k x$. □

If estimated operators $\widehat{L}_n$ are uniformly bounded and $(\widehat{L}_n - L_n)x \to 0$ for every fixed $x$, the same argument gives fixed-horizon convergence for $\widehat{L}_n$. These are explicit assumptions, not conclusions of the finite learning procedure.

This result is strong convergence, not operator-norm convergence and not a bound uniform in time. On $\ell^2$, finite coordinate projections converge strongly to $I$ while $\|I - P_n\| = 1$ for every $n$. Unbounded operators require domain control. Adaptive data-dependent partitions and fitted coefficients require additional consistency arguments. These remain separate mathematical tasks.

## 13. A complete small signal example

Consider the rational linear signal dynamics

$$x_{t+1} = Lx_t, \qquad L = \begin{pmatrix} \frac{1}{2} & 1 \\ \frac{1}{4} & \frac{1}{3} \end{pmatrix}, \qquad x_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The spatial archive stores $x_t$. A second atlas uses the orthogonal Haar matrix

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Its coefficients $c_t = Ux_t$ reconstruct as $x_t = U^* c_t$. The same dynamics in that atlas are $ULU^*$, not a separate physical law. Thus coordinate transport preserves the full dynamics.

Now retain only the first spatial coordinate. With $P = \operatorname{diag}(1, 0)$, the retained scalar $z_t$ and discarded scalar $q_t$ obey

$$z_{t+1} = \tfrac{1}{2} z_t + q_t, \qquad q_{t+1} = \tfrac{1}{4} z_t + \tfrac{1}{3} q_t.$$

The exact first-coordinate sequence begins

$$z_0 = 1, \qquad z_1 = \tfrac{1}{2}, \qquad z_2 = \tfrac{1}{2}, \qquad z_3 = \tfrac{11}{24}.$$

The compressed one-step matrix predicts $z_2 = 1/4$ instead of $1/2$. The lost return path contributes exactly

$$BC = 1 \cdot \tfrac{1}{4} = \tfrac{1}{4}.$$

There are two exact repairs. One is state refinement: start with $C_0 = (1, 0)$ and add $C_0 L = (1/2, 1)$. The resulting row space is the full two-dimensional space, so it closes. The other is memory:

$$z_{t+1} = \tfrac{1}{2} z_t + \tfrac{1}{4} \sum_{k=0}^{t-1} \left(\tfrac{1}{3}\right)^k z_{t-1-k},$$

since $q_0 = 0$. This is the exact same system expressed without retaining $q_t$ as a separate current-state variable. At $t = 2$, it yields

$$z_3 = \tfrac{1}{2} \cdot \tfrac{1}{2} + \tfrac{1}{4} \cdot \tfrac{1}{2} + \tfrac{1}{12} \cdot 1 = \tfrac{11}{24}.$$

To see a nontrivial reduced closure, adjoin a third independent coordinate with multiplier $5/6$. The smallest closed refinement of the first coordinate then contains the first two coordinates, not the third. Exact prediction of the chosen summary does not require reconstructing every component of the enlarged signal.

Finally, low energy is not the same as low predictive importance. For

$$L' = \begin{pmatrix} 0 & 1000 \\ 0 & 0 \end{pmatrix}, \qquad x = (0, 1/1000)^\top,$$

omitting the second coordinate loses norm $1/1000$, but changes the next first coordinate by exactly one. The correct certificate is the product of omitted magnitude and transfer amplification.

## 14. Connection to observable learning

Exact realization, complete decomposition and operator transport provide the representation laws. Propositions 6–10 separate predictive closure, constructive refinement and memory. The remaining requirement for learning is access to a sufficient predictive state. Chapter 2 addresses a finite linear observable sector; it does not infer its rank or coverage assumptions from a successful fit.

CHAPTER 2

# Learning, stability and coherent expansion

The first stage constructed predictive closures for declared operators. The second stage asks how those constructions relate to what a learner can actually observe. A closure that depends on an unavailable hidden coordinate is a mathematical description, not yet an accessible state. A fitted recurrence can be accessible but wrong outside the span represented by its data. A one-step approximation can be accurate and still fail under repeated composition.

The results below connect these distinctions. They give an exact learning result for a finite linear observable sector, deterministic perturbation bounds, a propagation law for typed defects, a sufficient stability condition for truncated memory, controlled nonlinear interaction bounds, coherent atlas reconciliation and posterior accounting for an expanding grammar. These are finite or explicitly bounded results. They do not assume that every unknown process has a short linear state.

The observable-state perspective has classical antecedents in predictive state representations [R5] and observable-operator learning [R7]. Finite invariant observable spaces also underlie Koopman reductions [R6]. The present development uses elementary statements proved below and identifies their specific place in Komorebi's realization and obstruction calculus. It does not import those papers' sample-complexity guarantees into this model.

## 15. From hidden state to an accessible predictive state

Let a finite-dimensional autonomous linear process have hidden state $x_t \in \mathbb{R}^d$ and observed output $y_t \in \mathbb{R}^p$:

$$x_{t+1} = Lx_t, \qquad y_t = Cx_t.$$

The observation map may be noninjective. For a positive integer $h$, define

$$O_h = \begin{pmatrix} C \\ CL \\ \vdots \\ CL^{h-1} \end{pmatrix}, \qquad V_h = \operatorname{im} O_h.$$

The coordinates of $O_h x_s$ are future observations relative to time $s$. They are not available at time $s$ merely because this matrix is written down. At time $t \geq h-1$, however, the same vector is the observed history

$$z_t = \begin{pmatrix} y_{t-h+1} \\ \vdots \\ y_t \end{pmatrix} = O_h x_{t-h+1}.$$

Thus a forward observation map becomes a causal state after a declared warm-up interval. The time index is part of the contract.

**Proposition 14. Exact causal delay closure.** The following statements are equivalent:

1. There is a linear map $D_h : V_h \to V_h$ satisfying $O_h L = D_h O_h$.
2. $\ker O_h \subseteq \ker(O_h L)$.
3. $\operatorname{rank} O_{h+1} = \operatorname{rank} O_h$.

When these statements hold, $D_{\mathrm{h}}$ is unique on $V_{\mathrm{h}}$. The history state obeys $z_{\mathrm{t+1}} = D_{\mathrm{h}} z_{\mathrm{t}}$, and its last output block is $y_{\mathrm{t}}$. Its iterates determine every later observed output. If a positive integer $d$ bounds the hidden dimension, $h = d$ suffices for closure, though a smaller positive $h$ may suffice.

**Proof.** The equivalence of the first two statements is the linear factorization criterion: define $D_h(O_h x) = O_h L x$ and require independence from the representative $x$. The rows of $O_h L$ are $CL, \ldots, CL^h$. All except the final block already occur in $O_h$. The kernel condition is therefore equivalent to $CL^h$ being in the row space of $O_h$, which is exactly the rank equality. It also implies $O_h L x \in V_h$ because this vector is $O_h$ applied to $Lx$. Uniqueness follows because every element of $V_h$ has a representative. The history recurrence follows directly. Cayley-Hamilton expresses $L^d$ as a linear combination of lower powers, so $CL^d$ is in the row space of $O_d$. □

This does not recover hidden variables in $\ker O_h$. It identifies the quotient needed for the declared outputs. An independent hidden component can remain unobservable forever without preventing exact output prediction.

The rank equality here is a property of the *true observation maps*. A plateau in a finite empirical data matrix is not the same statement. It may reflect unexcited directions rather than a closed observation space. Section 17 makes the difference exact.

For several actions $L_u$, a fixed delay vector does not automatically have one autonomous transition matrix. Products of the actual action matrices determine its observation map, which can change with the action history. One needs an action-aware state, a common closed observable family, or a separately justified belief-state construction. Proposition 14 is not silently generalized to arbitrary controlled or stochastic dynamics.

## 16. What finite data identify

Suppose candidate states and their successors have been recorded as columns of matrices $Z_-, Z_+ \in \mathbb{R}^{m \times N}$. Their construction may use output histories. No hidden-state coordinates need be supplied to the fitting procedure.

Write $W = \operatorname{im} Z_-$. Exact interpolation of these columns is a question about a linear map on $W$, not yet about every physically realizable state.

**Proposition 15. Data-span identification and its boundary.** A linear map $B$ on the ambient space with $BZ_- = Z_+$ exists if and only if

$$\ker Z_- \subseteq \ker Z_+.$$

Its restriction to $W$ is unique. It is an endomorphism of $W$ if and only if $\operatorname{im} Z_+ \subseteq W$. A particular ambient extension is

$$B_0 = Z_+ Z_-^\dagger,$$

where † denotes the Moore-Penrose inverse. Every other fitted extension agrees with $B_0$ on $W$. If $W$ is a proper subspace of the true realizable state space $V$, the data alone do not identify the operator on $V \setminus W$.

**Proof.** Necessity follows by applying $B$ to every linear dependence among the input columns. Under the kernel condition, define $B(Z_- a) = Z_+ a$. Two choices of $a$ differ by a kernel vector and give the same answer. This defines and uniquely determines the map on $W$. It maps $W$ into $W$ exactly under the image condition. The pseudoinverse formula gives that map because $a - Z_-^\dagger Z_- a \in \ker Z_-$. If $W \subsetneq V$, choose a nonzero linear functional $\varphi$ on $V$ vanishing on $W$ and a nonzero $v \in V$. The operators $B$ and $B + v\varphi$ agree on every observed column and differ elsewhere. □

The three checks are different. The kernel test checks consistency of a deterministic linear fit. The image test checks whether the observed span is closed under the fitted transitions. Equality $W = V$ is a *coverage condition*. Neither of the first two checks establishes it.

For a known fixed feature space of dimension $r$, coverage is witnessed by $r$ independent input states. For an unknown larger feature space, the same rank only gives a lower bound on dimension. A model card must say which interpretation is being used.

If several action-specific operators act on a common declared $r$-dimensional state, each action's paired data must determine that operator on the required span. Pooling all actions into one high-rank matrix does not identify an action that has never been tried on some direction.

### 16.1 Perturbations have an explicit amplification factor

A noisy rank test is not an exact certificate. Even after the state space is fixed, fitting can be ill-conditioned. The following deterministic bound isolates that issue without introducing a probability model for noise.

Work in a fixed $r$-dimensional coordinate system, where $Z_- \in \mathbb{R}^{r \times N}$ has full row rank. Let

$$Z_+ = BZ_- + R, \qquad \widetilde{Z}_- = Z_- + E, \qquad \widetilde{Z}_+ = Z_+ + F.$$

The matrix $R$ records deviations from the declared linear recurrence on these samples. The matrices $E$ and $F$ record input and successor perturbations. All norms in the next proposition are spectral norms.

**Proposition 16. Fixed-space identification bound.** Put $s = \sigma_{\min}(Z_-) > 0$. If $\|E\| < s$, then $\widetilde{Z}_-$ has full row rank, and

$$\widehat{B} = \widetilde{Z}_+ \widetilde{Z}_-^{\dagger}$$

satisfies the exact error identity and bound

$$\widehat{B} - B = (R + F - BE)\widetilde{Z}_-^{\dagger},$$
$$\|\widehat{B} - B\| \le \frac{\|R\| + \|F\| + \|B\|\,\|E\|}{s - \|E\|}.$$

**Proof.** For every unit $v \in \mathbb{R}^r$, $\|\widetilde{Z}_-^{\top} v\| \ge \|Z_-^{\top} v\| - \|E^{\top} v\| \ge s - \|E\|$. Thus the perturbed matrix has full row rank and its pseudoinverse has norm at most $(s - \|E\|)^{-1}$. Since $\widetilde{Z}_- \widetilde{Z}_-^{\dagger} = I_r$, subtract $B$ from the fitted expression and collect terms. The bound follows by submultiplicativity and the triangle inequality. □

This is a bound for a fixed state space and a fixed coordinate basis. If the basis or its dimension is also estimated from noisy data, subspace-estimation error must be added. No such error is hidden inside $R$ without a derivation. Correlated and adversarial perturbations are allowed by the displayed deterministic inequality, but there is no sample-complexity claim until their magnitudes are bounded under a sampling model.

The denominator explains why numerical rank alone is insufficient. Full rank can coexist with severe noise amplification. The reference calculations include examples attaining the $1/s$ dependence and a rank-collapse example at the excluded boundary $\|E\| = s$.

## 17. Why finite agreement is not unrestricted discovery

**Proposition 17. Finite-prefix indistinguishability.** For every integer $T \ge 0$, there are two finite-dimensional autonomous rational linear systems with identical scalar observations at times $0, \dots, T$ and different observations at time $T + 1$. Both transition matrices can be chosen nilpotent.

**Proof.** One system has scalar output identically zero. For the other, take dimension $T + 2$, initial state $x_0 = e_{T+2}$, observation row $C = e_1^{\top}$, and shift $Le_j = e_{j-1}$ for $j \ge 2$, $Le_1 = 0$. Then $CL^t x_0 = 0$ for $0 \le t \le T$ and $CL^{T+1} x_0 = 1$. The shift is rational and nilpotent. □

An invisible delayed component of this kind can also be adjoined to any fixed base model. The combined output agrees with the base output on the entire inspected prefix and then differs. This is not an argument against learning. It identifies why a claimed guarantee must state a dimension bound, a model class, a coverage condition, an intervention protocol or a probabilistic assumption.

Even nilpotent operators have only zero eigenvalues, so stability language alone does not remove this ambiguity. A finite rank plateau cannot certify the absence of a delayed mode when arbitrarily large hidden dimensions are admitted.

## 18. A complete output-only identification example

Return to the rational system from Section 13, now with an additional independent hidden coordinate:

$$L = \begin{pmatrix} 1/2 & 1 & 0 \\ 1/4 & 1/3 & 0 \\ 0 & 0 & 5/6 \end{pmatrix}, \qquad C = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}.$$

The third coordinate never influences the observed output. The first coordinate alone is not closed, but

$$O_2 = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \end{pmatrix}$$

has rank two and $O_3$ has the same rank. Proposition 14 gives an exact two-dimensional delay state. Choose $x_0 = (1, 0, q)^\top$ for any $q$. The first four observations are

$$y_0 = 1, \quad y_1 = 1/2, \quad y_2 = 1/2, \quad y_3 = 11/24.$$

A learner receives those four outputs and the declared delay length $h = 2$. It does **not** receive $L$, $C$, $O_2$ or $q$. From consecutive observed windows it forms

$$Z_- = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}, \qquad Z_+ = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 11/24 \end{pmatrix}.$$

Since $\det Z_- = 1/4$, these windows span the declared two-dimensional state. Exact arithmetic gives

$$\boxed{D_2 = Z_+ Z_-^{-1} = \begin{pmatrix} 0 & 1 \\ 1/12 & 5/6 \end{pmatrix}.}$$

The learned recurrence is

$$\boxed{y_{t+1} = \frac{5}{6} y_t + \frac{1}{12} y_{t-1}, \qquad t \geq 1.}$$

It predicts $y_4 = 61/144$, which was not used to identify the coefficients. Proposition 14 and the verified rank condition on the true observation maps establish why this recurrence continues for every initial state of this particular system. The four samples alone establish the fitted recurrence only within the declared closed two-state model class. Proposition 17 forbids confusing those two conclusions.

The reference script separates the roles. One routine fits only the observed samples. A different check, with access to the generating system, verifies the theoretical assumptions and compares subsequent outputs. This prevents hidden access to the true operator from being presented as learning.

The same example connects back to the exact memory kernel. Section 13 expressed the output through a decaying convolution of its history. Here the history is summarized by a two-component recursive state. Neither representation stores the independent third coordinate. Memory and state realization describe the same observable behavior in two different ways.

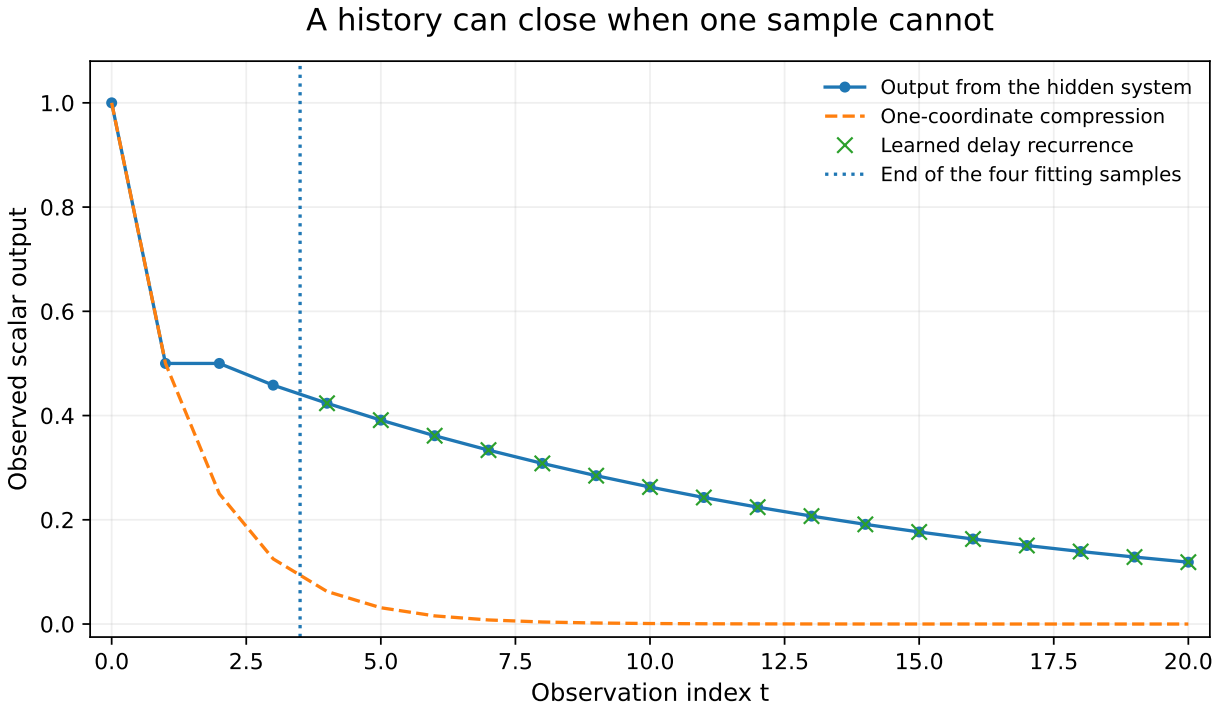



**Figure 2.1:** Output-only identification in the declared rational example. Four observations determine the delay recurrence. Later outputs are calculated independently from the hidden system. The one-coordinate compression omits the return path. This is a mathematical example, not a task-suite score.

## 19. A single propagation law for typed defects

A reconstruction error, an operator-fit error and a refinement error have different origins. Once each has been translated into an error in the state used by the next operation, their propagation can be analyzed by the same identity.

Allow the representation to change with time. Let $\gamma_t : X \to Z_t$ be the declared representation at step $t$, with a norm on each $Z_t$. Let $x_{t+1} = F_t(x_t)$ be the true transition and let $\widehat{f}_t : Z_t \to Z_{t+1}$ be the implemented step. Its defect is

$$r_t(x) = \gamma_{t+1}(F_t(x)) - \widehat{f}_t(\gamma_t(x)).$$

This definition already includes a representation switch. It is not restricted to one fixed feature vector.

**Proposition 18. Compositional defect propagation.** Suppose $\widehat{f}_t$ is $\lambda_t$-Lipschitz on a domain containing both $\gamma_t(x_t)$ and the predicted state $\widehat{z}_t$. Suppose $\|r_t(x_t)\| \le \delta_t$ and $\widehat{z}_{t+1} = \widehat{f}_t(\widehat{z}_t)$. Then, with $e_t = \|\gamma_t(x_t) - \widehat{z}_t\|$,

$$\boxed{e_T \le \left(\prod_{j=0}^{T-1} \lambda_j\right) e_0 + \sum_{k=0}^{T-1} \left(\prod_{j=k+1}^{T-1} \lambda_j\right) \delta_k.}$$

Empty products are one. If the output $h_t(x)$ has a decoder $H_t$ with Lipschitz constant $\kappa_t$ and $\|h_t(x_t) - H_t(\gamma_t(x_t))\| \le \eta_t$, then

$$\|h_t(x_t) - H_t(\widehat{z}_t)\| \le \eta_t + \kappa_t e_t.$$

**Proof.** Add and subtract $\widehat{f}_t(\gamma_t(x_t))$ to obtain $e_{t+1} \le \delta_t + \lambda_t e_t$. Repeated substitution yields the product-sum formula. Add and subtract $H_t(\gamma_t(x_t))$ for the output inequality. □

For a fixed linear representation $R$ and an estimated transition $\widehat{B}_u$, the defect is $(RL_u - \widehat{B}_u R)x$. For a reference reduced operator $B_u$, it splits exactly as

$$RL_u - \widehat{B}_u R = (RL_u - B_u R) + (B_u - \widehat{B}_u)R.$$

The first term is representational nonclosure. The second is operator-estimation error. A fit statistic for the second cannot certify that the first vanishes. Proposition 16 supplies a conditional bound on the second in its declared fixed-space setting.

If $\lambda_t \le \rho < 1$ and $\delta_t \le \delta$, then

$$e_T \le \rho^T e_0 + \delta \frac{1 - \rho^T}{1 - \rho}.$$

This is a genuinely recursive prediction bound. It is not merely a one-step comparison made using the true history at every step. If the Lipschitz constants exceed one, the same theorem remains valid but can permit error growth.

### 19.1 Stability under changing operations

For linear candidate operators, one sufficient way to obtain a common contraction norm is to exhibit a positive-definite matrix $M$ and $0 \le \rho < 1$ with

$$\widehat{B}_u^\top M \widehat{B}_u \preceq \rho^2 M \quad \text{for every declared action } u.$$

Then $\|z\|_M = (z^\top M z)^{1/2}$ gives $\|\widehat{B}_u z\|_M \le \rho \|z\|_M$ for all actions, directly from the displayed inequality. Proposition 18 therefore applies under arbitrary switching among them. This is the common quadratic Lyapunov construction from control theory [R8], used here as an explicit certificate rather than inferred from a list of eigenvalues.

Individual spectral stability does not suffice. The matrices

$$A = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix}$$

are both nilpotent, but $BA = \operatorname{diag}(0, 4)$. Alternating them amplifies an appropriate state by four every two steps. Even a single stable nonnormal matrix can produce large transient amplification in a specified norm.

If a learned policy chooses actions from predicted states while the reference process uses actions chosen from true states, the two action sequences may differ. Proposition 18 then needs a bound for that action mismatch, or a Lipschitz bound for the full closed-loop map. The result is not an automatic control-policy guarantee.

## 20. Truncating memory without losing the prediction bound

Section 8 bounded the omitted tail of the exact linear memory expression using the same resolved history. We now close the separate recursion argument.

Write the exact relation as

$$z_{t+1} = A z_t + \sum_{k=0}^{t-1} K_k z_{t-1-k} + f_t, \qquad K_k = B D^k C, \quad f_t = B D^t q_0.$$

For negative indices set $z_s = 0$, solely as a convention for missing prehistory in this initialized formula. A length-$H$ approximate model, with $H \ge 0$, is

$$\widehat{z}_{t+1} = \widehat{A} \widehat{z}_t + \sum_{k=0}^{H-1} \widehat{K}_k \widehat{z}_{t-1-k}.$$

An empty sum is zero. Let $d_t$ be the exact right-hand side minus this truncated right-hand side evaluated on the **true** history.

**Proposition 19. Uniform prediction bound for a stable finite-memory truncation.** Suppose

$$a = \|\widehat{A}\| + \sum_{k=0}^{H-1} \|\widehat{K}_k\| < 1, \qquad \sup_{t\ge 0} \|d_t\| \le \Delta.$$

Let $E_0$ bound the initialization errors at indices $-H, \dots, 0$. Then the recursively generated predictions satisfy

$$\boxed{\sup_{t\ge 0} \|z_t - \widehat{z}_t\| \le \max\left(E_0, \frac{\Delta}{1-a}\right).}$$

If $\|z_t\| \le Z$ for $t \ge 0$ and $\|D\| \le \rho < 1$, a valid choice of defect bound is

$$\Delta = Z\left(\|A - \widehat{A}\| + \sum_{k=0}^{H-1} \|K_k - \widehat{K}_k\|\right) + \|B\|\,\|C\| Z \frac{\rho^H}{1-\rho} + \|B\|\,\|q_0\|.$$

The last term can be removed if the exact initialized forcing $f_t$ is retained in the approximate recurrence.

**Proof.** Subtract the two recurrences and bound every retained history error by the largest of the errors at indices $t - H, \dots, t$. This gives $e_{t+1} \le a \max_{t-H\le s\le t} e_s + \Delta$. The ball of radius $E = \max(E_0, \Delta/(1-a))$ is invariant because $aE + \Delta \le E$. Induction proves the bound. The estimate for $\Delta$ separates coefficient errors, the geometric kernel tail and the initialized forcing, using Proposition 10. □

This norm-sum condition is sufficient, not necessary. A stable recurrence with cancellation can fail this test. Failure to find this certificate is not proof of instability.

For the scalar example with $q_0 = 0$, take exact retained coefficients and $z_0 = 1$. All outputs lie in $[0, 1]$: the full nonnegative memory coefficients sum to at most $7/8$, so induction bounds every new output by the previous maximum. Thus $Z = 1$ is valid. With

$$A = 1/2, \quad K_k = \frac{1}{4}\left(\frac{1}{3}\right)^k,$$

one obtains

$$a_H = \frac{7}{8} - \frac{3}{8}3^{-H}, \qquad \Delta_H = \frac{3}{8}3^{-H}, \qquad \boxed{\sup_t |z_t - \widehat{z}_t| \le \frac{3\,3^{-H}}{1 + 3\,3^{-H}}.}$$

The right-hand side is a theorem for every time in this initialized example. The figure compares it with a separately computed finite-horizon error. The computed maximum is not substituted for the uniform bound.

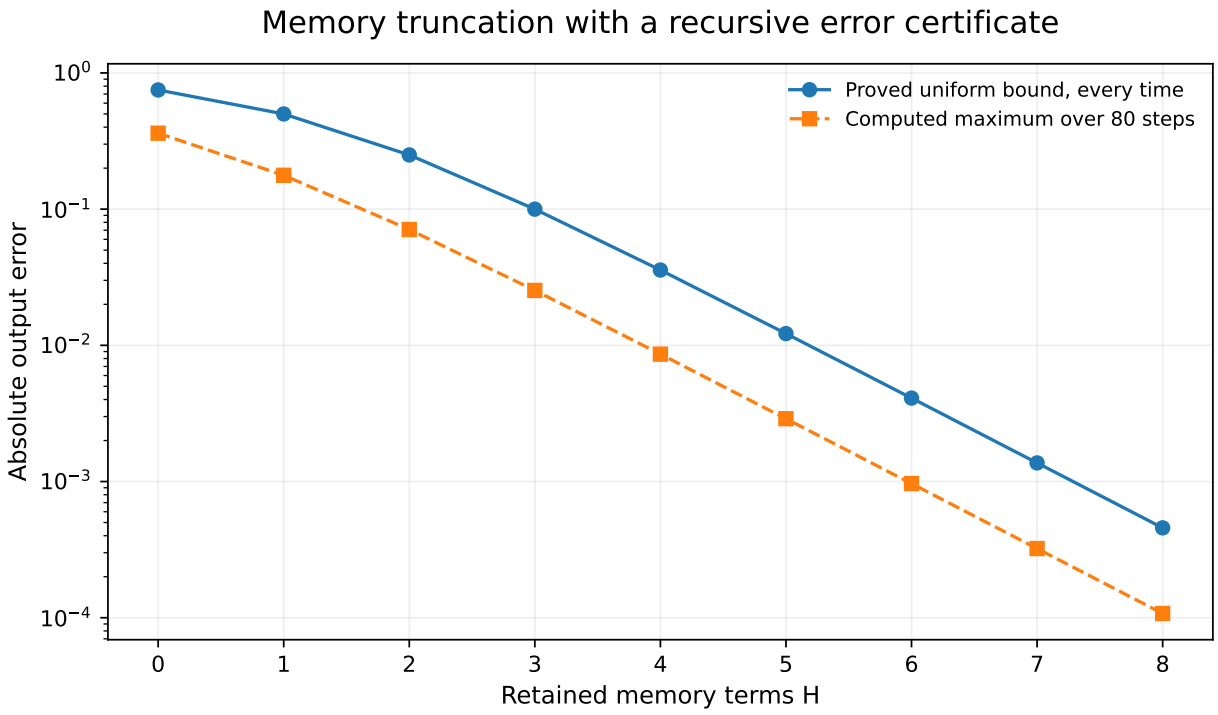



**Figure 2.2:** A memory-tail estimate becomes a recursive prediction certificate only after a stability margin is supplied. The measured curve is the maximum over 80 calculated steps. The upper curve is the proved bound for all steps under the stated scalar initialization.

## 21. Nonlinear packet truncation with an explicit remainder

Polynomial interactions were expanded exactly in Proposition 5. A computational model may retain only some ordered interaction tuples. It then needs a bound on what was omitted, rather than an assumption that packetwise evaluation is exact.

Let $P_i$ be an orthogonal resolution of a finite-dimensional real Hilbert space. Suppose

$$F(x) = b + \sum_{m=1}^{p} B_m(x, \dots, x),$$

with bounded multilinear maps. For each ordered tuple $\boldsymbol{i} = (i_1, \dots, i_m)$, let $\beta_{m,\boldsymbol{i}}$ bound the norm of $B_m$ restricted to $\operatorname{im} P_{i_1} \times \cdots \times \operatorname{im} P_{i_m}$. Let $\mathcal{I}_m$ be the retained tuple set, and let $F_{\mathcal{I}}$ retain the constant term and exactly those interactions.

**Proposition 20. Packet-interaction truncation bound.** Put $a_i = \|P_i x\|$. Then

$$\boxed{\|F(x) - F_{\mathcal{I}}(x)\| \le \sum_{m=1}^{p} \sum_{\boldsymbol{i} \notin \mathcal{I}_m} \beta_{m,\boldsymbol{i}} \prod_{j=1}^{m} a_{i_j}.}$$

On the ball $\|x\| \le R$, a valid Lipschitz bound for $F_{\mathcal{I}}$ is

$$\operatorname{Lip}(F_{\mathcal{I}}) \le \sum_{m=1}^{p} \sum_{\boldsymbol{i} \in \mathcal{I}_m} m \beta_{m,\boldsymbol{i}} R^{m-1}.$$

**Proof.** Subtract the retained terms from the exact multilinear expansion. Bound each omitted term by its restricted multilinear norm and sum. For the Lipschitz estimate, telescope the difference between two $m$-fold evaluations by replacing one argument at a time. Each changed argument has norm at most $\|x - y\|$ after projection, and each unchanged argument has norm at most $R$. There are $m$ terms for each retained tuple. □

If the declared polynomial itself approximates a nonpolynomial map with remainder at most $\varepsilon_R$ on the ball, add $\varepsilon_R$ to the defect bound. Without such a remainder estimate, the result is only a theorem for the polynomial, not for the unexpanded physical law.

For a retained state $z = Px$ and omitted state $q = (I - P)x$, dropping every term that uses $q$ gives the coarser majorant

$$\|F(z + q) - F(z)\| \le \sum_{m=1}^{p} \|B_m\| \left( (\|z\| + \|q\|)^m - \|z\|^m \right).$$

The tuple-level result is finer because it can distinguish a harmless omitted packet from a strongly coupled one. Ordered tuples must be counted with their multiplicities unless symmetry has explicitly combined them.

A prediction guarantee follows by inserting these defect and Lipschitz bounds into Proposition 18, provided the true and predicted trajectories remain in the declared domain. That domain condition cannot be inferred from a small initial error. One possible construction is to assume that the true map preserves a closed convex set $K$ and to project every approximate output back onto $K$. Euclidean projection is nonexpansive and fixes the true output, so it does not increase the one-step approximation bound. This is a declared modification of the predictor, not an identity for an unmodified recurrence.

For the scalar quadratic $(u + v)^2$, retaining only $u^2 + v^2$ omits $2uv$. The packet bound is $2|u||v|$ and is exact in magnitude. Small energy in one packet matters in proportion to both its amplitude and the cross-interaction coefficient.

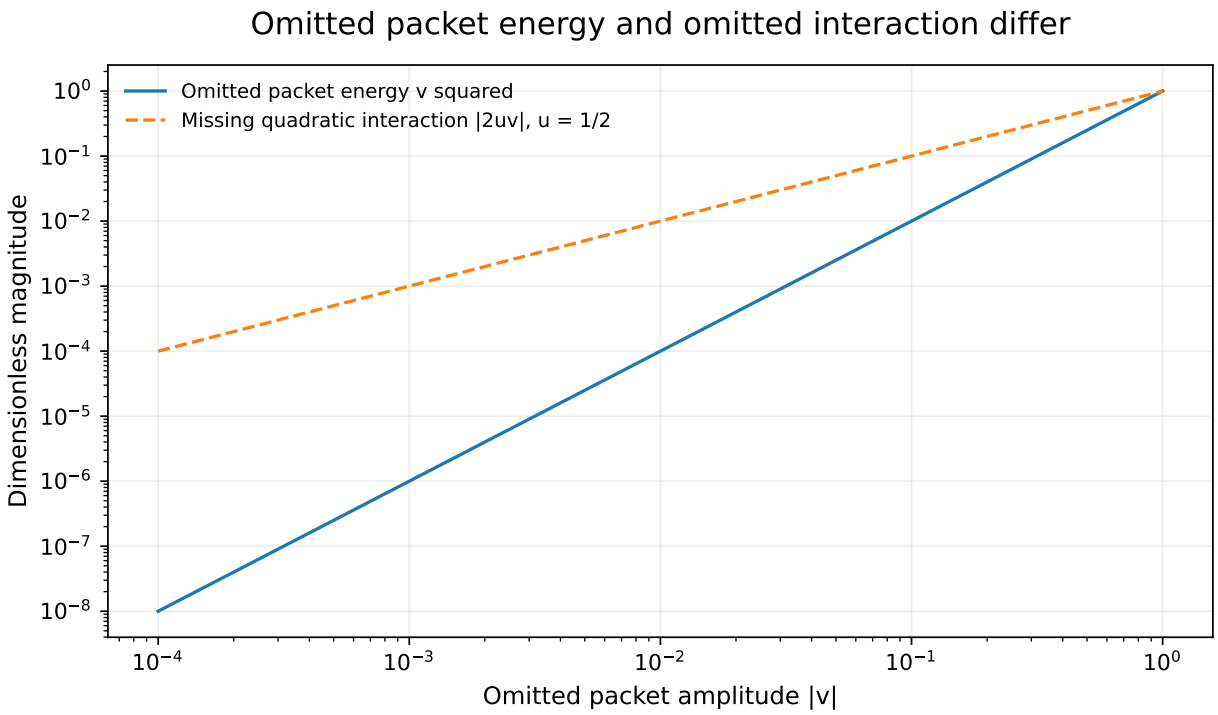



**Figure 2.3:** A nonlinear operation couples packets. For a square, the mixed term is proportional to the omitted amplitude while its energy is quadratic in that amplitude. The plot keeps those two scales visible. Refinement must account for transfer or interaction, not only reconstruction energy. A packet decomposition is not a collection of independent nonlinear subsystems. Dropping interaction terms requires its own approximation budget.

## 22. Learning several charts without inventing several worlds

The exact atlas identities constrain the meaning of a fitted model, but independently trained predictors need not satisfy them. Here is a simple coherent reconciliation with a precise objective.

Fix a finite-dimensional represented space $X$. It may be the observed signal space or a predictive quotient, but it must be the *same* space for all atlases in this section. Let $A_a : X \to C_a$ and $S_a : C_a \to X$ satisfy $S_a A_a = I$. Let $G_a : C_a \to C_a$ be linear fitted maps. Pull them back to object space:

$$L_a = S_a G_a A_a.$$

For positive weights summing to one, define

$$\overline{L} = \sum_a w_a L_a, \qquad \widetilde{G}_a = A_a \overline{L} S_a \quad \text{on } \mathcal{C}_a = \operatorname{im} A_a.$$

**Proposition 21. Coherent reconciliation and coordinate-stable transport.** The family $\widetilde{G}_a$ is consistent and satisfies every atlas intertwining identity. The operator $\overline{L}$ is the unique minimizer of

$$\sum_a w_a \|L - L_a\|_F^2$$

over linear maps on $X$, where $\|\cdot\|_F$ is the Frobenius norm. For any target operator $L_*$ and any chosen operator norm,

$$\|\overline{L} - L_*\| \le \sum_a w_a \|L_a - L_*\|.$$

Further, the norm $\|c\|_a = \|S_a c\|$ on $\mathcal{C}_a$ makes every exact transport $T_{ba} = A_b S_a$ an isometry. A represented transition $A_b L S_a$ has exactly the same induced Lipschitz constant in these norms as $L$ has in object space.

**Proof.** Coherence follows from Proposition 2 because every chart realizes the same $\overline{L}$. Expanding the squares gives

$$\sum_a w_a \|L - L_a\|_F^2 = \|L - \overline{L}\|_F^2 + \sum_a w_a \|L_a - \overline{L}\|_F^2.$$

The weighted cross term vanishes, proving the unique minimizer. The target bound follows from the triangle inequality. The pullback expression is a norm because $S_a$ is inverse to $A_a$ on $\mathcal{C}_a$. Finally, $\|T_{ba} c\|_b = \|S_b A_b S_a c\| = \|S_a c\|$. The same identity for a transition, and the bijection between $X$ and $\mathcal{C}_a$, establish equality of Lipschitz constants. □

This reconciliation is ordinary weighted least squares in a common semantic space. It does not prove that consensus is correct, reduce statistical error under arbitrary dependence, or turn deterministic recodings into independent evidence. The output must also be reanalyzed into a

realizable coefficient vector. An independently fitted redundant output can have an additional inconsistency residual $(I - A_a S_a) G_a A_a x$ before reconciliation.

For Euclidean coefficient errors, conditioning remains important. If $A_a$ has smallest singular value at least $m_a > 0$ and largest singular value at most $M_a$, its canonical synthesis $A_a^\dagger$ has norm at most $1/m_a$, and $\|T_{ba}\| \le M_b/m_a$. An arbitrary left inverse can behave much worse away from the consistent coefficient space, so the canonical-synthesis qualification matters.

A learned atlas can therefore be parameterized within prescribed singular-value bounds, with synthesis tied to its inverse or pseudoinverse. This gives a construction contract. It is not a claim that every unconstrained training procedure respects those bounds.

When atlases change over time but represent the same state, using the pullback norms avoids multiplying arbitrary chart condition numbers at every step. Exact transport telescopes to the direct transport. Numerical synthesis errors and genuinely inconsistent predictions remain defects and must still be bounded.

## 23. Refinement cannot manufacture missing information

An atlas change is an invertible recoding of the same represented information. Predictive refinement usually adds information. Treating these as identical operations would undo the central distinction of the foundation.

**Proposition 22. The upward-lift obstruction.** Let $\gamma : X \to Z$ be a coarse representation and $\eta : X \to W$ a proposed refined representation. There is a deterministic map $J : \gamma(X) \to \eta(X)$ satisfying $\eta = J\gamma$ if and only if

$$\gamma(x) = \gamma(x') \implies \eta(x) = \eta(x') \quad \text{for every } x, x'.$$

For linear representations, this is equivalent to $\ker \gamma \subseteq \ker \eta$. If two states share a coarse representation but their refined representations differ, any predicted refined value $w$ has worst-case error at least

$$\max\big(\|w - \eta(x)\|, \|w - \eta(x')\|\big) \ge \tfrac{1}{2}\|\eta(x) - \eta(x')\|.$$

**Proof.** A lift must be constant on each coarse fiber. Conversely, the stated condition defines it unambiguously on every fiber. The linear version is the kernel factorization criterion. The lower bound follows from the triangle inequality between the two refined values through $w$. □

For a strict refinement, the coarse state is recoverable from the refined one but not conversely. In the linear case this typically means $\ker \eta \subsetneq \ker \gamma$, exactly the opposite of the condition required for an upward lift.

For example, $\gamma(x_1, x_2) = x_1$ and $\eta(x_1, x_2) = (x_1, x_2)$. The states $(0, 1)$ and $(0, -1)$ have the same coarse state, but any common point initialization has worst-case error at least one in Euclidean norm. Padding a vector with zeros is a choice, not information recovery.

A refinement step therefore needs one of three explicitly distinguished inputs: an archive that already contains the relevant observations, new observations, or a set-valued or probabilistic state retaining uncertainty about the missing details. An exact archive of measured outputs can initialize a longer history when those samples were stored. It cannot recover a latent coordinate that never affected any measurement.

For adaptive prediction, an attempted lift $J_t$ contributes its own initialization defect

$$\eta_t(x_t) - J_t\gamma_t(x_t).$$

It belongs in Proposition 18's initial error or in the defect of the composed transition. It cannot be set to zero merely because a finer packet tree has been allocated.

A data-dependent representation schedule also requires care. The pathwise bound applies to the realized schedule when all of its hypotheses hold along that schedule. Comparing two different schedules chosen by true and predicted histories introduces a selection discrepancy. No uniform statistical consistency of adaptive selection follows solely from the deterministic propagation law.

## 24. A growing operator grammar with coherent evidence

A large algebraic signal model should be able to add operator programs without changing the meaning of its existing evidence. This requires more than assigning a new softmax weight when a candidate is introduced.

Let $\mathcal{G}$ be a countable declared program set, with one fixed probability mass function $\pi(g) > 0$ satisfying $\sum_g \pi(g) = 1$. Every program specifies a likelihood $L_g(D)$ for the **same** observed data event or density relative to a common base measure. Assume

$$0 < Z(D) = \sum_{g\in\mathcal{G}} \pi(g)L_g(D) < \infty.$$

Let finite active sets $\mathcal{G}_n$ increase to $\mathcal{G}$. Their unnormalized evidence is $Z_n = \sum_{g\in\mathcal{G}_n} \pi(g)L_g(D)$. Only stages with $Z_n > 0$ are considered.

**Proposition 23. Exact accounting for an enumerated grammar.** Define the full posterior $P(g \mid D) = \pi(g)L_g(D)/Z$ and extend the active posterior $P_n(g \mid D) = \pi(g)L_g(D)/Z_n$ by zero outside $\mathcal{G}_n$. Then

$$\boxed{\|P - P_n\|_{\mathrm{TV}} = \tau_n, \qquad \tau_n = 1 - \frac{Z_n}{Z}.}$$

For any common family of candidate predictive probability kernels, the corresponding predictive mixtures differ in total variation by at most $\tau_n$. If the omitted unnormalized evidence has a bound

$$\sum_{g \notin \mathcal{G}_n} \pi(g) L_g(D) \leq b_n,$$

then

$$\tau_n \leq \frac{b_n}{Z_n + b_n}.$$

**Proof.** On the active set, $P = (Z_n/Z)P_n$, so the total mass of the difference there is $\tau_n$. The mass outside is also $\tau_n$. Half the sum of absolute differences is therefore $\tau_n$. Applying the same probability kernel to two mixing measures cannot increase total variation: integrate any event probability, a function taking values in $[0, 1]$, against their difference. For the final bound, write $Z = Z_n + W_n$ with $0 \leq W_n \leq b_n$ and use that $W_n/(Z_n + W_n)$ increases with $W_n$. □

Since the active sets exhaust the grammar and the evidence sum is finite, $Z_n \uparrow Z$ and $\tau_n \downarrow 0$. This is posterior approximation *inside the declared grammar*. It does not establish that the real signal-generating mechanism belongs to it.

For a discrete data sequence, a probability likelihood is at most one, so the omitted prior mass is a valid $b_n$. A continuous-data density need not be at most one. It requires a density envelope or another justified bound on omitted evidence. A small unnormalized tail can also carry substantial posterior mass when the active evidence is very small.

The fixed prior need not be uniform. For an explicit enumeration $g_1, g_2, \ldots$, $\pi(g_k) = 2^{-k}$ is a normalized example, with omitted prior mass $2^{-n}$ after $n$ programs. It expresses a particular enumeration bias, not a canonical prior on mathematical meaning.

A program outside the enumerated prefix can be accounted for only when its prior and likelihood are defined or bounded. A genuinely outside-grammar component needs its own model. Assigning an unexplained residual probability is not a Bayesian treatment of missing mechanisms.

### 24.1 Refining a program without double counting it

A representation-rich language can contain many syntactic versions of one mechanism. It can also split one broad mechanism into more specific children. These operations must conserve prior mass when they are intended as a re-expression of the same model.

**Proposition 24. Prior-preserving program refinement.** Suppose a parent program $g$ is replaced by children $g_j$ with weights $\alpha_j \geq 0$, $\sum_j \alpha_j = 1$, and suppose its likelihood is the mixture

$$L_g(D) = \sum_j \alpha_j L_{g_j}(D).$$

Give child $j$ prior mass $\pi(g)\alpha_j$ and leave all other programs unchanged. Then total evidence is unchanged and the sum of the children's posterior masses equals the former parent's posterior mass. Posterior predictive distributions are also unchanged when the same mixture relation holds for the joint likelihood of $(D, y)$ for every predicted event $y$.

**Proof.** The children's total unnormalized evidence is $\sum_j \pi(g)\alpha_j L_{g_j}(D) = \pi(g)L_g(D)$. Thus both the total normalizer and every other program's posterior are unchanged. The same sum for joint evidence gives the predictive statement after normalization by the unchanged evidence of $D$. □

The mixture identity must hold for the **whole data record**. Averaging one-step likelihoods independently at each time generally describes a switching mechanism, not a fixed child selected once for the whole sequence.

If the children are only exact atlas recodings of the same operator and data model, their likelihoods are identical. Splitting the original prior mass leaves every prediction unchanged. Giving each duplicate a fresh full prior mass instead changes the model's prior preference. That may be an intentional new prior, but it is not invariance under representation.

The implementation's existing held-out predictive weights, described in Section 10, are not replaced by these identities merely by calling them posterior weights. The proposition supplies the probabilistic semantics that a future grammar-weighting implementation would need to enforce.

## 25. The finite learned-realization theorem

The preceding results combine into a concrete completion of one part of the learning problem. This is the integrated identification claim for the autonomous linear sector.

**Proposition 25. Observationally grounded finite linear realization.** Assume all of the following:

1. The data come from an autonomous finite-dimensional linear process with declared outputs $y_t = Cx_t$.
2. A delay length $h$ satisfies $\operatorname{rank} O_{h+1} = \operatorname{rank} O_h$.
3. Exact observed input windows in $Z_-$ span $V_h = \operatorname{im} O_h$, and $Z_+$ records their true one-step successor windows.

Then an algorithm using only $(Z_-, Z_+)$ recovers the unique transition on $V_h$. Initialized with a fully observed length-$h$ window, it predicts every later output exactly. Any lossless atlas of $V_h$, any complete packet decomposition of that space, and every coherently transported realization of the learned transition preserve the same predictions. Hidden variables outside the observable quotient need not be recovered.

If the fixed-space assumptions of Proposition 16 replace exact observations, that proposition bounds operator error. Whenever the resulting transition defects and Lipschitz constants satisfy Proposition 18, its product-sum bound controls the output prediction error. A common contraction margin yields the stated uniform bound.

**Proof.** Proposition 14 gives the unique true transition $D_h$ on $V_h$. Because the observed windows span $V_h$, Proposition 15 identifies exactly its restriction, which is the entire required operator. Equality of the learned and true states follows by induction after initialization. The current output is the last block of the history state. Propositions 1–4 preserve state reconstruction, transport

and complete decomposition, so they preserve that output sequence. The perturbation statement is the direct composition of Propositions 16 and 18 with the declared output decoder. □

This establishes an actual route from observations to a composable algebraic signal model under explicit finite assumptions. It is more than a fit on a few points. It is also less than unrestricted nonlinear world-model discovery. In particular, the algorithm does not infer assumptions 1–3 merely by returning a matrix. The accompanying certificate keeps the data-verified identities separate from those model assumptions.

## 26. Connection to controlled prediction

Proposition 25 completes the autonomous observable linear sector under its declared coverage assumptions. It does not establish a common delay operator for arbitrary actions. Chapter 3 constructs controlled predictive states from interventional experiment probabilities. The two results are compatible routes, not interchangeable learning assumptions.

CHAPTER 3

# Controlled stochastic realization

The next step is not to rename a deterministic predictor as a probabilistic one. A stochastic signal model must say what its state predicts, which actions intervene on the process, how observations update the state, and why every resulting number is a probability. These are separate obligations. A low-dimensional linear representation can satisfy its reconstruction equations while failing probability positivity.

Chapter 2 established observable learning for an autonomous linear signal. Here the observable objects are probabilities of finite action and observation experiments. We derive a finite realization from them, show when it is minimal, give an explicit probability-valid state domain, and separate exact reconstruction from noisy estimation. The final construction connects back to Komorebi's packets through a normalized closure-defect identity.

The classical antecedents are predictive state representations, observable operator models and finite-rank realization [R5, R7, R9]. Those traditions already use probabilities of tests and products of operators. The purpose here is to make their assumptions and consequences explicit within the realization calculus. The twelve propositions below are a connected development, not twelve claims of historical novelty. No statement here replaces a proof with a numerical check.

## 27. Actions and observations are different inputs

Let $\mathcal{A}$ and $\mathcal{O}$ be finite nonempty action and observation sets. A symbol $s = (a, o)$ records an action followed by an observation. A word $w = s_1 \cdots s_k$ has temporal order from left to right. The empty word is $\epsilon$.

Write

$$f(w) = \Pr(o_1, \ldots, o_k \mid \mathrm{do}(a_1, \ldots, a_k)).$$

The process is reset to one specified initial distribution before this experiment. The actions in this definition are fixed externally. This is not, in general, the distribution obtained by conditioning a passive record on an adaptively selected sequence of actions. We assume the consistency laws

$$f(\epsilon) = 1, \qquad f(w) \geq 0, \qquad \sum_{o \in \mathcal{O}} f(w(a, o)) = f(w) \quad \text{for every } w, a.$$

These laws specify causal probabilities on finite observation trees. A fixed history-dependent policy can be evaluated by selecting the action on each branch and multiplying the appropriate conditional probabilities. This does not establish optimality of that policy or its identification from insufficient data.

### 27.1 A finite hidden-state construction

A concrete realization has a belief $p \in \mathbb{R}^n_+$ with $\mathbf{1}^\top p = 1$. For each $s = (a, o)$, let $M_s \in \mathbb{R}^{n \times n}$ be entrywise nonnegative, with

$$\mathbf{1}^\top \sum_o M_{a,o} = \mathbf{1}^\top.$$

The matrix entry $M_{a,o}(j, i)$ is the probability of next hidden state $j$ and observation $o$, given current hidden state $i$ and chosen action $a$. This is a discrete probability instrument. It includes an action-specific transition followed by an observation, but does not require that factorization.

Put $M_w = M_{s_k} \cdots M_{s_1}$. Then

$$f(w) = \mathbf{1}^\top M_w p_0, \qquad p^+ = \frac{M_s p}{\mathbf{1}^\top M_s p}$$

on branches of positive probability. An observation with zero probability has no conditional posterior under that model. It must not be handled by silently dividing by an arbitrary numerical floor.

**Proposition 26. Finite instrument closure.** Starting with $W_0 = \operatorname{span}\{\mathbf{1}^\top\}$, define

$$W_{j+1} = W_j + \sum_{s \in \mathcal{A} \times \mathcal{O}} W_j M_s.$$

This stabilizes after at most $n - 1$ strict enlargements at

$$W_* = \operatorname{span}\{\mathbf{1}^\top M_w : w \text{ a finite word}\}.$$

Choose a full row rank matrix $R$ with row space $W_*$ and first row $\mathbf{1}^\top$. There are unique matrices $B_s$ with

$$RM_s = B_s R, \qquad \lambda = e_1^\top, \qquad \lambda R = \mathbf{1}^\top.$$

For $z = Rp$, both the next-observation likelihood and the posterior predictive state close:

$$\Pr(o \mid z, a) = \lambda B_{a,o} z, \qquad z^+ = \frac{B_{a,o} z}{\lambda B_{a,o} z}.$$

Every linear summary determining all word probabilities by linear readouts for every initial belief has row space containing $W_*$.

**Proof.** The nested-space argument from Proposition 8 applies to the finite family $M_s$. Multiplication of a test row by $M_s$ prepends the symbol $s$ to that test, so the stable span is exactly the displayed word span. Invariance supplies the matrices $B_s$, and full row rank makes them unique. Apply $R$ to the unnormalized and normalized hidden-state updates. If all test probabilities are linear functions of another summary, then each row $\mathbf{1}^\top M_w$ lies in its row space. Equality on all beliefs implies equality on their linear span, which is $\mathbb{R}^n$. □

This proposition uses a known hidden-state model. It is not yet an observable-data learner. It also concerns all initial beliefs, whereas a model initialized at one fixed $p_0$ can have a smaller reachable predictive dimension. The next construction handles that distinction without requiring recovery of hidden states.

## 28. The state is a function of possible future experiments

For a history $h$, define its unnormalized future-behavior column

$$v_h(t) = f(ht),$$

where $t$ ranges over every finite test word. Form the infinite history/test matrix

$$\mathscr{H}_{t,h} = f(ht).$$

Its rank means the dimension of the span of its columns, equivalently the supremum of the ranks of finite submatrices. The columns are functions, not vectors obtained by physically observing every future. Their finite coordinate realization becomes learnable only when enough experiment probabilities are available.

**Proposition 27. Minimal finite behavior realization.** If $\operatorname{rank} \mathscr{H} = r < \infty$, the vector space $V = \operatorname{span}\{v_h\}$ has dimension $r$. There are well-defined linear maps $B_s : V \to V$, an initial vector $v_\epsilon$, and a linear mass functional $\lambda(v) = v(\epsilon)$ satisfying

$$B_s v_h = v_{hs}, \qquad f(s_1 \cdots s_k) = \lambda B_{s_k} \cdots B_{s_1} v_\epsilon.$$

Every linear realization of $f$ has dimension at least $r$. Any two reachable and observable realizations of dimension $r$ differ by an invertible change of coordinates. For $f(h) > 0$, the normalized state $v_h / f(h)$ determines every future test probability conditional on $h$ and the declared future interventions.

**Proof.** If $\sum_h c_h v_h = 0$, then for every test $t$,

$$\sum_h c_h v_{hs}(t) = \sum_h c_h f(hst) = 0,$$

because $st$ is another test. Thus shifting columns preserves all linear relations and defines $B_s$. The probability formula follows by induction. For any $d$-dimensional realization, the history/test matrix factors into test rows and reachable state columns, so its rank is at most $d$.

For uniqueness, map the reachable vector after each history in one minimal realization to the corresponding vector in the other. If a linear combination vanishes in the first, every future test evaluates to zero on it. Those same evaluations vanish in the second, where observability makes the combination zero. The map is therefore well-defined and bijective, and intertwines every symbol operator and both boundary vectors. Finally $f(ht)/f(h)$ is exactly the conditional experiment probability. □

The coordinate dimension $r$ includes probability mass. Normalized states lie in the affine hyperplane $\lambda z = 1$, so their affine dimension is at most $r - 1$. A finite hidden-state model of size $n$ has behavior rank at most $n$. The theorem itself requires finite behavior rank, not a preferred hidden-state ontology.

Two realized histories with equal normalized future columns are equivalent for every declared future test. This is a predictive equivalence, not equality of their physical pasts or lossless archives.

## 29. Recovering the operators from an observable probability table

Select $r$ tests $t_1, \dots, t_r$ and $r$ histories $h_1, \dots, h_r$, with $t_1 = \epsilon$, for which the square matrix

$$H_{ij} = f(h_j t_i)$$

is invertible. Such a choice exists under the rank hypothesis of Proposition 27. The nonzero empty-test row can be extended to a row basis. Select independent history columns for that basis.

For each symbol $s$, also obtain the shifted table and initial test vector

$$(G_s)_{ij} = f(h_j s t_i), \qquad (b_0)_i = f(t_i).$$

**Proposition 28. Observable table reconstruction.** Assume the full behavior rank is at most $r$ and the displayed $r \times r$ matrix $H$ is invertible. Then the full rank is exactly $r$, and

$$\boxed{B_s = G_s H^{-1}, \qquad \lambda = e_1^\top, \qquad b_0 = (f(t_i))_i}$$

is a minimal realization of every word probability. Its state after history $h$ is

$$b(h) = (f(ht_i))_i = B_h b_0.$$

The normalized state is $b(h)/f(h)$ when $f(h) > 0$. Its update uses the formula in Proposition 26, with no hidden-state coordinates supplied to the reconstruction algorithm.

**Proof.** Invertibility gives rank at least $r$, so the assumed upper bound is attained. Evaluation on the selected tests is an isomorphism $V \to \mathbb{R}^r$, because it maps the chosen basis of history columns to the invertible matrix $H$. The shifted images of those basis columns are precisely the columns of $G_s$, so $B_s H = G_s$. The first selected test evaluates the mass. Initial evaluation gives $b_0$. Proposition 27 supplies correctness for every history, not only the selected columns. □

There are two ways to justify coverage in this statement. One can know that the selected row and column families span the true behavior space. Alternatively, a justified upper bound $r$ on the process rank, together with the invertible $r \times r$ block, proves that coverage. The observed invertibility by itself only gives a lower bound. An unknown higher-rank process can agree on a finite table and differ later.

Exact experiment probabilities are population quantities. The reference example below supplies rational tables from a declared process and keeps that table generator separate from the learner. Section 36 supplies one finite sampling protocol for estimating those entries. Neither a hidden matrix nor its probability cone is secretly passed to the observable learner.

## 30. Probability validity is not a rank condition

A real matrix product can yield a negative number. Therefore a fitted factorization is not a stochastic model merely because it is low rank or exactly interpolates a probability table. Spectral estimators can encounter negative-probability problems in practice [R10]. We address the structural issue directly.

In the canonical behavior space $V$, define

$$\mathcal{C} = \{v \in V : v(t) \geq 0 \text{ for every finite test } t\}.$$

**Proposition 29. Canonical probability cone.** The set $\mathcal{C}$ is a closed convex pointed cone that spans $V$. Every $B_s$ preserves it. The mass functional is strictly positive on its nonzero elements, and

$$\sum_o \lambda B_{a,o} = \lambda.$$

Consequently the normalized base

$$\mathcal{K} = \{v \in \mathcal{C} : \lambda v = 1\}$$

is a compact convex probability-valid state space. Every positive-probability branch maps $\mathcal{K}$ into itself by normalized updating. In the core-test coordinates of Proposition 28, $\mathcal{K} \subseteq \{1\} \times [0,1]^{r-1}$.

**Proof.** Each test evaluation is a continuous linear functional on finite-dimensional $V$, so the cone is closed and convex. If both $v$ and $-v$ belong to it, every test value is zero and $v = 0$. It spans $V$ because it contains every column $v_h$. Also $(B_s v)(t) = v(st)$, proving invariance.

The consistency law $\sum_o v(t(a,o)) = v(t)$ holds for each history column and hence throughout $V$. Iterating it along any fixed action schedule expresses $v(\epsilon)$ as a sum of nonnegative values for all observation words on that schedule. If $v(\epsilon) = 0$, each summand is zero, for every schedule and length, so $v = 0$. The same sums show $0 \le v(t) \le 1$ when $\lambda v = 1$. They also give mass conservation. The base is closed and bounded in the finite core coordinates and thus compact. Invariance and positive normalization prove the posterior assertion. □

This cone is defined by all finite tests. Its existence is not an algorithm for finding a finite list of inequalities. A finite list gives an outer approximation and can miss a longer-test violation. The cone also need not equal the convex hull of the states reachable from the fixed initial distribution.

A practical sufficient certificate can instead use a known closed cone $C_0$ with nonempty normalized base, a strictly positive mass functional, $b_0 \in C_0$, $B_s C_0 \subseteq C_0$, and mass conservation. A nonnegative hidden-state realization supplies one after a coordinate change. Such a certificate may use information unavailable to an observable-only learner and must be labeled accordingly.

Negative entries of $B_s$ do not, by themselves, prove invalidity. Coordinates can be signed while the represented state remains in the correct cone. Conversely, clipping negative coefficients can destroy the operator identities. Validity belongs to the pair consisting of the operators and their admissible cone, not to a coordinatewise heuristic.

## 31. Packets can reconstruct a belief without being beliefs

Let $P_1, \dots, P_m$ be a complete family of linear packet maps on a predictive space, with $\sum_i P_i = I$. Algebraic reconstruction gives $z = \sum_i P_i z$. It does not imply that $P_i z$ is an unnormalized probability state.

**Proposition 30. Positive packet decomposition and its boundary.** Suppose a cone $C$ has a strictly positive mass functional $\lambda$. If every packet map preserves $C$, then each normalized $z \in C$ decomposes as a genuine mixture

$$z = \sum_{i:w_i>0} w_i z_i, \qquad w_i = \lambda P_i z \ge 0, \qquad z_i = P_i z / w_i, \qquad \sum_i w_i = 1.$$

A zero weight has $P_i z = 0$. In the standard cone $\mathbb{R}^n_+$, however, a complete family of cone-preserving idempotents must consist of diagonal coordinate-selection matrices. In particular, a nontrivial Fourier or Haar packet family cannot generally be reinterpreted as a mixture of probability states merely because it is orthogonal and complete.

**Proof.** Cone preservation and strict positivity give the weights and the zero-weight claim. Apply $\lambda$ to the complete decomposition to obtain their sum. Normalization of each nonzero term gives the mixture.

On $\mathbb{R}^n_+$, cone preservation is entrywise nonnegativity. Since the matrices sum to $I$, every off-diagonal entry in every matrix is zero. Idempotence forces each diagonal entry to be zero or one, and completeness makes the selections disjoint and exhaustive. □

For a two-state belief $p = (3/4, 1/4)^\top$, the Haar detail projector

$$P_{\mathrm{d}} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

produces $(1/4, -1/4)^{\top}$. It is a valid signed detail packet, but not a subprobability distribution. Its total mass is zero despite being nonzero, which cannot happen inside a cone with a strictly positive mass functional.

This does not prohibit useful spectral packets. It fixes their type. Signed algebraic packets can be used for transfer, approximation and diagnostics. They must be synthesized before probability interpretation unless a separate positive decomposition has been established. That distinction prevents a reconstruction identity from being misused as a Bayesian mixture identity.

## 32. Action coverage is a mathematical requirement

**Proposition 31. Missing-intervention nonidentifiability.** There are finite controlled stochastic models that agree on every possible observation record generated using only action $a$, but disagree under another action $b$. Therefore even infinitely many passive records confined to $a$ cannot identify the unrestricted controlled model.

**Proof.** Use one hidden state and two observations. In both models, action $a$ gives probabilities $(1/2, 1/2)$. Under $b$, let the first model give $(1, 0)$ and the second $(0, 1)$. Every length-$k$ record under $a$ has probability $2^{-k}$ in both models. Their first observation under $b$ differs with probability one. □

This is a coverage obstruction, not an optimization problem. It remains even when the hidden dimension is known exactly. Richer transforms of the same action-$a$ observations do not create data about $b$.

Action selection based on past observations requires a further distinction. Under a known policy $\pi$, an action/observation word has joint probability

$$\Pr_{\pi}(w) = f(w) \prod_{j=1}^{|w|} \pi(a_j \mid h_{j-1}),$$

provided the action randomization has no access to hidden information beyond the declared history. The policy factors depend on the observed branch. Conditioning on a selected action sequence generally reweights branches rather than removing these factors. Recovering intervention probabilities from such data requires an appropriate coverage and reweighting argument. The reset experiments in Section 36 avoid that additional inference problem.

## 33. Learning the coordinate space under bounded noise

Chapter 2 bounded operator error after a state space and basis were fixed. Here we also bound one step of data-dependent space selection. This remains a finite-matrix statement. It does not certify that a chosen finite history/test table captures the full process.

Let $X \in \mathbb{R}^{m\times n}$ have rank $r \geq 1$, let $s = \sigma_r(X) > 0$, and observe $Y = X + E$ with $\|E\|_2 \leq \varepsilon < s/2$. Let $P$ project onto $\operatorname{im} X$, and let $\widehat{P}$ project onto the leading $r$ left singular vectors of $Y$.

**Proposition 32. Finite-table rank separation and subspace stability.** One has

$$\sigma_r(Y) \geq s - \varepsilon, \qquad \sigma_{r+1}(Y) \leq \varepsilon, \qquad \|P - \widehat{P}\|_2 \leq \frac{2\varepsilon}{s} < 1.$$

The $(r+1)$st singular value is interpreted as zero when $r$ is the maximum possible rank. If $U$ and $\widehat{U}$ are orthonormal bases of the two ranges, then

$$\sigma_{\min}(\widehat{U}^\top U) \geq \sqrt{1 - (2\varepsilon/s)^2}.$$

Thus $x \mapsto \widehat{U}^\top x$ is injective on the true column space. Its inverse on that space is

$$y \longmapsto U(\widehat{U}^\top U)^{-1} y,$$

with norm at most $[1 - (2\varepsilon/s)^2]^{-1/2}$.

**Proof.** The min-max characterization of singular values and the triangle inequality imply that adding $E$ changes each singular value by at most $\varepsilon$. The leading-$r$ singular subspace gives $\|(I - \widehat{P})Y\|_2 = \sigma_{r+1}(Y) \leq \varepsilon$. Hence

$$\|(I - \widehat{P})X\|_2 \leq 2\varepsilon, \qquad \|(I - \widehat{P})P\|_2 = \|(I - \widehat{P})XX^\dagger\|_2 \leq 2\varepsilon/s.$$

For equal-rank orthogonal projections, the latter norm equals $\|P - \widehat{P}\|_2$. To see the equality, diagonalize $U^\top \widehat{U}$ by singular value decomposition. Each nontrivial paired principal plane contributes a sine to both norms, while common directions contribute zero. On the same planes the singular values of $\widehat{U}^\top U$ are the corresponding cosines. This proves the last bound and the inverse formula. □

A threshold $\tau$ with $\varepsilon < \tau < s - \varepsilon$ recovers the rank of this finite matrix. The unknown true gap must be justified or replaced by a valid confidence interval. Rank thresholds do not exclude modes absent from the selected tests or unexcited by the selected histories.

The inverse formula contains the unknown true basis $U$. It proves identifiability and conditioning relative to that space, not that a learner already knows a perfect decoder. When using the estimated subspace as coordinates for a realization, all boundary vectors, symbol operators and probability constraints must be transformed together. Signed singular-vector coordinates are not new independent observations.

## 34. Noisy operator reconstruction in a fixed core

After choosing a core block, the square reconstruction formula yields a particularly direct error identity. It is useful even when the core was chosen on a separate discovery data set.

Let $G_s = B_s H$ with $s_H = \sigma_{\min}(H) > 0$. Observe

$$\widetilde{H} = H + E, \qquad \widetilde{G}_s = G_s + F_s, \qquad \|E\|_2 < s_H.$$

**Proposition 33. Noisy core-block realization.** The estimate $\widehat{B}_s = \widetilde{G}_s \widetilde{H}^{-1}$ satisfies

$$\widehat{B}_s - B_s = (F_s - B_s E)\widetilde{H}^{-1},$$

and therefore

$$\boxed{\|\widehat{B}_s - B_s\|_2 \le \frac{\|F_s\|_2 + \|B_s\|_2 \|E\|_2}{s_H - \|E\|_2}.}$$

A bound using only the fitted operator and the same perturbation budgets is

$$\|\widehat{B}_s - B_s\|_2 \le \frac{\|F_s\|_2 + \|\widehat{B}_s\|_2 \|E\|_2}{s_H}.$$

If $\widetilde{s} = \sigma_{\min}(\widetilde{H})$ is known and $\|E\|_2 \le e < \widetilde{s}$, the last denominator may be replaced by $\widetilde{s} - e$.

**Proof.** Since $\widetilde{H}$ is invertible, subtract $B_s$ from its fitted expression and collect the terms. The least singular value of $\widetilde{H}$ is at least $s_H - \|E\|_2$. For the second identity, use

$$(\widehat{B}_s - B_s)H = F_s - \widehat{B}_s E$$

and multiply by $H^{-1}$. Finally $s_H \ge \widetilde{s} - e$ by singular-value perturbation. □

When the first test is empty, the mass row $\lambda = e_1^\top$ is fixed rather than estimated. Set the first component of the initial vector to one as required by $f(\epsilon) = 1$. Errors in other initial test probabilities remain initial-state error. This convention removes an avoidable normalizer-estimation ambiguity but does not enforce cone preservation or mass conservation for the fitted symbol family.

Numerical implementation should solve the linear systems rather than explicitly form a floating-point inverse. The theorem describes the exact algebraic estimator. Finite-precision solve error is an additional residual, which the reference script measures separately.

## 35. Conditioning a prediction can amplify its error

Let $C$ be a probability cone with compact normalized base $K$ and mass row $\lambda$. Work in a declared norm and its dual. Suppose true normalized states have norm at most $R$. Let a true branch have

$$u = B_s z, \qquad p = \lambda u \geq \eta > 0, \qquad z^+ = u/p \in K.$$

For a predicted state $\widehat{z}$, form $\widehat{u} = \widehat{B}_s \widehat{z}$, $\widehat{p} = \lambda \widehat{u}$, and define $d = \widehat{u} - u$. Write $e = \|\widehat{z} - z\|$, $\|\widehat{B}_s\| \leq L$, and $\|\widehat{B}_s - B_s\| \leq \epsilon_B$. Then

$$\|d\| \leq Le + \epsilon_B R =: D.$$

**Proposition 34. Normalization-aware filtering stability.** If $\|\lambda\|_* D < \eta$, then $\widehat{p} > 0$ and

$$\widehat{z}^+ - z^+ = \frac{(I - z^+\lambda)d}{\widehat{p}},$$

so

$$\boxed{\|\widehat{z}^+ - z^+\| \leq \frac{(1 + R\|\lambda\|_*)D}{\eta - \|\lambda\|_* D}.}$$

Alternatively, if each realized fitted denominator is checked to satisfy $\widehat{p}_t \geq \widehat{\eta}_t > 0$, the pathwise recursion obeys

$$e_{t+1} \leq \alpha_t e_t + \beta_t, \qquad \alpha_t = \frac{(1 + R\|\lambda\|_*)L_t}{\widehat{\eta}_t}, \qquad \beta_t = \frac{(1 + R\|\lambda\|_*)\epsilon_{B,t} R}{\widehat{\eta}_t}.$$

The product-sum bound of Proposition 18 applies. A uniform bound follows if these constants have a common contraction margin and bounded forcing. The algebraic bound alone does not certify that a fitted state or every unchosen branch is probability-valid.

**Proof.** The branch likelihood differs by $\lambda d$, giving the lower bound on $\widehat{p}$. Subtract $u/p$ from $(u + d)/(p + \lambda d)$ and use $u = pz^+$. The induced norm of $I - z^+\lambda$ is at most $1 + R\|\lambda\|_*$. Applying that identity at each step gives the recursion. □

Rare evidence can create a large posterior change from a small absolute error. For example, take a true unnormalized vector $(\eta, 0)$ and a perturbed vector $(\eta, \eta)$. Their $\ell^1$ difference is $\eta$, while their normalized vectors differ by one in $\ell^1$. Both arise as possible branch outputs of valid probability instruments when $0 < \eta \leq 1/2$, after assigning the remaining mass to another branch. There is no uniform posterior Lipschitz constant independent of a probability margin.

When a known closed convex feasible base is available, Euclidean projection of a tentative posterior back onto that base cannot increase its distance from the true posterior. This is a declared repair to the estimator, not an identity for an unmodified update. It also does not repair negative branch probabilities of arbitrary fitted operators before normalization.

## 36. A finite experimental route to the probability table

The following deliberately simple protocol makes a finite-sample statement possible without confusing a deterministic identity with a statistical guarantee.

Fix positive integers $M$ and $N$ and choose $M$ finite words before estimation. For each word, reset to the same initial distribution, execute its action sequence, and record whether all observations match. Obtain $N$ independent trials per word. Estimates for different words may be dependent, but each word's $N$ trials must be independent with the declared success probability.

**Proposition 35. Reset-experiment probability certificate.** Let $f(w)$ be the sample mean for each selected word. For $0 < \delta < 1$, with probability at least $1 - \delta$, all $M$ estimates satisfy

$$|\widehat{f}(w) - f(w)| \le \epsilon_N, \qquad \epsilon_N = \sqrt{\frac{\log(2M/\delta)}{2N}}.$$

If the selected words cover every entry of $H$, every $G_s$, and $b_0$, then on the same event

$$\|\widetilde{H} - H\|_2 \le r\epsilon_N, \qquad \|\widetilde{G}_s - G_s\|_2 \le r\epsilon_N, \qquad \|\widetilde{b}_0 - b_0\|_2 \le \sqrt{r}\epsilon_N.$$

These bounds feed Propositions 32 and 33 whenever their respective gap conditions hold. The empty-word probability may be set exactly to one instead of sampled.

**Proof.** For a centered Bernoulli variable $X - p$, write $g(t) = \log \mathbb{E} \exp(t(X - p))$. Then $g(0) = g'(0) = 0$, and $g''(t)$ is the variance of a Bernoulli under exponential tilting, at most $1/4$. Integrating twice gives $g(t) \le t^2/8$. Independence and the exponential Markov inequality give an upper-tail bound $\exp(-2N\epsilon^2)$ after optimizing $t = 4\epsilon$. The lower tail has the same bound. A union bound over the $M$ words proves the simultaneous result. The spectral norm is bounded by the Frobenius norm, yielding the matrix and vector estimates. □

This is the Bernoulli form of Hoeffding's inequality [R11], proved here for the protocol used. It is a conservative entrywise-to-spectral bound, not a claim of optimal sample complexity. The number and lengths of chosen experiments are part of its cost.

If core tests and histories were selected from data, use a separate independent discovery sample and freeze the selected finite experiment list before these estimation trials. Conditional on that selection, the proposition applies. Alternatively, a uniform bound over a fixed larger candidate list can support data-dependent selection within that list. Reusing a post-selected table without either argument changes the claim.

A single dependent trajectory, an adaptive stopping rule, unknown resets, drifting sensors or experiments chosen without coverage do not satisfy this proposition as stated. Those settings require their own concentration and identifiability arguments. Synthetic reset sampling in the reference example demonstrates the protocol, not its validity for every measurement source.

## 37. A filtering defect with a packet interpretation

The relation to the realization calculus is particularly direct before normalization. Let $R$ be a linear summary of a hidden probability space, let $\lambda R = \mathbf{1}^\top$, and let $B_s$ be a proposed reduced branch operator. For a true belief $p$, put

$$z = Rp, \qquad \delta_s = (RM_s - B_sR)p.$$

The defect $\delta_s$ is a vector in the reduced representation, not a probability of model failure. It records the failure of the unnormalized instrument to commute with the summary.

**Proposition 36. Exact normalized closure defect.** Suppose the true and proposed branch likelihoods

$$a = \mathbf{1}^\top M_s p > 0, \qquad b = \lambda B_s z > 0$$

are positive, and let $\widehat{z}^+ = B_s z/b$. Then

$$\boxed{R\left(\frac{M_s p}{a}\right) - \widehat{z}^+ = \frac{(I - \widehat{z}^+\lambda)\delta_s}{a}, \qquad a - b = \lambda\delta_s.}$$

If $M_s \neq 0$ and both likelihoods and normalized updates agree on every strictly positive belief, then $RM_s = B_sR$. Agreement of normalized updates alone need not force this equality.

**Proof.** The true unnormalized summary is $B_sz + \delta_s$, with mass $a = b + \lambda\delta_s$. Subtract the two normalized vectors and collect terms. Since a nonzero nonnegative branch has positive likelihood on every strictly positive belief, agreement permits multiplication by the common mass. This gives equality of unnormalized summaries on the interior of the probability simplex, whose linear span is the whole hidden vector space. For the last claim, a one-state model can assign different positive branch probabilities while its normalized state always remains the scalar one. □

For an orthogonal retained projector $P$ preserving the mass row, put $Q = I - P$, $R = P$ on its image and $B_s = PM_sP$. Then

$$\delta_s = PM_sQp.$$

This is the unresolved-to-retained transfer block from Chapter 1. Its posterior effect is not simply its norm: normalization applies the oblique map $I - \widehat{z}^+\lambda$ and divides by the true branch probability. A component proportional to $\widehat{z}^+$ changes only the branch mass and disappears from the normalized defect, while still changing likelihood-based inference over mechanisms.

A certificate should therefore retain both $\lambda\delta_s$ and the posterior defect. A posterior that looks correct can hide a wrong evidence model. This is the stochastic counterpart of the distinction between reconstruction and predictive closure.

## 38. A complete two-action rational example

Use two hidden states only to generate a reference process. The learner does not receive these matrices. Let

$$T_0 = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}, \qquad T_1 = \begin{pmatrix} 1/4 & 2/3 \\ 3/4 & 1/3 \end{pmatrix}, \qquad p_0 = \begin{pmatrix} 2/3 \\ 1/3 \end{pmatrix}.$$

The observation matrices are

$$O_0 = \operatorname{diag}(1/4, 3/4), \qquad O_1 = \operatorname{diag}(3/4, 1/4), \qquad M_{a,o} = O_o T_a.$$

The two transitions do not commute. Thus the process is genuinely action-dependent rather than one autonomous recurrence with an action label attached.

Choose tests $(\epsilon, (0,1))$ and histories $(\epsilon, (0,1))$. The observable tables give

$$H = \begin{pmatrix} 1 & 13/24 \\ 13/24 & 5/16 \end{pmatrix}, \qquad \det H = 11/576, \qquad b_0 = \begin{pmatrix} 1 \\ 13/24 \end{pmatrix}.$$

The selected coordinate $q$ is the probability of seeing observation one after choosing action zero. In the hidden realization,

$$R = \begin{pmatrix} 1 & 1 \\ 5/8 & 3/8 \end{pmatrix}, \qquad z = Rp = (1, q)^\top.$$

The observable reconstruction $B_s = G_s H^{-1}$ yields

$$B_{0,0} = \begin{pmatrix} 1 & -1 \\ 11/32 & -1/4 \end{pmatrix}, \quad B_{0,1} = \begin{pmatrix} 0 & 1 \\ -3/32 & 3/4 \end{pmatrix},$$
$$B_{1,0} = \begin{pmatrix} 5/48 & 5/6 \\ 23/192 & 5/24 \end{pmatrix}, \quad B_{1,1} = \begin{pmatrix} 43/48 & -5/6 \\ 37/64 & -5/8 \end{pmatrix}.$$

The coefficients include negatives. Nevertheless, every operator preserves the cone $R\mathbb{R}^2_+$, because $B_s R = R M_s$. A sufficient normalized base is

$$K_0 = \{(1, q)^\top : 3/8 \le q \le 5/8\}.$$

This cone certificate uses the known generating model and is checked separately. It is not inferred merely from the finite probability table, and $K_0$ is not claimed to be the maximal canonical base of Proposition 29.

The likelihoods and posterior maps can now be read from the matrices. For action zero and observation one,

$$p(1 \mid z, 0) = q, \qquad q^+ = \frac{-3/32 + (3/4)q}{q} = \frac{3}{4} - \frac{3}{32q}.$$

Starting at $q_0 = 13/24$, this gives probability $13/24$ and $q_1 = 15/26$. If action one then produces observation zero, its probability is

$$\frac{5}{48} + \frac{5}{6}\frac{15}{26} = \frac{365}{624},$$

and the new predictive coordinate is

$$q_2 = \frac{23/192 + (5/24)(15/26)}{365/624} = \frac{599}{1460}.$$

The joint probability of this two-step action/observation word is $365/1152$. The reference checks compare these values and every word up to a declared finite test length against the hidden generator in exact rational arithmetic. The proposition, under its rank and coverage assumptions, is what extends the result to all word lengths.

The model's online state is not a sequence of newly measured future probabilities. It is a recursively maintained prediction initialized from learned experiment probabilities and updated after each actual action and observation. Uncertainty in those learned parameters is separate from the stochastic uncertainty in the next observation.

## 39. Integrated controlled stochastic realization

**Proposition 37. Conditional observation-based stochastic completion.** Suppose a causal finite-alphabet signal process has behavior rank at most $r$, and a selected $r \times r$ history/test block containing the empty test is invertible. Exact values of that block, every one-symbol shifted block, and the initial core-test vector determine a minimal realization of every future experiment probability. Its normalized state is recursively accessible after each positive-probability history, without reconstructing the hidden physical state.

The exact realization has the canonical probability cone of Proposition 29. Under any invertible atlas, transform the state, symbol operators, mass row and cone together:

$$z' = Tz, \qquad B'_s = TB_sT^{-1}, \qquad \lambda' = \lambda T^{-1}, \qquad C' = TC.$$

Every likelihood and posterior prediction is then unchanged. A complete packet decomposition preserves these predictions after full synthesis. Packetwise probability interpretation requires the extra positivity condition of Proposition 30.

For estimates obtained under Proposition 35's reset protocol, the simultaneous event there bounds the finite table perturbations. A separated finite-table spectrum yields Proposition 32's subspace certificate. A well-conditioned fixed core yields Proposition 33's operator certificate. On branches satisfying Proposition 34's likelihood margins, the resulting normalized state errors

obey its recursive bound. These perturbation statements alone do not certify a learned cone or global positivity of an estimated model.

**Proof.** Proposition 28 reconstructs the unique minimal behavior realization, and Proposition 27 gives its conditional interpretation. Proposition 29 supplies probability validity for the exact realization. Substituting the transformed quantities cancels every $T^{-1}T$, including in the scalar normalization. Full packet synthesis is the identity, whereas Proposition 30 characterizes the additional probability-mixture condition. The perturbation conclusions are the stated chain of propositions with their separate assumptions retained. No missing gap, coverage or positivity assumption follows from matrix inversion alone. □

This theorem closes a finite controlled and stochastic sector of the foundational model. It does not claim unrestricted discovery of a nonlinear environment from a single trajectory. It does identify what a learned predictive state represents, why its operations compose, where its evidence comes from, and which defects change likelihoods or posteriors.

## 40. From realization to admissible domains

Proposition 37 identifies a finite controlled realization from a covered population table. Existence of its canonical probability cone does not provide a finite certificate for a fitted model. Chapter 4 supplies such certificates in finite positive-lift sectors, while keeping internal validity distinct from statistical coverage.

CHAPTER 4

# Certified probability domains and robust learning

A matrix learned from observations is not yet a probability model. Chapter 3 made that distinction exact: a finite predictive realization needs an admissible cone preserved by its action and observation operators. It supplied a canonical cone defined by all possible future tests, but did not supply a finite way to verify that cone for an estimated model.

This chapter develops a constructive finite sector of that missing step. It gives a certificate that can be checked by rational arithmetic, a search procedure that does not need hidden-state matrices, and a robust version that covers an entire declared uncertainty set. It also proves why a finite polyhedral certificate is not available for every valid finite-dimensional process. The resulting distinction is between an exact finite certificate, an uncertified candidate, and an explicitly modified approximation. Those are different mathematical outcomes.

The positive-realization and invariant-cone ingredients are classical. Jaeger's account includes the polyhedral-cone characterization and an irrational-rotation obstruction [R12]. Polyhedral invariance is also a standard subject of linear systems and optimization [R13]. The present contribution is a self-contained connection of these ingredients to Komorebi's observable learner, probability packets, confidence sets and approximation defects. The controlled rational rotation example is an adaptation of that classical obstruction, not a claim of discovering the phenomenon.

The next twelve propositions continue the numbering of the earlier chapters. All have written proofs below. The exact computations check finite instances and counterexamples. The formal coverage of each proposition is recorded separately in the appendices.

## 41. The object to be certified

Fix a real predictive space $E = \mathbb{R}^r$, a nonzero row $\lambda$, and a finite family $B_s$, indexed by $s = (a, o)$. The action and observation alphabets are finite and nonempty. Every declared observation branch is present, including branches represented by the zero operator. Temporal words act from right to left:

$$B_{s_1 \cdots s_k} = B_{s_k} \cdots B_{s_1}, \qquad B_\epsilon = I.$$

The mass identity is an exact constraint,

$$\lambda \sum_o B_{a,o} = \lambda \quad \text{for every action } a.$$

An initial state has $\lambda z_0 = 1$. The proposed probability of a word is $\lambda B_w z_0$. An invertible coordinate change must transform $z_0$, $B_s$, $\lambda$ and the admissible cone together, as in Proposition 37.

A **probability-domain certificate** specifies a closed, pointed, generating cone $C \subset E$, with $\lambda$ strictly positive on $C \setminus \{0\}$, such that

$$z_0 \in C, \qquad B_s C \subseteq C.$$

Its normalized base is $K = \{z \in C : \lambda z = 1\}$. These conditions imply nonnegative word probabilities and their per-action normalization at every depth. Positive-probability updates remain in $K$. Zero-probability branches have no conditional posterior and must not be assigned one by denominator flooring.

There are three separate claims. **Internal validity** says the fitted model is a stochastic model. **Data compatibility** says it agrees with the observations within a specified error budget. **Coverage of the generating process** says the assumptions and uncertainty set actually include the process that generated those observations. The first can have a deterministic certificate even when the other two remain conditional.

## 42. Finite tests give an outer hierarchy, not automatic closure

Define the test cones

$$C_h = \{z : \lambda B_w z \geq 0 \text{ for every } |w| \leq h\}, \qquad C_0 = \{z : \lambda z \geq 0\}.$$

Each $C_h$ is polyhedral because it is an intersection of finitely many homogeneous half-spaces. It need not be a pointed cone or have a bounded normalized base.

**Proposition 38. Test-cone hierarchy and a finite stopping certificate.** The cones satisfy

$$C_{h+1} = C_h \cap \bigcap_s B_s^{-1} C_h, \qquad C_{h+1} \subseteq C_h.$$

Their intersection $C_\infty$ is the largest cone contained in $C_0$ and invariant under all $B_s$. If an exact equality $C_{h+1} = C_h$ occurs, then $C_h = C_\infty$ and the finite inequalities certify all future word probabilities. If the representation is observable, meaning that the rows $\lambda B_w$ span $E^*$, then $C_\infty$ is pointed and $\lambda$ is strictly positive on its nonzero elements. If reachable vectors $B_w z_0 \in C_\infty$ span $E$, the cone is also generating.

**Proof.** For a first symbol $s$ followed by a word $v$, $B_{sv} = B_v B_s$. The inequalities for all words of length at most $h+1$ are therefore exactly the displayed recurrence. If $z \in C_\infty$, every test on $B_s z$ is another test on $z$, so $C_\infty$ is invariant. Conversely, invariance of a cone contained in $C_0$ makes

every successive test nonnegative, hence places it in $C_\infty$. Equality at one finite step establishes invariance immediately and therefore gives the intersection.

If both $z$ and $-z$ lie in $C_\infty$, every test row annihilates $z$, so observability gives $z = 0$. For $z \in C_\infty$ with $\lambda z = 0$, sum word probabilities along any fixed action sequence. Mass conservation makes the sum zero. Every term is nonnegative, so every word test vanishes and again $z = 0$. Spanning by reachable cone elements gives the final assertion. □

A useful negative witness is a state admitted by the current inequalities and a specific longer word with negative proposed probability. It demonstrates that this state needs to be excluded. It does not prove that the initial state, or every reachable state, is invalid.

Finite stabilization is sufficient, not promised. Even three predictive coordinates can require a nonpolyhedral domain, as Section 51 shows. Checking a long list of trajectories is not the same as checking the invariance identity that makes every trajectory valid.

## 43. A finite certificate with a positive lift

Let $V = [v_1 \ \cdots \ v_m] \in \mathbb{R}^{r\times m}$ have full row rank and satisfy

$$\lambda V = \mathbf{1}_m^\top.$$

Its columns generate $C = V\mathbb{R}_+^m$ and its normalized base is $K = \operatorname{conv}\{v_1, \dots, v_m\}$. Finite generation makes the cone closed. Its normalized generators and full row rank make it pointed and generating, with a strictly positive mass functional.

**Proposition 39. Finite positive-lift certificate.** This cone is invariant under $B_s$ if and only if there is an entrywise nonnegative matrix $A_s \in \mathbb{R}^{m\times m}$ such that

$$\boxed{B_s V = V A_s, \qquad A_s \geq 0.}$$

If these identities hold for every symbol and $z_0 = V\alpha_0$ for $\alpha_0 \geq 0$ with $\mathbf{1}^\top \alpha_0 = 1$, then

$$\mathbf{1}^\top \sum_o A_{a,o} = \mathbf{1}^\top, \qquad \lambda B_w z_0 = \mathbf{1}^\top A_w \alpha_0.$$

Thus the finite witness defines a nonnegative probability instrument whose image under $V$ is the predictive model. It certifies all word probabilities without enumerating them.

**Proof.** Invariance places each $B_s v_j$ in the conical hull of the columns of $V$. Choose its nonnegative coefficients as column $j$ of $A_s$. Conversely, $B_s V\alpha = VA_s\alpha \in C$ for every $\alpha \geq 0$. Multiplying the intertwining identity by $\lambda$ and summing over observations gives

$$\mathbf{1}^\top \sum_o A_{a,o} = \lambda \sum_o B_{a,o} V = \lambda V = \mathbf{1}^\top.$$

Iterating $B_s V = V A_s$ gives $B_w V = V A_w$. Applying the mass row and the initial coefficients proves the word identity. Strict positivity of $\lambda$ follows because a nonzero conical combination has positive total coefficient mass. □

The $m$ nonnegative coordinates need not be physical hidden states, and $m$ need not equal the minimal predictive dimension $r$. They are one finite realization witnessing validity. When $V$ has dependent columns, the lift can be nonunique. Those redundant coordinates do not change the predictive state after synthesis.

There is a dual finite test. If a complete inequality description of the same cone is $C = \{z : Gz \geq 0\}$, then $GB_s v_j \geq 0$ for every generator $j$ is equivalent to invariance. One may also certify $GB_s = H_s G$ with $H_s \geq 0$. The latter certificate is sufficient by direct multiplication and necessary by the polyhedral cone form of Farkas' lemma [R13]. The reference calculation uses the direct generator test, not an unreported numerical optimization certificate.

## 44. Robustness is a quantified invariance statement

A certificate for one fitted operator does not automatically cover neighboring operators. For uncertainty, the quantifier must range over the whole proposed model set, including correlations required by mass conservation.

Let $\mathcal{U}$ be the convex hull of finitely many **complete operator families** $\{B_s^{(j)}\}_s$. Every family must satisfy the exact per-action mass identities. The convex coefficient is shared across that family's symbols. A product of independent uncertainty sets can instead be handled by its joint vertices.

**Proposition 40. Robust finite-domain certificate.** For a fixed normalized generator matrix $V$, suppose every vertex family has witnesses

$$B_s^{(j)} V = V A_s^{(j)}, \qquad A_s^{(j)} \geq 0.$$

Then every family in $\mathcal{U}$ preserves $C = V\mathbb{R}_+^m$ and is mass conserving. If its initial state lies in $K$, it defines valid finite observation-tree probabilities at every horizon. This remains true when a family in $\mathcal{U}$ is selected separately at each history.

For a complete facet description $C = \{z : Gz \geq 0\}$, let $g_i^\top$ be its rows. A sufficient norm-ball certificate around $\widehat{B}_s$ is

$$\boxed{g_i^\top \widehat{B}_s v_j \geq \epsilon_s \|g_i\|_2 \|v_j\|_2 \quad \text{for all } s, i, j.}$$

It covers all perturbations with $\|B_s - \widehat{B}_s\|_2 \leq \epsilon_s$ that also satisfy the exact mass constraints.

**Proof.** For convex weights $\theta_j$, take $A_s = \sum_j \theta_j A_s^{(j)}$. The equality and nonnegativity survive convex combination, as does mass conservation. At each history, every selected family maps the current cone state back into the cone. Applying the mass identity at that node proves recursive normalization, regardless of how the family was selected. For the norm ball,

$$|g_i^\top (B_s - \widehat{B}_s) v_j| \le \epsilon_s \|g_i\|_2 \|v_j\|_2.$$

The margin makes every perturbed generator image satisfy all facets. Proposition 39, or the direct generator test, then supplies invariance. □

The switching conclusion is structural validity, not an assertion that time-varying selection represents one fixed unknown parameter. A likelihood for a fixed mechanism and a likelihood for a mechanism selected anew at each history are different statistical models.

The norm-ball inequality is sufficient and can be conservative, especially at boundary images with zero slack. Failure of this margin is not a proof that no robust certificate exists. The exact vertex test retains uncertainty correlations that a separate bound for each entry may discard.

## 45. Fitting inside a certified domain

Once $V$ is fixed, certificate-constrained fitting is a convex problem. Unknowns include the entries of $\widetilde{B}_s$, $A_s$, $\widetilde{z}_0$ and $\alpha_0$. The constraints

$$\widetilde{B}_s V = V A_s, \quad A_s \ge 0, \quad \lambda \sum_o \widetilde{B}_{a,o} = \lambda,$$
$$\widetilde{z}_0 = V \alpha_0, \quad \alpha_0 \ge 0, \quad \mathbf{1}^\top \alpha_0 = 1$$

are linear. Entrywise confidence intervals and fixed-input regression tolerances add linear inequalities. Norm balls add convex constraints. Jointly optimizing $V$ and $A_s$ is different: the product $VA_s$ is then bilinear. No general convexity is claimed for learning the domain and dynamics together.

**Proposition 41. Certified correction and its error budget.** For fixed $V$, every feasible solution of the displayed constraints is an internally valid model. Suppose a center estimate $\widehat{B}_s$ and a feasible corrected estimate $\widetilde{B}_s$ satisfy

$$\|B_s^* - \widehat{B}_s\| \le \epsilon_s, \qquad \|\widetilde{B}_s - \widehat{B}_s\| \le \epsilon_s.$$

Then $\|\widetilde{B}_s - B_s^*\| \le 2\epsilon_s$. The same conclusion holds for initial states in their specified norm. If the true model lies in these confidence sets and itself admits the fixed-domain certificate, the convex constraints are feasible. Feasibility alone does not establish that the true model preserves this domain.

**Proof.** Validity is Proposition 39. The error bound is the triangle inequality through the center estimate. Under the final assumptions, insert the true operators, true initial state and their nonnegative witnesses into the constraints. They provide a feasible point. The converse fails because existence of one valid point does not say that every point in a confidence set, or its true member, has the same invariant domain. □

This is why existential correction and robust certification must be reported separately. A corrected model can be a well-defined approximation even when no common true-model cone has

been proved. The trajectory bounds below require both models to be valid on a common specified cone.

### 45.1 An explicit positivity repair in a simplicial cone

When $V$ is square and invertible, put $A_s = V^{-1}\widehat{B}_s V$. Assume mass conservation has already been imposed. Let $N_s$ be an entrywise positive reference instrument with the same per-action mass identities. For a mixing weight in $[0, 1]$, the mixture

$$\widetilde{A}_s = (1-\eta)A_s + \eta N_s$$

is nonnegative provided

$$\eta \geq \max_{s,i,j:(A_s)_{ij}<0} \frac{-(A_s)_{ij}}{(N_s)_{ij} - (A_s)_{ij}}.$$

An empty maximum is zero. If every negative entry has magnitude at most $d$ and every reference entry is at least $m_0 > 0$, the simpler sufficient choice is $\eta = d/(m_0 + d)$. These statements follow by solving each scalar inequality. Mass conservation is unchanged by convex mixing.

This is an explicit model modification, not coefficient clipping or an exact coordinate change. Relative to a true instrument on the same cone, the one-step error is at most $(1-\eta)\epsilon + 2\eta$ in the summed base-norm metric defined next, when the original error was at most $\epsilon$. The factor two follows because two valid instruments each distribute unit total mass. The reference example retains the repair amount as part of its approximation record.

## 46. A norm that measures probability mass before conditioning

Assume $C$ is a closed, pointed, generating cone and $\lambda$ is strictly positive on its nonzero elements. Its base $K$ is compact. Define the base norm

$$\|x\|_C = \inf\{\lambda u + \lambda v : x = u - v,\ u, v \in C\}.$$

The infimum is finite because the cone is generating. Compactness of $K$ makes this a genuine norm. Its unit ball is $\mathrm{conv}(K \cup (-K))$. In particular $\|u\|_C = \lambda u$ on $C$, and $|\lambda x| \leq \|x\|_C$. On the ordinary probability orthant this is the $\ell^1$ norm. These are standard ordered-space constructions [R15], used here to avoid mixing Euclidean coefficient magnitude with probability mass.

**Proposition 42. Instrument contraction before normalization.** If $\{B_{a,o}\}_o$ is cone preserving and mass conserving for a fixed action, then

$$\boxed{\sum_o \|B_{a,o}x\|_C \leq \|x\|_C.}$$

For a polyhedral base $K = \mathrm{conv}\{v_j\}$, define the discrepancy between two instruments by

$$\epsilon_a = \sup_{z \in K} \sum_o \|(B_{a,o} - \widetilde{B}_{a,o})z\|_C.$$

The supremum equals the maximum of the same expression at the finitely many generators $v_j$. When both instruments are valid on $C$, $\epsilon_a \leq 2$.

**Proof.** For any $x = u - v$ with $u, v \in C$, cone preservation gives

$$\sum_o \|B_{a,o}x\|_C \leq \sum_o \lambda B_{a,o}(u + v) = \lambda u + \lambda v.$$

Take the infimum. For the generator statement, the discrepancy is convex in $z$. At a convex combination its value is at most the same combination of generator values, hence at most their maximum. The generators themselves belong to $K$, proving equality. Finally, for $z \in K$, the triangle inequality bounds the sum by the total masses of both instruments, which are one each. □

For a polyhedral certificate, the norm itself has a finite linear-program formulation:

$$\|x\|_C = \min_{a,b \geq 0} \{\mathbf{1}^\top (a + b) : V(a - b) = x\}.$$

If $V$ is square and invertible, this reduces to $\|V^{-1}x\|_1$. The exact worked interval certificate is of this type. With dependent generators, choosing one pseudoinverse does not in general compute the base norm.

## 47. Whole observation laws can be stable without rare-event margins

The posterior after one rare observation can be sensitive, but that does not force the entire distribution of observation records to be equally sensitive. Work before normalization and retain the probability of each branch.

Let two models share $(C, \lambda)$ and have initial states $z_0, \widetilde{z}_0 \in K$. At each time use the same policy, a specified function or randomization depending only on the observed history. Thus on any common history the action probabilities are identical in the two comparisons. This is not a comparison of two policies optimized separately for the two models.

**Proposition 43. Observation-tree total-variation bound.** If both instruments preserve $C$ and conserve mass, and if every action at time $t$ has the discrepancy in Proposition 42 bounded by $\epsilon_t$, then the distributions $P_T, \widetilde{P}_T$ of length-$T$ observation records under the common policy satisfy

$$\boxed{\|P_T - \widetilde{P}_T\|_{\mathrm{TV}} \leq \min\left(1, \frac{1}{2}\|z_0 - \widetilde{z}_0\|_C + \frac{1}{2}\sum_{t=0}^{T-1} \epsilon_t\right).}$$

The same bound holds for joint action and observation records when the common policy is randomized. No positive lower bound on individual branch probabilities is required.

**Proof.** Let $u_h, \widetilde{u}_h \in C$ be the unnormalized states on each history at depth $t$, including the common policy factors. Their masses sum to one in each model. Define $D_t = \sum_h \|u_h - \widetilde{u}_h\|_C$. At a history with selected action $a$, write

$$B_{a,o}u_h - \widetilde{B}_{a,o}\widetilde{u}_h = B_{a,o}(u_h - \widetilde{u}_h) + (B_{a,o} - \widetilde{B}_{a,o})\widetilde{u}_h.$$

Sum the first term over observations and use Proposition 42. When $\lambda\widetilde{u}_h > 0$, normalize only to apply the definition of $\epsilon_t$ to the second term. Its sum is bounded by $\epsilon_t\lambda\widetilde{u}_h$. If the mass is zero, strict positivity implies $\widetilde{u}_h = 0$. Summing histories yields $D_{t+1} \le D_t + \epsilon_t$. Averaging over common policy probabilities gives the same inequality for randomized actions. Finally each word-probability difference is bounded by its state difference because $|\lambda x| \le \|x\|_C$. Total variation is half the sum of those absolute differences. □

This bound separates total law error from conditional-state error. It can grow linearly with the horizon and eventually become uninformative. It is not a uniform-in-time approximation theorem. It also does not control the regret of a changed planning policy without additional decision-theoretic assumptions.

## 48. Conditioning still has an unavoidable cost

Let $P, Q$ be probability distributions on a finite record space. A record event $H$ might be a particular observed prefix. Conditional continuation laws are obtained by restriction to $H$ and normalization.

**Proposition 44. Sharp event-probability dependence of conditioning.** If $\|P - Q\|_{\mathrm{TV}} \le \delta$, $p = P(H) > 0$ and $q = Q(H) > 0$, then

$$\boxed{\|P(\cdot \mid H) - Q(\cdot \mid H)\|_{\mathrm{TV}} \le \min(1, \delta/p).}$$

When $p > \delta$, positivity of $q$ follows automatically. The factor $1/p$ cannot be removed in general.

**Proof.** Write the subprobability restrictions on $H$ as $u, v$, of masses $p, q$. On the complementary event, the sum of absolute probability differences is at least $|p - q|$. Therefore

$$\|u - v\|_1 + |p - q| \le 2\delta.$$

Adding and subtracting $v/p$ gives

$$\frac{1}{2}\|u/p - v/q\|_1 \le \frac{\|u - v\|_1 + |p - q|}{2p} \le \delta/p.$$

Also $|p - q| \le \delta$, establishing the positivity condition. For sharpness, put total mass $p$ on an event with two outcomes. Let $P$ assign all of it to the first, and let $Q$ move mass $\delta \le p$ to the second, leaving the complement unchanged. The unconditional total variation is $\delta$ and the conditional distance is $\delta/p$. □

Combining this proposition with Proposition 43 gives a conditional continuation bound when the evidence event has adequate probability. It does not give a coordinate-state error bound unless test probabilities control that coordinate norm. Validity at every branch and stable inference at every branch remain different properties.

## 49. Searching for a domain from observable operators

A cone certificate need not be imported from a hidden-state generator. With exact rational operators it can be searched for directly, then checked independently.

**Proposition 45. Rational finite-certificate search.** Suppose $\lambda$, $B_s$ and $z_0$ are rational and satisfy the exact mass identities. There is an enumeration procedure that returns a valid finite certificate whenever a certificate of Proposition 39 exists with a rational normalized full-row-rank generator matrix $V$. The procedure is sound whenever it returns. No termination guarantee is asserted when no such rational certificate exists.

**Proof.** Enumerate all finite rational matrices $V$ in an order that eventually includes every size and every bounded numerator and denominator range. Retain those of full row rank with $\lambda V = \mathbf{1}^\top$. For each retained matrix, the requirements $B_s V = V A_s$, $A_s \geq 0$, $V\alpha_0 = z_0$, $\alpha_0 \geq 0$ and $\mathbf{1}^\top \alpha_0 = 1$ are rational linear feasibility problems. A nonempty rational polyhedron has a rational feasible point, as follows from elimination of rational linear inequalities. These problems can therefore be decided by exact rational linear programming, or witnesses can be enumerated in a dovetailed rational search. Recheck every returned equality and sign exactly. Proposition 39 proves soundness. If a rational generator certificate exists, its finite representation is eventually considered and its feasible witnesses are found. □

The general enumeration is a mathematical existence algorithm, not a practical complexity claim. A bounded candidate family is often the useful reference experiment. Failure within that family means only that the chosen family did not produce a certificate.

### 49.1 A certificate learned without hidden matrices

The observable tables of Section 38 determine the two-dimensional operators $B_s$ and initial vector $(1, 13/24)^\top$. The new reference procedure receives only these fitted operators and that vector. It enumerates intervals with rational endpoints in $[0, 1]$ of reduced denominator at most eight, ordered by increasing width and then by endpoints. The coordinate restriction $[0, 1]$ comes from its declared test-probability meaning.

The first accepted interval is

$$\boxed{K = \{(1, q)^\top : 2/5 \leq q \leq 3/5\}.}$$

It uses

$$V = \begin{pmatrix} 1 & 1 \\ 2/5 & 3/5 \end{pmatrix}, \qquad \alpha_0 = \begin{pmatrix} 7/24 \\ 17/24 \end{pmatrix}, \qquad A_s = V^{-1} B_s V.$$

Every entry of every $A_s$ is at least $1/64$. The full per-action column sums are exactly one. These rational identities certify all word probabilities of the fitted model, not only the words used in the fit. The certificate was not supplied to the search, and no hidden generator is imported by the standalone example.

The interval is the narrowest accepted interval **in this specified finite grid**, not a claim of the smallest invariant interval among all real endpoints. It also need not contain every predictive state associated with an arbitrary initialization of the old hidden model. It contains the declared initial state and preserves every subsequent branch of this fitted realization, which is the requirement being certified.

The positive slack gives a robust statement: any perturbation of the lifted entries by at most $1/64$, coupled to retain the exact per-action column sums, remains nonnegative. Interior perturbations of size $1/128$ are used in the worked comparison. Repeated composition is valid for all horizons, while Proposition 43 bounds the changed observation law for each fixed horizon.

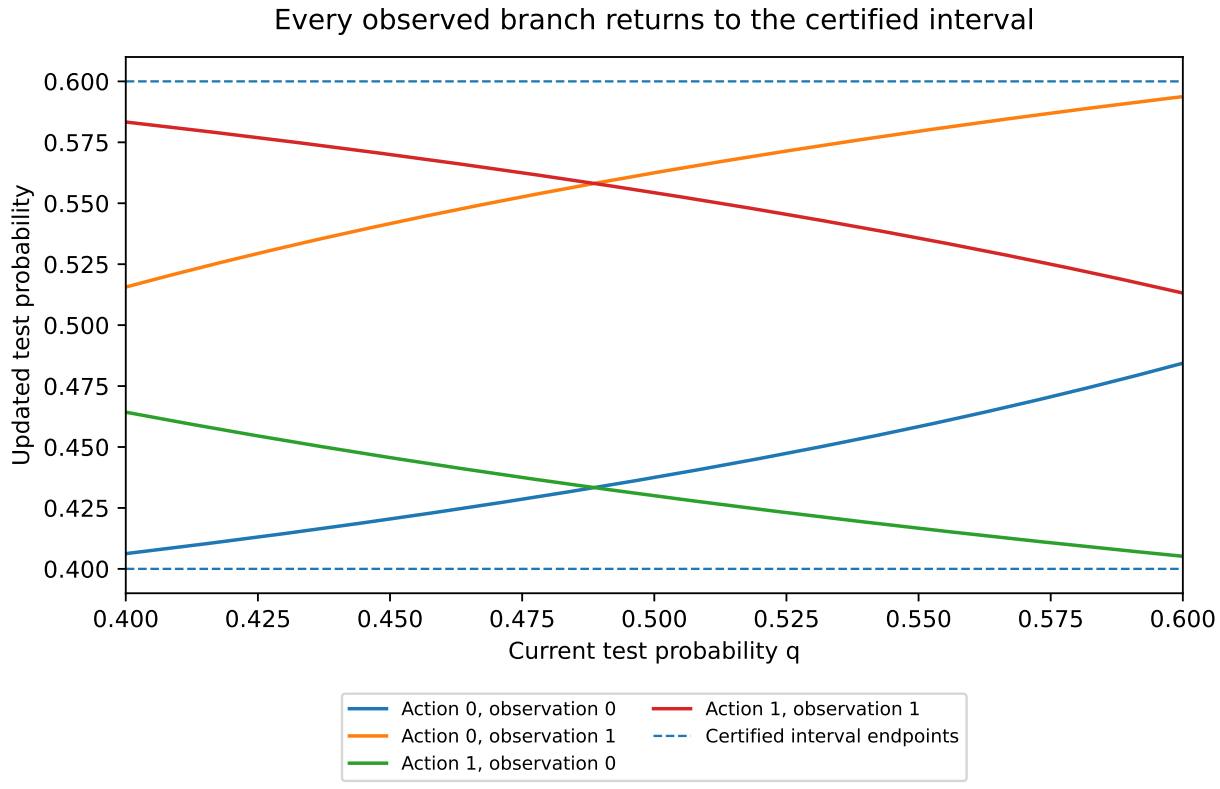



**Figure 4.1:** The fitted two-coordinate model maps every branch back into the interval from $2/5$ to $3/5$. Its exact positive-lift witness checks generators, nonnegative branch matrices, mass conservation and the initial weights. The plotted maps illustrate the same certified interval. The plotted grid is not the all-horizon proof. Intertwining and nonnegativity establish validity of the submitted model. They do not establish that it is the true process.

## 50. Selection and noise do not invalidate an exact witness

A data-dependent cone proposal is not itself a statistical confidence statement. Once the matrices and witness are fixed, its exact invariance identities hold or fail independently of how it was proposed. What remains statistical is whether the chosen uncertainty set covers the generating model.

For a finite-sample interpretation, first obtain a valid uncertainty set for a fixed identifiable model class. Proposition 35 supplies one route under independent reset experiments and declared rank coverage. Proposition 33 translates those probability bounds into a fixed-core operator bound. A validity witness can then be sought for a fitted member or robustly for every member. Only the latter identifies a common true-model domain on the coverage event.

The admissible family must enforce mass conservation exactly. A free box around unconstrained matrices generally contains nonnormalized families. Discarding those families is legitimate only because the generating process is already assumed to obey the exact mass law. Other cuts on a confidence set need their own justification.

## 51. Why finite predictive rank need not mean finitely many positive states

The following controlled rational example makes the boundary explicit. It is a worked adaptation of the irrational-rotation obstruction in [R12, Section 7]. Related distinctions among finite linear, positive and quantum realizations are discussed in [R14]. Only the elementary cone argument below is used.

Let $z = (t, x, y)^\top$, $\lambda = (1, 0, 0)$, $z_0 = (1, 1, 0)^\top$, and

$$R = \begin{pmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{pmatrix}, \qquad T = \operatorname{diag}(1, R).$$

There are two actions and two observations. A rotate action has branches

$$B_{\mathrm{r},0} = T, \qquad B_{\mathrm{r},1} = 0.$$

A probe action has branches

$$B_{\mathrm{p},0} = z_0(1/2, 1/2, 0), \qquad B_{\mathrm{p},1} = z_0(1/2, -1/2, 0).$$

The rotate outcome zero is certain. The probe has probabilities $(t + x)/2$ and $(t - x)/2$ and resets every nonzero branch to $z_0$ after normalization. Both actions conserve the mass $t$.

**Proposition 46. A rational rank-three model with no polyhedral probability-domain certificate.** The Lorentz cone

$$C_{\mathrm{L}} = \{(t, x, y) : t \geq \sqrt{x^2 + y^2}\}$$

is probability valid for these operators. It is the canonical all-test cone. Every closed convex invariant cone containing $z_0$ and admitting the nonnegative mass interpretation equals $C_{\mathrm{L}}$. In particular no finite normalized generator certificate exists for this exact model. Its reachable and observable predictive dimension is three.

**Proof.** Rotation preserves $t$ and $x^2 + y^2$. On $C_{\mathrm{L}}$ both probe effects are nonnegative and each probe image is a nonnegative multiple of $z_0$. Thus the displayed cone is invariant and valid.

Write the rotation angle as $\theta$. It is not a rational multiple of $2\pi$. Otherwise $\zeta = e^{i\theta}$ would be a root of unity, so $\zeta + \zeta^{-1} = 6/5$ would be a rational algebraic integer. Every rational algebraic integer is an integer, by the rational-root argument for a monic integer polynomial, contradicting $6/5 \notin \mathbb{Z}$. The forward powers of an irrational circle rotation are dense. Compactness supplies powers approaching the identity, so their closure contains the inverse rotation and is a closed subgroup. An infinite closed subgroup has arbitrarily small positive angles. Integer multiples of such angles approximate every point on the circle, so that subgroup is the entire circle.

A state passing every test must pass a probe after any number $k$ of rotations. Hence

$$t \pm e_1^{\top} R^k (x, y)^{\top} \geq 0 \quad \text{for every } k \geq 0.$$

Density gives $t \geq \sqrt{x^2 + y^2}$. The reverse inclusion was already proved, so the canonical cone is exactly $C_{\mathrm{L}}$.

Any invariant cone with nonnegative mass lies inside this canonical cone, by Proposition 38. If it is closed, convex and contains $z_0$, it contains the orbit $T^k z_0$ and the closed convex hull of that orbit. On $t = 1$, this hull is the full unit disk. Conical scaling gives all of $C_{\mathrm{L}}$. The unit disk has infinitely many extreme points, so it is not the convex hull of finitely many points.

Finally the reachable columns $z_0, Tz_0, T^2 z_0$ are independent. The mass row and probe rows before zero and one rotations span the dual space because the rotation has nonzero sine. Reachability and observability establish predictive dimension three. □

This is not a failure of finite-dimensional prediction. The exact model has only three coordinates and a simple quadratic cone. It is a failure of the narrower requirement that every valid model admit a finite *polyhedral* domain. A certificate language limited to finitely many positive generators would exclude a legitimate low-rank signal model.

No general impossibility of all finite certificates follows. This very example has a finite quadratic certificate. The obstruction concerns polyhedral certificates, not invariant cones described by other mathematics.

## 52. A finite approximation with an explicit price

A curved probability domain can sometimes be replaced by a finite one after changing the dynamics. That change must be quantified, not described as exact refinement.

Let $P_n$ be the regular $n$-gon inscribed in the unit disk, with a vertex at $(1,0)$ and $n \geq 3$. Replace the rotate branch by

$$T_\alpha = \operatorname{diag}(1, \alpha R), \qquad 0 \leq \alpha < 1,$$

while retaining the same probe branches. The spatial coordinates are attenuated, but total probability mass remains unchanged.

**Proposition 47. Polyhedral approximation of the rotation model.** If

$$\alpha \leq \cos(\pi/n),$$

then the cone over $P_n$ is invariant under all modified operators and contains $z_0$. Consequently the modified model admits a finite positive lift. The original and modified models are both valid on the larger cone $C_{\mathrm{L}}$, and for every common policy their length-$T$ laws satisfy

$$\boxed{\|P_T - P_T^{(\alpha)}\|_{\mathrm{TV}} \leq \min(1, T(1-\alpha)/2).}$$

The angle-independent sufficient condition can be improved for a fixed polygon and rotation. For the rational diamond $|x| + |y| \leq t$, the present rotation is invariant after attenuation whenever $\alpha \leq 5/7$. Thus $\alpha = 2/3$ gives an entirely rational four-generator certificate.

**Proof.** The regular polygon contains the disk of radius $\cos(\pi/n)$ and lies in the unit disk. Every $\alpha R v$ with $v \in P_n$ has Euclidean norm at most $\alpha$, hence lies in $P_n$ under the stated condition. The probe effects remain nonnegative because $P_n$ lies in the disk, and their normalized images are the retained vertex $(1,0)$.

The base norm of $C_{\mathrm{L}}$ is

$$\|(t,x,y)\|_{C_{\mathrm{L}}} = \max(|t|, \sqrt{x^2+y^2}).$$

Indeed the convex hull of the unit disks in the planes $t = 1$ and $t = -1$ is the cylinder with this unit ball. The rotate-branch discrepancy on a normalized state is at most $1-\alpha$ in this norm. Probe discrepancies are zero. Proposition 43 gives the law bound. For the diamond, the induced $\ell^1$ norm of $R$ is $7/5$, so $\alpha\|R\|_1 \leq 1$ proves invariance. □

For a target attenuation loss $0 < \eta \leq 1$, choosing $\alpha = 1-\eta$ and sufficiently many polygon vertices satisfies the certificate. The inequality $1 - \cos u \leq u^2/2$ shows that $n \geq \pi/\sqrt{2\eta}$ is sufficient, with an integer ceiling and $n \geq 3$. This is a constructive upper bound on certificate size, not a lower bound or a proof of optimality.

The distinction is exact. At $\alpha = 1$ there is no finite polyhedral certificate for this model. At $\alpha < 1$ a sufficiently fine polygon works, at the cost of a declared change in the observation law. The reference figures show that tradeoff rather than labeling the approximation an identical model.

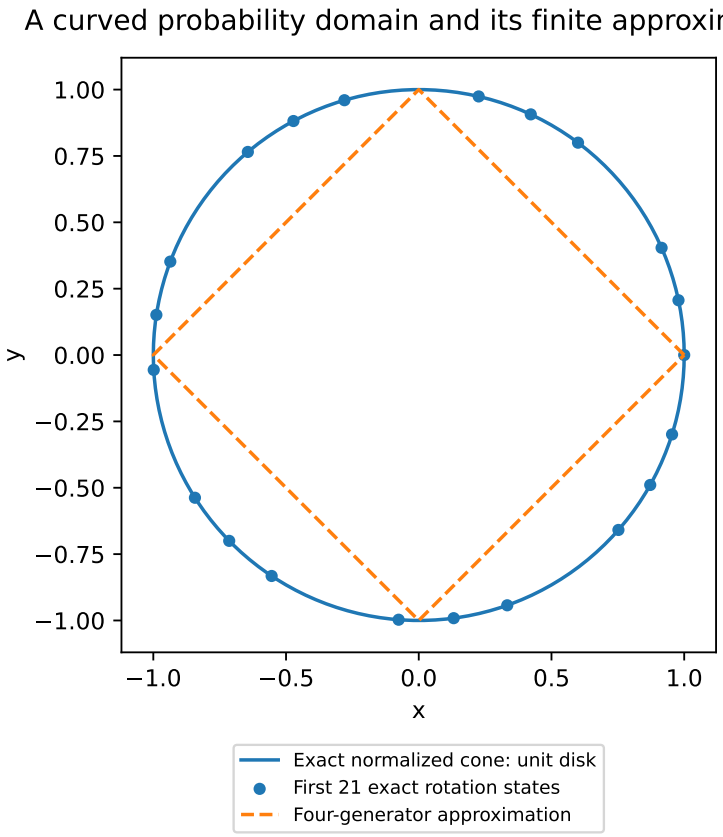



**Figure 4.2:** The rational rotation example has a three-coordinate realization whose normalized probability domain is the disk. The displayed orbit is finite, while the proof establishes density and rules out an exact finite polyhedral domain for this model. The inscribed diamond becomes useful only after a declared modification of the dynamics. Replacing the disk by a finite picture is not exact state refinement.

## 53. Repeated proposal and verification with a confidence budget

Model development may choose another core, propose another domain or collect another independent batch after an earlier certificate is withheld. Statistical coverage must survive that selection process.

**Proposition 48. Sequential confidence accounting for certified proposals.** At stage $n$, let the proposal be measurable from previous data, and let a verification procedure return a set $\mathcal{U}_n$ for the true model in the declared identifiable coordinate system. Assume conditional coverage

$$\Pr(\text{true model} \notin \mathcal{U}_n \mid \text{previous data}) \le \delta_n, \qquad \sum_{n \ge 1} \delta_n \le \delta.$$

Then with probability at least $1 - \delta$, every returned uncertainty set covers its true model. On this event, any data-selected domain with an exactly verified robust certificate for **all** members of its corresponding $\mathcal{U}_n$ is also valid for the true model, provided the true initial state is covered by the certified base. The conclusion remains valid at any stopping time chosen from the recorded results.

**Proof.** Take expectations of the conditional failure bounds, then apply the countable union bound. Off the union of failure events, all coverages hold simultaneously. Robust invariance over an entire set then applies to its true member by deterministic implication, independently of how its domain was chosen. A stopped stage is one of the already covered stages. □

For example, $\delta_n = \delta 2^{-n}$ spends a finite total confidence budget. Fresh reset trials for a proposal fixed by past data can use Proposition 35 conditionally, with this stage budget. An exact witness needs no statistical correction for how many candidate cones were tried. The coverage statement does need correction for repeatedly constructing confidence sets or inspecting post-selected data.

The proposition does not infer a global rank bound, validate an unknown reset operation, or transform dependent observations into independent trials. Those are assumptions of the coverage procedure. A known coordinate transformation can transport a confidence set exactly. An estimated transformation requires the subspace and conditioning controls already separated in Chapter 3.

## 54. Integrated certificate-carrying realization

**Proposition 49. Conditional learning with an all-horizon probability certificate.** Suppose the observable identification assumptions of Proposition 37 hold for a declared finite-rank controlled process. Suppose a justified confidence procedure produces an operator and initial-state set containing the true realization on an event of probability at least $1 - \delta$. Assume the per-action mass constraints are exact.

If a fitted realization admits a checked finite witness of Proposition 39, it is an internally valid stochastic model at every horizon. If the witness is robust over the whole confidence set and covers its initial states, then on the coverage event the same probability domain is valid for both the fitted and true models. When their summed base-norm instrument discrepancies are at most $\epsilon_t$ and their initial discrepancy is at most $e_0$, every common policy has the bound

$$\|P_T - \widehat{P}_T\|_{\mathrm{TV}} \le \min\left(1, \frac{e_0 + \sum_{t<T} \epsilon_t}{2}\right).$$

Conditioning on a positive-probability event then obeys Proposition 44. Exact coordinate transport preserves the certificate. A signed packet decomposition preserves it after complete synthesis but does not turn individual packets into probability states without Proposition 30's extra positivity requirement.

**Proof.** Observable identification supplies the specified realization and the confidence procedure supplies the stated event, not an inferred event of its own. Proposition 39 supplies internal validity for the fitted model. Proposition 40 extends the same domain to every member of the

uncertainty set, hence to its true member on that event. Propositions 42 and 43 give the total-variation bound, and Proposition 44 gives the conditional bound. For a coordinate change $T$, use $V' = TV$, $B'_s = TB_sT^{-1}$ and $\lambda' = \lambda T^{-1}$. Then $B'_sV' = V'A_s$ and $\lambda'V' = \mathbf{1}^\top$ with the unchanged nonnegative witnesses. Full packet synthesis is the identity. □

This closes a concrete foundational gap. A learner can now return not just a matrix family but a verifiable domain in which that family has probability meaning. It can attach a robust uncertainty statement when one is justified, or return an internally valid approximation with its correction budget explicitly recorded.

It does not turn a successful fit into unrestricted knowledge of the world. The model class, observation protocol, rank coverage, exact mass law, domain witness and approximation budget remain visible parts of the result. Failure of a finite search is a withheld certificate, not a proof of invalidity. The nonpolyhedral example explains why this distinction is necessary even for very small exact models.

## 55. From polyhedral to curved certificates

A finite positive lift is one sufficient certificate language. Proposition 46 shows that a low-dimensional model can require a nonpolyhedral domain. Chapter 5 represents that same model exactly through a curved matrix domain. A failed finite-generator search is therefore not a general invalidity theorem.

CHAPTER 5

# Curved domains and continuous observations

Chapter 4 separated a valid predictive model from a successful fit. Its finite positive lift gives a useful certificate, but its rank-three rotation example has no finite polyhedral probability domain. That chapter also left continuous observations outside the finite-alphabet construction. This part addresses both boundaries in a declared finite sector.

A curved domain can have a short matrix certificate even when it has infinitely many extreme states. A continuous observation can have a finite operator description even when its possible values are uncountable. Neither statement follows from finite predictive dimension alone. The constructions below require an explicit domain, a reference measure and a restricted observation-kernel family.

This chapter connects matrix positivity, integral observation semantics, population-moment identification, exact coarsening and approximation. It also identifies a defect that is easy to overlook: averaging a matrix transformation before applying its quadratic action does not equal averaging that action. The difference is a positive covariance contribution. Omitting it can destroy normalization even in a very small example.

There are established antecedents. Congruence representations and their positive semidefinite Gram matrices belong to the theory of completely positive maps [R16]. Polynomial Gram certificates are standard sum-of-squares machinery [R17]. Continuous-valued observable operator models already occur in Jaeger's work [R18]. The proofs here develop their specific connection to the realization and defect contracts of the large signal model. They do not claim to originate those ingredients.

This is a discrete-time theory with continuous-valued observations. It is not a theorem about continuous-time stochastic dynamics.

## 56. A matrix certificate for a curved domain

Let $E = \mathbb{R}^r$ and let $\mathbb{S}^d$ be the real symmetric $d \times d$ matrices. Fix an injective linear map

$$\mathcal{J} : E \longrightarrow \mathbb{S}^d, \qquad \lambda z = \operatorname{tr} \mathcal{J}(z).$$

Suppose some $e \in E$ satisfies $\mathcal{J}(e) \succ 0$. Define

$$C_{\mathcal{J}} = \{z : \mathcal{J}(z) \succeq 0\}, \qquad K_{\mathcal{J}} = \{z \in C_{\mathcal{J}} : \lambda z = 1\}.$$

The map $\mathcal{J}$ is an embedding, not a many-to-one decoder of a larger positive coordinate space. The cone is an inverse image of the positive semidefinite cone. This is different from the image cone $V\mathbb{R}^m_+$ used in Proposition 39.

**Proposition 50. Matrix-embedding probability certificate.** The cone $C_{\mathcal{J}}$ is closed, pointed and generating, and $\lambda$ is strictly positive on its nonzero elements. Its normalized base is compact. Suppose a finite-alphabet predictive model has $z_0 \in K_{\mathcal{J}}$, exact mass conservation and identities

$$\boxed{\mathcal{J}B_s = \Phi_s \mathcal{J},} \qquad \Phi_s(X) = \sum_{j=1}^{m_s} w_{sj} K_{sj} X K_{sj}^{\mathsf{T}}, \quad w_{sj} \geq 0.$$

Then every $B_s$ preserves $C_{\mathcal{J}}$, and every finite observation word has a nonnegative, normalized probability. Zero maps may use an empty sum. An invertible change of predictive coordinates transports the certificate by

$$B'_s = TB_sT^{-1}, \qquad \mathcal{J}' = \mathcal{J}T^{-1}, \qquad \lambda' = \lambda T^{-1}.$$

**Proof.** Closedness follows from continuity. If both $z$ and $-z$ are in the cone, $\mathcal{J}(z)$ is both positive and negative semidefinite, hence zero. Injectivity gives $z = 0$. A sufficiently small neighborhood of $e$ maps into positive definite matrices, so the cone has an interior point and is generating. A nonzero positive semidefinite matrix has positive trace. The trace-one positive semidefinite matrices form a compact set. The image of $\mathcal{J}$ is a closed finite-dimensional subspace and its inverse on that image is continuous, so $K_{\mathcal{J}}$ is compact.

Every term $wKXK^{\mathsf{T}}$ is positive semidefinite when $X$ is. The intertwining identity therefore proves invariance. Iteration gives invariance of every branch state, and the mass identity normalizes every finite observation tree. Substituting the transported quantities proves the final statement. □

This is a sufficient certificate language. It is not asserted that every invariant cone has this form, or that every positive map on a matrix subspace extends through the displayed congruence representation. The extension requirement is stronger than positivity on the represented states alone.

The base norm of Section 46 is also not automatically the ambient trace norm. One always has

$$\|\mathcal{J}(z)\|_1 \leq \|z\|_{C_{\mathcal{J}}},$$

because every positive decomposition in $E$ gives one in the matrix space. Equality holds when $\mathcal{J}$ maps onto all of $\mathbb{S}^d$, by the positive and negative spectral parts. For a proper matrix subspace, those spectral parts may leave the subspace. Norm transport therefore needs its own stated hypothesis.

## 57. The earlier disk obstruction has an exact finite matrix witness

For the rank-three model of Section 51, put

$$\mathcal{J}(m,x,y) = \frac{1}{2}\begin{pmatrix} m+x & y \\ y & m-x \end{pmatrix}.$$

Then $\operatorname{tr}\mathcal{J} = m$ and

$$\det\mathcal{J} = \frac{m^2 - x^2 - y^2}{4}.$$

Its eigenvalues are $(m \pm \sqrt{x^2+y^2})/2$, so positivity is exactly the Lorentz-cone condition. The normalized base is the unit disk.

**Proposition 51. Exact rational matrix witness for the rotation model.** Let $T = \operatorname{diag}(1, R)$ with the unchanged rotation

$$R = \begin{pmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{pmatrix}, \qquad W = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}.$$

The rotate branch and the two probe branches from Proposition 46 have the identities

$$\begin{aligned} \mathcal{J}(Tz) &= \frac{1}{5} W \mathcal{J}(z) W^{\mathsf{T}}, \\ \mathcal{J}(B_{\mathrm{p},0} z) &= E_{00}\mathcal{J}(z)E_{00}^{\mathsf{T}}, \\ \mathcal{J}(B_{\mathrm{p},1} z) &= E_{01}\mathcal{J}(z)E_{01}^{\mathsf{T}}, \end{aligned}$$

where $E_{ij}$ is the corresponding $2 \times 2$ matrix unit. The other rotate branch is zero. Thus this exact model has a finite rational matrix certificate without attenuation or a polygonal replacement.

**Proof.** Direct multiplication gives the rotate identity. The two matrix-unit congruences retain, respectively, the upper and lower diagonal entries of $\mathcal{J}(z)$ and place them in its upper diagonal entry. These are $(m+x)/2$ and $(m-x)/2$, exactly the two probe masses resetting to $(1,1,0)$. Also $W^{\mathsf{T}}W = 5I$ and

$$E_{00}^{\mathsf{T}}E_{00} + E_{01}^{\mathsf{T}}E_{01} = I.$$

The actions conserve trace. The initial state embeds as $\operatorname{diag}(1,0)$, and $(1,0,0)$ is a strictly positive interior witness. Proposition 50 applies. □

Nothing has been changed in the observation law. Proposition 46 still rules out a finite positive-generator certificate for this model. What changes is the certificate language. A curved matrix domain and a finite set of positive coordinate states are not the same representation class.

This use of symmetric positive semidefinite matrices is mathematical. It neither requires quantum hardware nor claims a quantum description of the measured system. It borrows a well-defined operator representation rather than a physical interpretation.

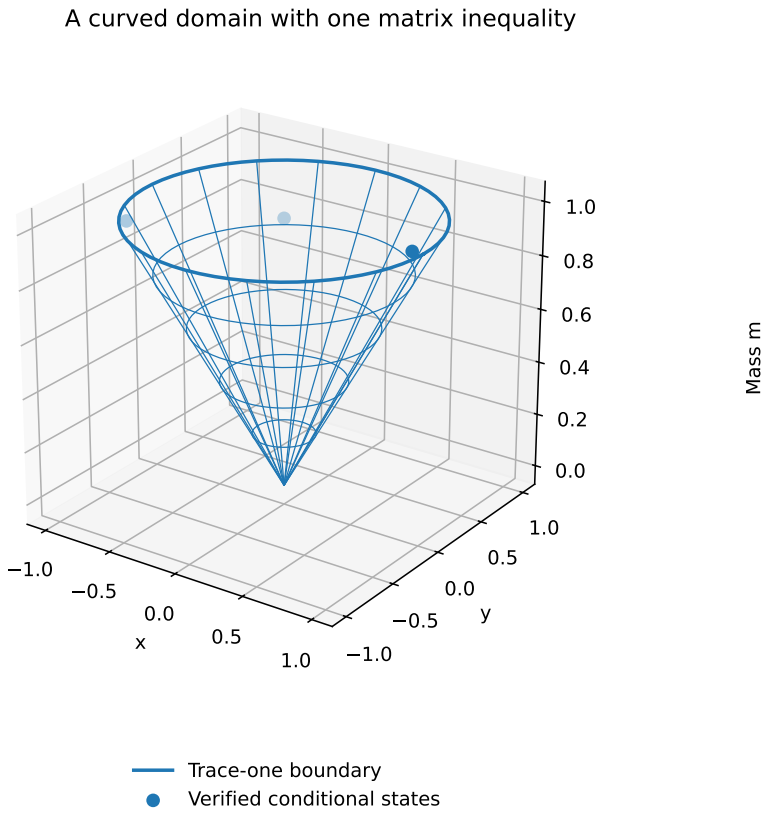



**Figure 5.1:** The Lorentz cone admits a two-by-two positive-semidefinite matrix representation. Congruence identities certify the unchanged rotation and observation operations. This resolves the certificate-language limitation without damping the exact model. The matrix representation is a mathematical positivity witness. No quantum hardware, physical quantum process or quantum advantage is asserted.

## 58. An observation density is not a branch probability

Let $(Y, \mathcal{Y})$ be a standard Borel observation space and let $\mu$ be a fixed probability reference measure. Keep a finite action set. For each action $a$, let $B_a(y) : E \to E$ be measurable and integrable in operator norm. Let $(C, \lambda)$ be a proper probability cone and mass functional as in Section 46. Assume that, outside one $\mu$-null set for each action,

$$B_a(y)C \subseteq C, \qquad \lambda \int_Y B_a(y)\, \mu(dy) = \lambda.$$

The second identity is an equality of linear functionals. The first is a statement on the whole cone, not only on a finite selection of states.

**Proposition 52. Continuous-observation instrument semantics.** For $z \in K = \{z \in C : \lambda z = 1\}$,

$$\boxed{P_a(A \mid z) = \int_A \lambda B_a(y) z \, \mu(dy), \qquad z^+(y) = \frac{B_a(y)z}{\lambda B_a(y)z}}$$

define a probability kernel and, for $\mu$-almost every observation with positive denominator, a conditional state. They generate consistent finite record laws under any common measurable history-dependent policy. For prescribed actions, the joint observation density is

$$\lambda B_{a_T}(y_T) \cdots B_{a_1}(y_1) z_0$$

with respect to $\mu^{\otimes T}$. Replacing the reference measure by $\mu' = r\mu$, with $r > 0$ almost everywhere and $\int r \, d\mu = 1$, requires $B'_a(y) = B_a(y)/r(y)$. Event probabilities and normalized states then remain unchanged.

**Proof.** Cone positivity makes the density nonnegative and the integrated mass identity makes its integral one. A positive-density update lies in $K$. Assign any fixed state of $K$ on the common exceptional null set and at zero-density points to obtain a measurable version of the update. This arbitrary choice has no effect on the probability kernel. Repeated integration, or conditioning one step at a time, proves normalization and consistency of the finite laws. Measurable policy factors are nonnegative and sum to one at each history. The reference-measure transformation cancels $r$ in each integral and cancels $1/r$ in the normalized update. □

When $\mu$ has no atoms, an individual point has probability zero even when its density is positive. A density may exceed one. Neither fact is a numerical pathology. Flooring a vanishing denominator changes the model and is not part of this theorem.

A continuous density at an exact point is a version-dependent object on null sets. The update formula specifies a convenient version wherever its denominator is positive. Event probabilities are the invariant statistical objects. Comparing model densities requires a common measure or a declared change-of-measure calculation.

A representation that lacks an operator-norm integrable density needs a more general operator-valued measure. Such representations are not silently covered by the density hypothesis above.

## 59. A finite certificate over an entire observation interval

Specialize to $Y = [-1, 1]$ and $\mu(du) = du/2$. Suppose $B_a(u)$ is a matrix polynomial. A finite congruence certificate can prove positivity for every $u$, not merely for values on a grid.

For a real $d^2 \times d^2$ symmetric matrix $F$, define

$$[\Phi_F(X)]_{ij} = \sum_{p,q=1}^{d} F_{(i,p),(j,q)} X_{pq}.$$

Vectorization orders the pairs with output index first inside each input column. If

$$F = \sum_{\nu} w_\nu \operatorname{vec}(K_\nu) \operatorname{vec}(K_\nu)^{\mathsf{T}}, \qquad w_\nu \geq 0,$$

then $\Phi_F(X) = \sum_\nu w_\nu K_\nu X K_\nu^{\mathsf{T}}$. Every positive semidefinite $F$ has a real factorization of this kind. The matrix is the usual Choi/Gram representation, with an explicit index convention [R16].

**Proposition 53. Polynomial Gram certificate for interval observations.** Fix the injective embedding $\mathcal{J}$ and its strict interior witness from Proposition 50. Set

$$Z_q(u) = \begin{pmatrix} I_{d^2} \\ u I_{d^2} \\ \vdots \\ u^q I_{d^2} \end{pmatrix}.$$

Suppose each action has constant positive semidefinite matrices $G_{a,0}, G_{a,1}$, with the second term omitted when $q = 0$, such that

$$\begin{aligned} F_a(u) &= Z_q(u)^{\mathsf{T}} G_{a,0} Z_q(u) \\ &\quad + (1-u^2) Z_{q-1}(u)^{\mathsf{T}} G_{a,1} Z_{q-1}(u), \\ \mathcal{J} B_a(u) &= \Phi_{F_a(u)} \mathcal{J}. \end{aligned}$$

Assume also $z_0 \in K_{\mathcal{J}}$ and the exact integrated mass identity. Then the model is internally probability-valid for every observation value in $[-1, 1]$ and every finite horizon. The intertwining condition is finitely many coefficient equalities. The normalization condition uses only

$$\int_{-1}^{1} u^k \frac{du}{2} = \begin{cases} 0 & k \text{ odd}, \\ 1/(k+1) & k \text{ even}. \end{cases}$$

**Proof.** Both Gram expressions are positive semidefinite on the interval because $1 - u^2 \geq 0$ there. Factoring $F_a(u)$ at a fixed $u$ gives a positive congruence map, whether or not its individual factors depend polynomially on $u$. The intertwining identity therefore proves $B_a(u) C_{\mathcal{J}} \subseteq C_{\mathcal{J}}$. Polynomials on a compact interval are integrable. The mass identity and Proposition 52 prove the assertion. Equality of two finite matrix polynomials is equivalent to equality of every coefficient. The moment values follow by elementary integration. □

For fixed $\mathcal{J}$ and fixed degree bounds, unknown entries of $B_a$ and the Gram matrices obey linear equalities and positive semidefinite constraints. A convex data-fit objective therefore gives a semidefinite feasibility or optimization problem in this fixed representation. Jointly choosing $\mathcal{J}$ changes the problem: the intertwining equation then contains products of unknown coefficients.

This is a sufficient certificate, not a completeness theorem for all positive kernels or all curved domains. A negative answer at one degree is not a proof of model invalidity. Approximate Gram feasibility is also not an exact witness. Rational factorizations with nonnegative rational weights, or rigorous matrix-positivity bounds, must survive a separate replay.

The reference checker accepts a factored version:

$$\Phi_u(X) = \sum_\nu w_\nu g_\nu(u) K_\nu(u) X K_\nu(u)^{\mathsf{T}}, \qquad g_\nu(u) \in \{1, 1-u^2\}.$$

It checks the polynomial identities, weights, support, embedding, initial state and normalization exactly. It does not run a general semidefinite optimizer. The operator identity holds on the image of $\mathcal{J}$; trace preservation of an arbitrary extension outside that image is not required.

## 60. Learning from integral observations rather than point probabilities

A finite polynomial family can be identified through moment operators. This gives a continuous-observation analogue of the core-table construction without pretending that a point has positive probability.

Assume the true operator density belongs to the declared class

$$B_a(u) = \sum_{k=0}^{D} u^k C_{a,k}, \qquad M_{a,j} = \int_{-1}^{1} u^j B_a(u) \, \frac{du}{2}.$$

Define the moment Gram matrix $W_D$ by

$$(W_D)_{jk} = \int_{-1}^{1} u^{j+k} \, \frac{du}{2}, \qquad 0 \le j,k \le D.$$

A history is now a prescribed action sequence with observations constrained to measurable events of positive or zero probability. It produces an unnormalized predictive vector $h_j$. A future event experiment produces a row functional $t_i$. Choose $r$ of each and form

$$H_{ij} = t_i h_j, \qquad (G_{a,j})_{i\ell} = t_i M_{a,j} h_\ell.$$

The index $j$ in $G_{a,j}$ is a moment order. Odd-order entries can be signed expectations. They are not branch probabilities and are never normalized as if they were.

**Proposition 54. Conditional finite-moment identification.** Suppose the true predictive dimension is $r$, the degree bound $D$ is correct, and the selected $r \times r$ core $H$ is invertible. Exact population values of $H$, every $G_{a,j}$ and the initial test vector determine the complete polynomial operator family in test coordinates:

$$\boxed{\widehat{M}_{a,j} = G_{a,j} H^{-1}, \qquad \widehat{C}_{a,k} = \sum_{j=0}^{D} (W_D^{-1})_{kj} \widehat{M}_{a,j}.}$$

If every moment-operator error is at most $\eta$ in one fixed induced norm, then coefficient reconstruction alone obeys

$$\|\widetilde{C}_{a,k} - \widehat{C}_{a,k}\| \le \eta \sum_j |(W_D^{-1})_{kj}|.$$

**Proof.** A nonzero polynomial has strictly positive squared integral, so its coefficient vector $c$ satisfies $c^{\mathsf{T}} W_D c > 0$. Thus $W_D$ is invertible. Let $T$ have rows $t_i$ and $S$ have columns $h_j$. Since $H = TS$ is an invertible square core in an $r$-dimensional space, both $T$ and $S$ are invertible. Therefore $G_{a,j} H^{-1} = T M_{a,j} T^{-1}$. The relation $M_{a,j} = \sum_k (W_D)_{jk} C_{a,k}$ is exact integration of the polynomial expansion. Invert it and conjugate by $T$ to obtain the displayed reconstruction. The error bound is the triangle inequality for that same fixed linear inversion. □

The rank and degree assumptions are not learned from the determinant of $H$. The determinant establishes coverage only after the ambient predictive dimension has been justified. The moment Gram can become poorly conditioned at higher degree. An orthogonal polynomial basis can improve this representation without removing an incorrect degree assumption.

The entries have an operational population meaning. A reset experiment for an event word, with one observation weighted by $u^j$, returns a bounded random variable in $[-1, 1]$. Its expectation is the corresponding entry of $G_{a,j}$. With $N$ independent trials for each of $M$ predeclared entries, the same exponential-moment argument as Proposition 35 gives the simultaneous bound

$$\epsilon_N = \sqrt{\frac{2 \log(2M/\delta)}{N}}$$

with failure probability at most $\delta$. The factor differs from the Bernoulli bound because the value range is two. Dependent trajectories, unknown resets and post-selection are not covered by that observation protocol. The worked example uses exact constructed population tables.

## 61. Finite moments cannot validate an unrestricted density class

A finite observation basis is a modeling assumption. Positivity certification does not turn that assumption into a consequence of data.

**Proposition 55. A valid continuous counterexample to finite-moment identification.** Let

$$L_3(u) = \frac{5u^3 - 3u}{2}, \qquad p_\pm(u) = 1 \pm \epsilon L_3(u), \qquad 0 < \epsilon < 1.$$

Relative to $du/2$, these are distinct strictly positive normalized densities with the same moments of orders zero, one and two. Their total-variation distance is

$$\boxed{\|p_+ \mu - p_- \mu\|_{\mathrm{TV}} = \frac{13\epsilon}{40}.}$$

Taking independent observations gives two rank-one processes whose products of these low-order moments agree at every sequence length.

**Proof.** The endpoints and stationary points of $L_3$ show $|L_3(u)| \le 1$ on the interval. Direct integration gives $\int u^j L_3(u)\,du/2 = 0$ for $j = 0,1,2$. This proves normalization and moment agreement, while strict positivity follows from $\epsilon < 1$. Its nonnegative root other than zero is $\sqrt{3/5}$. Integrating $|L_3|$ on the two subintervals of $[0,1]$ and using symmetry gives $\int |L_3|\,du/2 = 13/40$. The density difference is $2\epsilon L_3$, so its half-integral is the stated distance. Independence makes every product moment factor into the identical one-step moments. □

This is not a defect in Proposition 54. These densities lie outside its quadratic class. It is a concrete reason to retain an explicit approximation or misspecification term when increasing the observation basis. Internal probability validity and correct density identification remain different claims even in one predictive dimension.

## 62. Exact observation bins preserve their own event law

Let $A_1, \dots, A_n$ be a measurable partition of $Y$. Define the cell operators

$$B_{a,j} = \int_{A_j} B_a(y)\,\mu(dy).$$

Their role is to represent observed events. They are not point samples of the original density.

**Proposition 56. Exact bin coarsening and its posterior.** Every $B_{a,j}$ preserves the probability cone, and $\lambda \sum_j B_{a,j} = \lambda$. For a policy measurable with respect to bin histories, the finite model $\{B_{a,j}\}$ reproduces exactly the binned observation law of the continuous model at every finite horizon. On an observed cell with $p_j = \lambda B_{a,j} z > 0$,

$$\frac{B_{a,j} z}{p_j} = \frac{1}{p_j} \int_{A_j} z^+(y) \lambda B_a(y) z\,\mu(dy).$$

**Proof.** Integrals of vectors in a closed convex cone remain in the cone. Summing over a partition gives the original mass identity. For a fixed bin history, its policy factors are constant across the corresponding cells. Successive integration of the operator product therefore gives the product of the cell operators, in the original temporal order. This establishes every finite binned record probability. In the posterior identity, substitute $z^+(y)\lambda B_a(y) z = B_a(y) z$ on the positive-density set. The remaining outputs are zero because they lie in the cone and have zero mass. □

The coarsened posterior is a conditional average, not the posterior at a representative observation value. It preserves exactly the information retained by the event history. An arbitrary policy that reacts differently to two observations inside the same cell is not bin-measurable, so the preceding exact-policy conclusion does not apply to it.

Cell refinement is compatible with this semantics: a parent operator is the sum of its children. In this setting those cell operators are positive probability branches. This positivity is an extra property of event integration, not a property of every signed spectral packet in the signal archive.

## 63. Averaging transformations has a covariance defect

A common discretization replaces a family of matrices $K(y)$ by its mean and then applies $X \mapsto KXK^{\mathsf{T}}$. This is generally the wrong mean operation.

**Proposition 57. Exact covariance correction for an averaged congruence.** Let $\nu$ be a probability measure and suppose $K(y)$ has square-integrable entries. Put $\bar{K} = \int K(y)\,\nu(dy)$. For every symmetric $X$,

$$\boxed{\int K(y)XK(y)^{\mathsf{T}}\,\nu(dy) = \bar{K}X\bar{K}^{\mathsf{T}} + \int (K(y)-\bar{K})X(K(y)-\bar{K})^{\mathsf{T}}\,\nu(dy).}$$

For positive semidefinite $X$, so is the correction. For $K(u) = K_0 + uK_1$ with scalar $u$ of finite second moment, it is $\mathrm{Var}_\nu(u)K_1XK_1{}^{\mathsf{T}}$. A uniform interval of width $h$ has variance $h^2/12$.

**Proof.** Substitute $K = \bar{K} + (K - \bar{K})$ and expand. The two cross terms integrate to zero. The remaining integrand is a positive congruence when $X$ is positive. For an affine matrix function the centered term is $(u - \mathbb{E}_\nu u)K_1$, giving the variance formula. Integrating the square of a centered uniform variable gives $h^2/12$. □

This is a defect between two operations: averaging a transformation parameter and applying its quadratic action. It is not the same as the residual from dropping a signal packet, although both fit the framework's requirement to record failures of commutation.

For the example in Section 66, equal bins with midpoint evaluation lose total mass

$$\boxed{\frac{3}{4n^2}}$$

when there are $n$ bins. With one bin, midpoint evaluation retains only one quarter of the required mass. Exact integration or the covariance correction restores the missing contribution. Dividing an inaccurate approximation by a guessed normalizer would define another model rather than establish this identity.

## 64. A valid density approximation with a resolution bound

Exact coarsening gives a finite event model. To compare complete continuous observation laws, lift that model back to a density on the same observation space. Assume all cells have positive reference mass and put

$$B_a^{\Pi}(y) = \frac{B_{a,j}}{\mu(A_j)} \quad (y \in A_j).$$

This uses the normalized reference measure inside each cell. It is an explicit density approximation, not an assertion that the original conditional density was constant there.

**Proposition 58. Positive cell-average approximation.** The family $B_a^{\Pi}$ is cone preserving and mass conserving. It has exactly the same cell operators as $B_a$. Let $Y$ have a metric and suppose

$$\|B_a(y) - B_a(v)\|_{C\to C} \le L\, d(y,v)$$

in the induced base norm, with every cell of diameter at most $h$. Then

$$\epsilon_{\Pi} := \sup_{a,z\in K} \int \|(B_a(y) - B_a^{\Pi}(y))z\|_C \, \mu(dy) \le Lh.$$

For equal intervals of width $h$ under the uniform measure on $[-1,1]$, the sharper bound is $\epsilon_{\Pi} \le Lh/3$.

**Proof.** Each cell average is a positive scalar multiple of a positive cell operator. Its integral over that cell is $B_{a,j}$, proving both normalization and exact cell agreement. For $y \in A_j$, subtract the cell average and use

$$\|(B_a(y) - B_a^{\Pi}(y))z\|_C \le \frac{L}{\mu(A_j)} \int_{A_j} d(y,v)\, \mu(dv),$$

because $\|z\|_C = 1$. Integrating over $y$ proves the diameter bound. Two independent uniform points in an interval of width $h$ have mean absolute separation $h/3$, obtained by integrating over its two triangles. Applying this inside every equal interval gives the sharper bound. □

The Lipschitz constant is a declared operator bound. It is not a fit residual inferred from adjacent samples. The comparison needs a common domain and a common reference measure. If two kernels are only close in Euclidean coefficient norm, the relevant norm-conversion constant must be included.

In the full $2 \times 2$ symmetric-matrix example, the pullback base norm is

$$\|(m,x,y)\|_C = \max\{|m|, \sqrt{x^2+y^2}\} = \|\mathcal{J}(m,x,y)\|_1.$$

There $\|K(u)\|_2 \le 2$ and $\|K(u) - K(v)\|_2 \le (3/2)|u-v|$. The identity

$$K(u)XK(u)^{\mathsf{T}} - K(v)XK(v)^{\mathsf{T}} = (K(u)-K(v))XK(u)^{\mathsf{T}} + K(v)X(K(u)-K(v))^{\mathsf{T}}$$

and $\|AXB\|_1 \le \|A\|_2\|X\|_1\|B\|_2$ give $L = 6$. The rotated action uses an orthogonal matrix before this map and has the same bound. Equal bins therefore give $\epsilon_{\Pi} \le 4/n$, capped by two for two valid instruments. The inequality is sufficient, not asserted sharp.

## 65. Full continuous record laws inherit the stability argument

The base-norm proof from Chapter 4 has an integral version. Its hypotheses must be supplied rather than inferred by replacing a finite sum with an integral in a program.

**Proposition 59. Continuous-record total-variation bound.** Let $B_a(y)$ and $\widetilde{B}_a(y)$ satisfy Proposition 52 on a common $(C, \lambda, \mu)$, and let the two models use the same measurable policy on every common observation history. Suppose their initial states satisfy $\|z_0 - \widetilde{z}_0\|_C \le e_0$ and their time-$t$ instruments satisfy

$$\sup_{a,z\in K} \int \|(B_{a,t}(y) - \widetilde{B}_{a,t}(y))z\|_C \, \mu(dy) \le \epsilon_t.$$

The length-$T$ record laws, including actions when the policy is randomized, obey

$$\boxed{\|P_T - \widetilde{P}_T\|_{\mathrm{TV}} \le \min\left\{1, \frac{e_0 + \sum_{t=0}^{T-1} \epsilon_t}{2}\right\}.}$$

No lower bound on each observation density is required. For identical initial states and the equal-bin approximation in Section 64, the example has the sufficient bound $\min(1, 2T/n)$.

**Proof.** For any positive decomposition $x = u - v$, cone preservation and mass conservation give

$$\int \|B_a(y)x\|_C \, \mu(dy) \le \int \lambda B_a(y)(u+v) \, \mu(dy) = \lambda u + \lambda v.$$

Taking the infimum proves integrated instrument contraction. At each continuous history let $v_h, \widetilde{v}_h$ be the unnormalized state densities, including the policy factors. Split the next difference as

$$B_a(y)(v_h - \widetilde{v}_h) + (B_a(y) - \widetilde{B}_a(y))\widetilde{v}_h.$$

Integrate the first term using contraction. The second is bounded by $\epsilon_t \lambda \widetilde{v}_h$ after normalizing only its nonzero positive mass. Its integral over histories is at most $\epsilon_t$. Tonelli's theorem applies to these nonnegative norm bounds, and operator integrability supplies well-defined finite-dimensional integrals. The integrated state discrepancy therefore satisfies $D_{t+1} \le D_t + \epsilon_t$. Finally $|\lambda x| \le \|x\|_C$ bounds the record-density difference, and total variation is half its absolute integral. □

This result compares one common policy, not two policies optimized separately. It concerns the whole law and can become uninformative at long horizons. It is not a uniform posterior contraction theorem.

For a measurable event $H$ of probability $p = P_T(H) > 0$, the restriction-and-normalization argument of Proposition 44 works for measures as well as finite vectors. When both conditional laws exist, their total-variation distance is at most $\min(1, \delta/p)$ if the full-law distance is at most $\delta$. The event probability is invariant under a reference-measure change. A density value at one point is not a substitute for that event probability.

The full-law comparison between $B$ and $B^{\Pi}$ permits any common measurable policy because both are continuous-output models on the same space. The exact equivalence to the finite bin

model in Proposition 56 instead requires a bin-measurable policy. These are different conclusions with different observation information.

## 66. A complete continuous-observation example

Use the same three-coordinate matrix embedding as Section 57. Set

$$K(u) = \frac{1}{2}\begin{pmatrix} 1+3u & 0 \\ 0 & 1-3u \end{pmatrix}, \qquad -1 \le u \le 1, \qquad \mu(du) = du/2.$$

The direct action uses $X \mapsto K(u)XK(u)^{\mathsf{T}}$. The rotated action first applies the unchanged rotation congruence from Section 57, giving

$$X \longmapsto \frac{1}{5}K(u)WXW^{\mathsf{T}}K(u)^{\mathsf{T}}.$$

Both are exactly normalized because $\int K(u)^{\mathsf{T}}K(u)\,du/2 = I$ and $W^{\mathsf{T}}W = 5I$. In predictive coordinates, write

$$a(u) = \frac{1+9u^2}{4}, \qquad b(u) = \frac{3u}{2}, \qquad c(u) = \frac{1-9u^2}{4},$$

$$B_0(u) = \begin{pmatrix} a(u) & b(u) & 0 \\ b(u) & a(u) & 0 \\ 0 & 0 & c(u) \end{pmatrix}, \qquad B_1(u) = B_0(u)T.$$

Although $c(u)$ is negative for part of the interval, both actions preserve the disk cone. Entrywise sign is not the positivity criterion in these coordinates. The initial state is $z_0 = (1,0,0)^{\mathsf{T}}$, or the matrix $I/2$.

### 66.1 Event tables and moment reconstruction

Let the three future tests be the empty test, a nonnegative observation after action zero, and a nonnegative observation after action one. Their rows form

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 3/8 & 0 \\ 1/2 & 9/40 & -3/10 \end{pmatrix}.$$

Use histories consisting of the empty history, a nonnegative observation under action zero, and that history followed by a nonnegative observation under action one. Their unnormalized columns are

$$S = \begin{pmatrix} 1 & 1/2 & 107/320 \\ 0 & 3/8 & 3/10 \\ 0 & 0 & -3/40 \end{pmatrix}.$$

The learner is supplied only the resulting integral tables and the declared class. Its core is

$$\boxed{H = \begin{pmatrix} 1 & 1/2 & 107/320 \\ 1/2 & 25/64 & 179/640 \\ 1/2 & 107/320 & 823/3200 \end{pmatrix}, \qquad \det H = \frac{81}{25600}.}$$

The core is not required to be symmetric. Its rows are tests and its columns are histories, which are different experiments. The three moment orders use

$$W_2 = \begin{pmatrix} 1 & 0 & 1/3 \\ 0 & 1/3 & 0 \\ 1/3 & 0 & 1/5 \end{pmatrix}, \qquad W_2^{-1} = \begin{pmatrix} 9/4 & 0 & -15/4 \\ 0 & 3 & 0 \\ -15/4 & 0 & 45/4 \end{pmatrix}, \qquad \det W_2 = \frac{4}{135}.$$

Proposition 54 reconstructs $QB_a(u)Q^{-1}$ from $H$ and the six inserted-moment tables. The code that performs this reconstruction imports no generating matrices. The separate generator produces the tables and is used only in the verification suite. These are population values from an exact example, not learned sample estimates.

### 66.2 Two densities and a genuine event probability

For a direct observation value $u = 1/2$, the density and updated state are

$$p_1 = \frac{13}{16}, \qquad z_1 = (1, 12/13, 0)^{\mathsf{T}}.$$

If the rotated action next gives $u = -1/2$, they are

$$p_2 = \frac{413}{1040}, \qquad z_2 = (1, -312/413, -240/413)^{\mathsf{T}}.$$

Their product $413/1280$ is a joint density with respect to $(du/2)^{\otimes 2}$, not the probability of those two exact values. The two point values each have zero probability in this nonatomic model.

By contrast, the event $u_1 \geq 0$ under action zero followed by $u_2 \geq 0$ under action one has actual probability

$$\boxed{\frac{107}{320},}$$

and its conditional state is $(1, 96/107, -24/107)^{\mathsf{T}}$. Exact bin operators produce this value without quadrature.

The final point-conditioned state remains strictly inside the disk:

$$\det \mathcal{J}(z_2) = \frac{15625}{682276} > 0.$$

A pure boundary state can instead have zero density at isolated observations. The checker rejects conditioning on such a zero-density branch. It does not manufacture a posterior by flooring its denominator.

### 66.3 The missing covariance is exactly calculable

On an interval with midpoint $v$ and width $h$, the average matrix instrument is

$$\mathbb{E}[K(u)XK(u)^{\mathsf{T}}] = K(v)XK(v)^{\mathsf{T}} + \frac{h^2}{12} K_1 X K_1^{\mathsf{T}}, \qquad K_1 = \frac{3}{2}\,\mathrm{diag}(1, -1).$$

The correction is itself a positive congruence. The cell operator multiplies this mean by the cell mass $h/2$. Equal bins have width $2/n$, so omitting the covariance loses trace $3/(4n^2)$ over the full partition. The reference calculation checks both the exact matrix identity and this mass defect for several rational partitions.

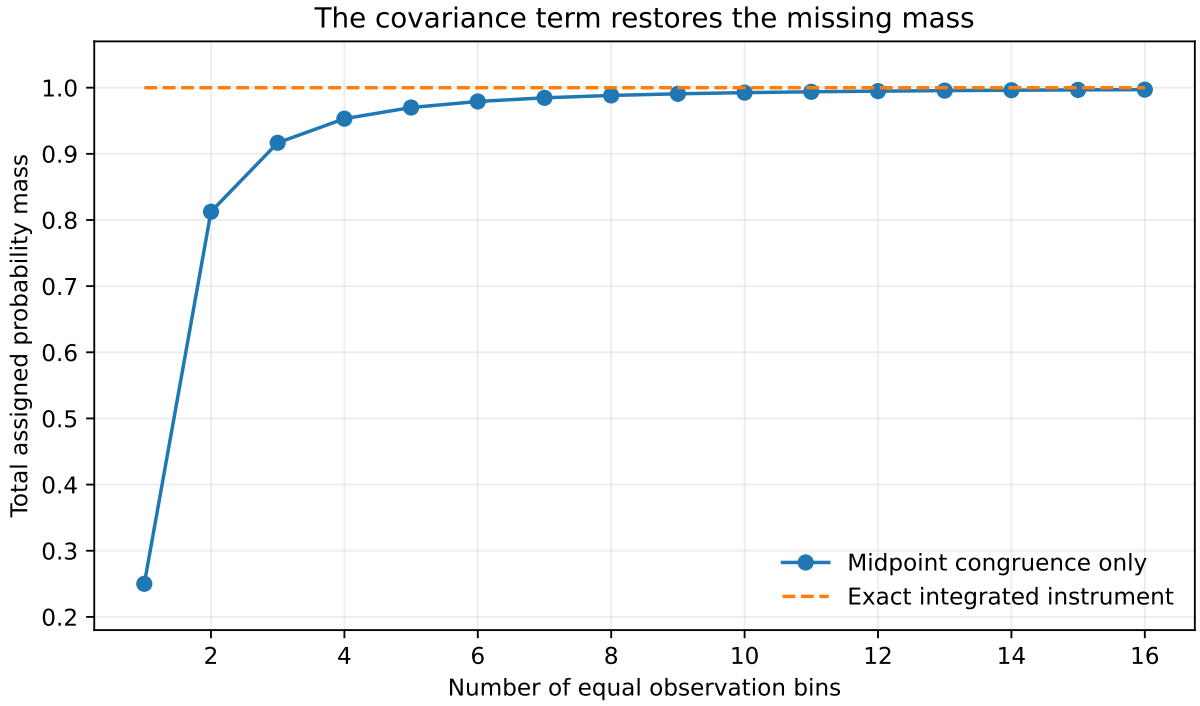



**Figure 5.2:** Midpoint congruences omit a covariance contribution. For this worked model, using n equal observation bins loses mass $3/(4n^2)$. Exact integration retains the missing term and preserves normalization. At one bin the naive approximation retains only one quarter of total mass. Event coarsening is exact for the stated bin-based policy. A policy that responds to values within a bin is a different comparison.

## 67. A declared template can be recognized without being rediscovered

General curved-domain learning remains difficult. There is a useful finite case between supplying a finished coordinate certificate and claiming unrestricted domain discovery.

Suppose a structural library supplies a certified template with polynomial coefficients $C_{a,k}$, mass row $\lambda$ and embedding $\mathcal{J}$. Suppose observed population tables produce coefficients $\widehat{C}_{a,k}$ and mass row $\widehat{\lambda}$ of the same dimension.

**Proposition 60. Exact recognition of a fixed certified template.** The equations

$$\boxed{\widehat{C}_{a,k}T = TC_{a,k}, \qquad \widehat{\lambda}T = \lambda}$$

are linear in $T$. Every invertible solution transports the template certificate to the fitted coordinates by $\widehat{\mathcal{J}} = \mathcal{J}T^{-1}$. If the fitted coefficient family is observable, meaning the rows $\widehat{\lambda}\widehat{C}_w$ span the dual space, the solution is unique. The fitted initial state must still be tested in the transported normalized cone.

**Proof.** All coefficient matrices are fixed, so the displayed constraints are linear. An invertible solution yields $\widehat{B}_a(u) = TB_a(u)T^{-1}$ coefficient by coefficient. Substitution in the template intertwining identity proves the transported certificate. To prove uniqueness, let $D$ be the difference of two solutions. It satisfies $\widehat{C}_{a,k}D = DC_{a,k}$ and $\widehat{\lambda}D = 0$. Therefore $\widehat{\lambda}\widehat{C}_wD = 0$ for every coefficient word. Observability makes every column of $D$ zero. An initial state outside the transported cone is not repaired by this operator identity, which explains the separate check. □

For the worked example, the linear system has the unique solution $T = Q$ from Section 66. The learner receives the population tables and a fixed structural template. It reconstructs the operators, solves the linear match, transports the embedding and exports a certificate. Independent replay then needs only that certificate. It does not refit, search, or import the template or generator.

This is template recognition, not unrestricted geometry discovery. The supplied template includes the specific operator family. An incompatible model is rejected rather than changed to fit the template. The reference routine also withholds a result for an underdetermined or singular match. Noise would require a declared approximate match and a separate validity or correction argument. The recognition example is exact.

For fixed embeddings, polynomial Gram optimization offers a broader construction route than fixed-template matching. That convex proposal problem is stated in Section 59 but is not implemented as a general solver here. The independent checker remains useful regardless of whether a proposal came from a template, symbolic derivation or an optimization routine.

## 68. The combined continuous polynomial sector

**Proposition 61. Conditional continuous realization with a finite domain certificate.** Assume a controlled continuous-observation process has the declared finite predictive dimension and polynomial operator-density degree, a fixed uniform reference measure on $[-1, 1]$, and an invertible complete event-test core. Exact moment tables identify its operator density in test coordinates by Proposition 54.

If the fitted family and initial state admit a checked matrix/Gram certificate of Proposition 53, the supplied realization is internally valid at every observation value and finite horizon. A successful invertible fixed-template recognition is one sufficient way to construct such a certificate. Exact coordinate transport and complete packet synthesis preserve the represented event probabilities. Individual signed state packets still need the separate positivity condition before probabilistic interpretation.

Exact observation-cell integration gives the precise finite event model for bin-measurable policies. Positive cell-average density approximations have the resolution bound of Proposition 58. Two certified models on a common cone and common reference measure satisfy Proposition 59's full-law bound whenever the declared instrument discrepancy is controlled.

**Proof.** Moment identification uses the true dimension and degree hypotheses to apply Proposition 54. The independent Gram identity establishes positivity on the whole declared interval, while the exact integral mass identity supplies normalization through Proposition 52. Proposition 60 supplies a transported witness in the template case. Coordinate cancellation and full packet synthesis are exact identities. Proposition 56 proves the cell-model statement. Propositions 58 and 59 provide the two approximation conclusions with their common-domain and common-policy assumptions unchanged. □

For a noisy fitted model, the certificate alone proves only internal validity. A guarantee about the generating process requires a valid uncertainty set covering it. A common-domain certificate over a convex uncertainty family can be checked at complete-family vertices with the same embedding, because the operator identities, Gram positivity and mass constraints survive convex combination. This does not justify an uncertainty set that excludes the true degree, rank or dynamics.

The construction leaves visible what is often collapsed into the phrase signal model: a represented state, an operator-valued observation density, a reference measure, a valid probability domain, an observation protocol and an approximation budget. Each is needed for a different reason.

## 69. From continuous semantics to discovery

Chapter 5 gives event-based probability semantics, exact binning and conditional finite-moment reconstruction. Fixed-template recognition in Proposition 60 is not unrestricted geometry discovery. Chapters 6–7 add restricted discovery and robust verification with their own assumptions. The continuous reference measure and the observation basis remain explicit inputs.

CHAPTER 6

# Certificate discovery, approximation and sequential evidence

A signal model must distinguish a representation that preserves probability from a representation that predicts the observed process accurately. Chapter 5 gave an exact curved certificate and a way to recognize a supplied operator template. It also made the polynomial-degree assumption explicit. This chapter advances both boundaries while retaining the distinction between them.

The first development is a restricted discovery procedure. It recovers an ellipsoidal probability domain from fitted polynomial operators, without a supplied template, matrix embedding or hidden generator. The second is a positive approximation theorem for operator densities outside a finite polynomial class. The third is sequential evidence under dependent observations, including the cost of action coverage. Together they separate structural validity, approximation error and observable disagreement.

The invariant-form and quadratic-certificate ingredients belong to classical systems theory [R8, R13]. Positive polynomial averaging uses the classical Bernstein–Durrmeyer kernel [R19]. The sequential argument is an elementary martingale confidence sequence, not a new concentration principle [R20]. Their connection to the large signal model is through explicit representation contracts and independently replayable examples. Propositions 62–70 have written proofs below. Their formal status is recorded in the appendices.

## 70. An ellipsoidal certificate with an orientation test

Work in mass-first coordinates $z = (m, x) \in \mathbb{R} \times \mathbb{R}^d$, with $\lambda = (1, 0, \dots, 0)$. For a center $c \in \mathbb{R}^d$ and a symmetric positive definite matrix $H$, define

$$C_{c,H} = \{(m, x) : m \geq 0,\ (x - mc)^{\mathsf{T}} H (x - mc) \leq m^2\}.$$

Its normalized base is an ellipsoid. Set

$$S_c = \begin{pmatrix} 1 & 0 \\ -c & I \end{pmatrix}, \qquad Q = S_c^{\mathsf{T}} \begin{pmatrix} 1 & 0 \\ 0 & -H \end{pmatrix} S_c.$$

Then $C_{c,H} = \{z : m \geq 0,\ z^{\mathsf{T}} Q z \geq 0\}$. The cone is proper and generating, and $m$ is strictly positive on every nonzero cone element. These facts follow by an invertible linear transformation to the Lorentz cone. In particular, an output with zero mass in the cone is the zero vector.

Let $B_a(u)$ be polynomial operator densities on $[-1,1]$, relative to $du/2$. Write

$$\lambda B_a(u) = (b_{a,0}(u), b_a(u)^{\mathsf{T}}), \qquad \beta_a(u) = b_{a,0}(u) + b_a(u)^{\mathsf{T}} c.$$

**Proposition 62. Oriented quadratic probability certificate.** Suppose that for each action and every $u \in [-1,1]$ there are $\tau_a(u) \geq 0$ and $R_a(u) \succeq 0$ satisfying

$$\boxed{B_a(u)^{\mathsf{T}} Q B_a(u) - \tau_a(u) Q = R_a(u),}$$

and that

$$\boxed{\beta_a(u) \geq 0, \qquad \beta_a(u)^2 - b_a(u)^{\mathsf{T}} H^{-1} b_a(u) \geq 0.}$$

If $z_0 \in C_{c,H}$ has mass one and $\lambda \int B_a(u)\, du/2 = \lambda$, then every operator preserves $C_{c,H}$. The supplied model has valid probabilities at every finite horizon. Polynomial sum-of-squares witnesses for the scalar inequalities and for $R_a$ make this a finite certificate.

**Proof.** For $z \in C_{c,H}$, the quadratic identity gives

$$(B_a z)^{\mathsf{T}} Q (B_a z) = \tau_a z^{\mathsf{T}} Q z + z^{\mathsf{T}} R_a z \geq 0.$$

For a nonzero $z$, write $x = m(c + w)$ with $w^{\mathsf{T}} H w \leq 1$. Its output mass is $m(\beta_a + b_a^{\mathsf{T}} w)$. Cauchy–Schwarz in the $H$ inner product bounds this below by

$$m\left(\beta_a - \sqrt{b_a^{\mathsf{T}} H^{-1} b_a}\right) \geq 0.$$

Thus both inequalities defining the desired sheet of the quadratic cone hold. The zero input causes no difficulty. Integrated mass conservation gives normalization, and induction gives the finite-horizon conclusion. For polynomial data, identities between coefficients and nonnegative sums of polynomial squares establish the displayed hypotheses throughout the interval. Terms multiplied by $1 - u^2$ are also nonnegative on that interval. □

The orientation test is essential. The matrix $-I$ preserves every quadratic form, but maps positive mass to negative mass. A quadratic inequality without its mass orientation can certify the wrong sheet. Likewise, testing finitely many values of $u$ is not a substitute for a polynomial identity and its positivity witness.

This is a sufficient language, not a completeness theorem for all cones. The reference checker accepts rational scalar and matrix sum-of-squares witnesses. Its discovery routine uses the narrower conformal case $R_a = 0$, but the checker does not confuse that search restriction with the theorem's scope.

## 71. Geometry can be proposed from the operators themselves

An operator identity can reveal a useful coordinate geometry without specifying that geometry in advance. Suppose two fitted operators evaluated at an anchor $u_*$ give an invertible ratio

$$F = B_b(u_*)B_a(u_*)^{-1}, \qquad \lambda F = \lambda.$$

This ratio is an algebraic object. It need not be a physically available action or a positive map by itself. In mass-first coordinates write

$$F = \begin{pmatrix} 1 & 0 \\ b & A \end{pmatrix}.$$

If $I - A$ is invertible, its fixed center is $c = (I - A)^{-1}b$. The centered spatial action is $A$. A candidate invariant metric obeys

$$A^{\mathsf{T}} H A = H.$$

With $A$ fixed, this is a linear equation in the independent entries of symmetric $H$. No coordinate template appears in that equation.

**Proposition 63. Conditional operator-only ellipsoid discovery.** Suppose the symmetric solution space of $A^{\mathsf{T}} H A = H$ is one-dimensional and contains a positive definite matrix $H_0$. Let a rank-one fitted branch at a declared observation have a nonzero-mass image, normalized to $v = (1, x_v)$. If $x_v \neq c$, define

$$H = \frac{H_0}{(x_v - c)^{\mathsf{T}} H_0 (x_v - c)}.$$

This gives a unique positive definite metric in that solution space for which $v$ is on the normalized boundary. If this proposed geometry passes every hypothesis of Proposition 62 for the complete fitted family and its initial state, it is an internally valid probability-domain certificate. The proposal requires no supplied template. Failure at any step withholds this certificate and does not establish model invalidity.

**Proof.** Invertibility of $I - A$ gives the unique center. One-dimensionality makes every symmetric invariant form a scalar multiple of $H_0$. The nonzero displacement has strictly positive squared $H_0$ length, so the normalization is well-defined, positive and independent of the initially chosen positive scale. The final acceptance test is exactly Proposition 62. Neither invariance under the ratio nor the choice of a boundary image alone is asserted to certify the full family. □

The rank-one image proposes a scale. It does not prove that every possible probability domain must place that image on its boundary. Some valid models have many admissible domains. The procedure can also fail because there is no suitable ratio, because invariant forms are not unique, because no rational rank-one anchor is found, or because the chosen ellipsoid fails another branch. These are explicit search outcomes.

### 71.1 The continuous example no longer needs its template

Use the same exact population moment tables as Section 66. The existing moment learner first reconstructs the two polynomial operators in test coordinates. The new discovery routine receives those operators, their mass row and initial vector. It does not read `continuous_template.json` or import the population generator.

At $u_* = 0$, the operator ratio gives the required invariant-form equation. A rational rank-one observation supplies a boundary image. The recovered center and metric are

$$\boxed{c = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}, \qquad H = \begin{pmatrix} 100/9 & -20/3 \\ -20/3 & 100/9 \end{pmatrix}.}$$

The two metric eigenvalues are $40/9$ and $160/9$. For both actions the exact certificate is

$$\begin{aligned} B_a(u)^{\mathsf{T}} Q B_a(u) &= \frac{(1-9u^2)^2}{16} Q, \\ \beta_a(u) &= \frac{1+9u^2}{4}, \\ \beta_a(u)^2 - b_a(u)^{\mathsf{T}} H^{-1} b_a(u) &= \frac{(1-9u^2)^2}{16}. \end{aligned}$$

Every scalar on the right has an exact rational sum-of-squares representation. Integrated mass is unchanged. The fitted initial state is the ellipsoid center, and the two densities $13/16$ and $413/1040$ from Section 66 are unchanged. This is a new certificate construction for the same model, not a change to its predictions.

The exported JSON contains the fitted operators, center, metric, initial state and all witnesses. Independent replay needs only that JSON. The discovery calculation is also checked on three additional rational coefficient families and affine coordinate changes. Those finite examples exercise the stated sector and do not establish unrestricted discovery.

The procedure currently uses three predictive coordinates, a two-dimensional ellipsoid, a finite rational anchor list, a one-dimensional invariant-form space and rational rank-one roots. It proposes only exactly conformal polynomial branches whose scalar nonnegativity can be resolved by its factor-based sum-of-squares routine. General noisy or jointly optimized curved-domain learning remains outside this implementation.

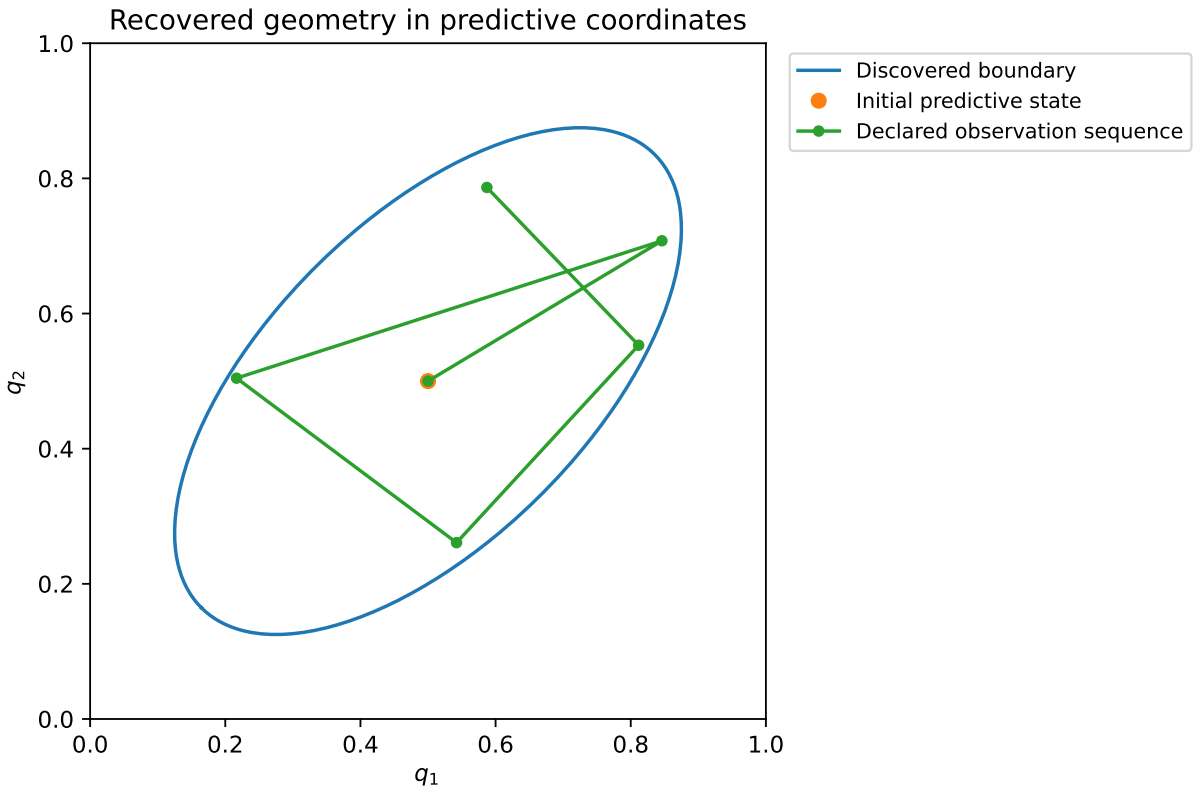



**Figure 6.1:** The exact example constructs an ellipse from observable operators, rather than taking its metric from a supplied template. The trajectory remains in the discovered domain. The accepted witness independently checks every operator, orientation and initial state. This is restricted exact invariant-form discovery, not general discovery of hidden state or noisy geometry from arbitrary observations.

## 72. Approximate an unknown density without assuming its exact degree

Chapter 5 identified a polynomial operator density under an exact degree assumption. A different question is how to approximate a valid density that is not a polynomial at all.

In this section observations lie in $[0,1]$ with Lebesgue reference measure. This is an explicit change from the preceding $[-1,1]$ convention. Under $v = (u+1)/2$, the probability measure $du/2$ becomes $dv$. A Lipschitz constant with respect to $u$ becomes twice as large with respect to $v$.

Write the Bernstein basis and the classical Bernstein–Durrmeyer kernel as

$$b_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k}, \qquad D_n(x,v) = (n+1)\sum_{k=0}^{n} b_{n,k}(x) b_{n,k}(v).$$

Both variables range over $[0,1]$. The kernel is nonnegative and integrates to one in either variable. Let $B_a(v)$ be any operator-norm integrable density preserving a common probability cone $C$, with mass functional $\lambda$ and normalized base $K$. Define

$$B_a^{[n]}(x) = \int_0^1 D_n(x,v) B_a(v)\, dv = \sum_{k=0}^{n} b_{n,k}(x) C_{a,n,k},$$

where

$$C_{a,n,k} = (n+1)\int_0^1 b_{n,k}(v) B_a(v)\, dv.$$

**Proposition 64. Valid positive polynomial approximation.** Every $B_a^{[n]}$ preserves $C$, has degree at most $n$, and satisfies exact integrated mass conservation. If

$$\|B_a(v) - B_a(x)\|_{C\to C} \le L|v-x|$$

for all actions, then

$$\boxed{\sup_{a,z\in K}\int_0^1 \|(B_a^{[n]}(x) - B_a(x))z\|_C\, dx \le \frac{L}{\sqrt{3(n+2)}}.}$$

With identical initial states, the corresponding length-$T$ laws under a common policy differ in total variation by at most

$$\boxed{\min\left(1, \frac{TL}{2\sqrt{3(n+2)}}\right).}$$

No finite true-degree hypothesis is needed.

**Proof.** Nonnegative integration preserves the closed convex cone. Since $\int_0^1 b_{n,k} = 1/(n+1)$ and $\sum_k b_{n,k} = 1$, the kernel is doubly stochastic. Its integral in $x$ gives $\int B_a^{[n]} = \int B_a$, proving exact mass conservation.

For the error, subtract $B_a(x)$ inside the kernel integral, apply the Lipschitz inequality and use $\|z\|_C = 1$. The integrated discrepancy is at most $L\mathbb{E}|V - X|$ under the joint probability density $D_n(x, v)$. This joint law can be constructed by first choosing $J$ uniformly from $\{0, \dots, n\}$, and then drawing $X, V$ independently, conditionally on $J = k$, with density $(n+1)b_{n,k}$.

That conditional beta law has variance

$$\operatorname{Var}(X \mid J = k) = \frac{(k+1)(n-k+1)}{(n+2)^2(n+3)}.$$

The average of $(k+1)(n-k+1)$ over $k = 0, \dots, n$ is $(n+2)(n+3)/6$, obtained by summing the first two integer powers. Hence

$$\mathbb{E}(V - X)^2 = 2\mathbb{E}\operatorname{Var}(X \mid J) = \frac{1}{3(n+2)}.$$

Cauchy–Schwarz gives the claimed integrated bound. The whole-law conclusion is Proposition 59 applied to the two valid instruments on the same domain. □

This construction approximates the operator density itself. It does not average a congruence factor and omit its covariance. Its coefficients are positive operators even when their entries in predictive coordinates are signed.

Operationally, it replaces the raw observation $v$ by a noisy observation $x$ drawn from $D_n(x, v)$. It preserves total mass and underlying unconditioned evolution, not the entire original observation law. That changed law is what the bound measures. A prediction policy must be compared as the same function of the reported history.

### 72.1 A nonpolynomial density and a failed ordinary truncation

Consider the valid scalar density

$$p(v) = \frac{1}{2} + 2\left|v - \frac{1}{2}\right|.$$

It integrates to one, takes values between $1/2$ and $3/2$, and has Lipschitz constant two. It is not a single polynomial on $[0, 1]$. Proposition 64 nevertheless supplies normalized positive polynomial approximations at every degree, with an explicit approximation budget.

An ordinary orthogonal truncation need not preserve that validity. The valid density $p(v) = 3v^2$ has degree-one $L^2$ projection

$$q(v) = 3v - \frac{1}{2},$$

which is negative near zero. The positive construction at degree one instead gives

$$p^{[1]}(v) = v + \frac{1}{2}.$$

The two approximations solve different mathematical problems. Neither has permission to call its approximation error zero. The reference checks derive both expressions exactly.

## 73. Keep approximation bias separate from coefficient error

The positive approximation theorem assumes access to exact integral coefficients. A fitted model introduces another discrepancy. Its coefficient estimates need their own validity constraints.

**Proposition 65. Certified coefficient error and approximation bias.** Suppose every fitted coefficient $\widetilde{C}_{a,n,k}$ preserves the same cone $C$, and

$$\lambda \frac{1}{n+1} \sum_{k=0}^{n} \widetilde{C}_{a,n,k} = \lambda.$$

Then $\widetilde{B}_a^{[n]}(x) = \sum_k b_{n,k}(x)\widetilde{C}_{a,n,k}$ is internally valid. If

$$\sup_{z \in K} \|(C_{a,n,k} - \widetilde{C}_{a,n,k})z\|_C \le \eta_{a,k},$$

its integrated discrepancy from the original Lipschitz density is at most

$$\boxed{\frac{L}{\sqrt{3(n+2)}} + \max_a \frac{1}{n+1} \sum_{k=0}^{n} \eta_{a,k}.}$$

A known additional model-correction discrepancy is added to this budget, not absorbed into the approximation term.

**Proof.** Nonnegative Bernstein weights preserve cone positivity, and their integrals prove normalization. For any $z \in K$, the triangle inequality gives

$$\int \| \textstyle\sum_k b_{n,k}(x)(C_{a,n,k} - \widetilde{C}_{a,n,k})z\|_C \, dx \le \frac{1}{n+1} \sum_k \eta_{a,k}.$$

Add the approximation error from Proposition 64. A separate correction is handled by another application of the triangle inequality. □

This is a conditional approximation-and-estimation theorem, not a claim that more coefficients always improve a finite-data estimator. Increasing $n$ reduces the stated bias bound while changing the number and conditioning of the quantities to be estimated. A cone witness proves that the resulting object is a probability model. It does not prove that the coefficient error budgets cover the generating process.

The moments needed for $C_{a,n,k}$ are bounded Bernstein-weighted experiments. Their interpretation as full operator coefficients still requires suitable predictive-state coverage or a justified observable core. The next section provides sequential evidence for conditional experiments, not an automatic replacement for those identification assumptions.

## 74. Evidence from dependent observations

Let $\mathcal{F}_{t-1}$ denote the available information immediately before observing the next measurement, including the chosen action, current model and any test selected using past data. A predictable test is a measurable function $\phi_{t,j}$ of the upcoming measurement, chosen from this information, with values in $[0,1]$. Put

$$X_{t,j} = \phi_{t,j}(Y_t), \quad m_{t,j} = \mathbb{E}[X_{t,j} \mid \mathcal{F}_{t-1}], \quad p_{t,j} = \text{the model's prediction for that test.}$$

The model prediction is also predictable. It may change as the model learns. The true conditional mean $m_{t,j}$ may depend on the entire past. No independence, stationarity or known reset state is assumed in this paragraph.

Choose budgets $\alpha_j > 0$ for a finite or countable prospectively specified family of test rules, with $\sum_j \alpha_j \le \delta$. Define

$$R_{t,j} = \sum_{s=1}^{t} (X_{s,j} - p_{s,j}), \qquad D_{t,j} = \sum_{s=1}^{t} (m_{s,j} - p_{s,j}).$$

**Proposition 66. Time-uniform predictable-residual certificate.** With probability at least $1-\delta$, for every time $t \ge 1$ and every registered rule $j$,

$$\boxed{|R_{t,j} - D_{t,j}| \le b_{t,j}, \qquad b_{t,j} = \sqrt{\frac{t}{2} \log \frac{2t(t+1)}{\alpha_j}}.}$$

For variables taking values in an interval of width $c$, multiply the boundary by $c$. The simultaneous guarantee remains valid when the observation time is chosen from the data.

**Proof.** For a random variable $X \in [0,1]$ with conditional mean $m$, its conditional centered logarithmic moment-generating function has value and first derivative zero at the origin. Its second derivative is the variance under exponential tilting. A variable supported on $[0,1]$ has variance at most $1/4$, because its squared distance from $1/2$ is at most $1/4$. Integrating twice gives

$$\mathbb{E}[e^{\theta(X-m)} \mid \mathcal{F}] \le e^{\theta^2/8}.$$

Iterating this conditional inequality gives the same exponential bound for the martingale sum $\sum_{s \le t}(X_{s,j} - m_{s,j})$. Optimizing the exponential Markov inequality gives the two-sided bound $2\exp(-2b^2/t)$. At the displayed $b_{t,j}$ this is $\alpha_j / [t(t+1)]$. Sum over $t$, using $\sum_{t \ge 1} 1/[t(t+1)] = 1$,

and then over $j$. The resulting event holds at every time, so it also holds at a data-selected time. Scaling the variable proves the range version. □

This elementary boundary pays a logarithmic time-union cost. It is not claimed to be optimal. Stronger confidence-sequence constructions are available in the literature [R20]. The proof here is supplied because the exact assumptions and the estimated object matter more than a sharp constant in this foundation.

The certificate estimates accumulated **conditional predictive drift**, not one fixed unconditional population probability. A prospectively chosen rule may use history, but an analyst cannot inspect an entire data sequence and then retroactively choose an unrestricted test with the same budget. New rules can be registered with fresh budgets and applied to future observations, using elapsed time after registration.

## 75. Actions need coverage, not only more observations

For intervention evidence use the information $\mathcal{H}_{t-1}$ just before choosing the action. Let the action be randomized from a known history-dependent policy $\pi_t$. The true next-observation kernel under action $a$ is $P_t(\cdot \mid \mathcal{H}_{t-1}, a)$. Randomization uses no unrecorded future information. For a predictable test $\phi_t \in [0,1]$, define

$$\mu_t^a = \int \phi_t(y) P_t(dy \mid \mathcal{H}_{t-1}, a), \qquad W_t^a = \frac{\mathbf{1}_{\{A_t=a\}}\phi_t(Y_t)}{\pi_t(a \mid \mathcal{H}_{t-1})}.$$

**Proposition 67. Covered adaptive-action evidence.** If the target action has $\pi_t(a \mid \mathcal{H}_{t-1}) \geq \pi_{\min} > 0$ at every time, then

$$\mathbb{E}[W_t^a \mid \mathcal{H}_{t-1}] = \mu_t^a, \qquad 0 \leq W_t^a \leq 1/\pi_{\min}.$$

Consequently Proposition 66 gives a time-uniform certificate for $\sum_{s\leq t} \mu_s^a$ from $\sum_{s\leq t} W_s^a$, with its boundary multiplied by $1/\pi_{\min}$. The target is the average action-conditioned quantity along the encountered contexts, not a reset-population table or the state distribution under a different long-term policy.

**Proof.** Condition first on the action and then on its observation. Only action $a$ contributes, with probability $\pi_t(a \mid \mathcal{H}_{t-1})$, cancelling the denominator. The range bound follows directly. Apply the scaled conditional moment-generating-function argument to the resulting martingale differences. □

This simple bound is conservative. Variance-aware weighting can improve it, but cannot remove the need for information about the action of interest. When $\pi_{\min} = 0$, two models can give exactly the same law under the observed policy and disagree on a never-selected action. For example, both can emit a Bernoulli observation with parameter $1/2$ under action zero, while their action-one parameters are $1/10$ and $9/10$. Repeating action zero cannot distinguish them.

The action randomization and its recorded probabilities are part of the experimental assumption. An observational association with a hidden selection mechanism is not automatically

an intervention estimate. Long-horizon off-policy predictions require additional occupancy or trajectory-coverage arguments that this one-step statement does not provide.

## 76. What a residual can prove about misspecification

A certified operator family can still be the wrong model for the observed process. Predictable residuals offer an explicit way to find that mismatch without pretending that internal validity proves empirical adequacy.

Let $P_{t,j}$ and $Q_{t,j}$ denote the true and proposed next-observation laws on the same actual history and action. The proposed law must be defined there. Since $0 \le \phi_{t,j} \le 1$,

$$|m_{t,j} - p_{t,j}| \le \|P_{t,j} - Q_{t,j}\|_{\mathrm{TV}}.$$

**Proposition 68. Sequential evidence gives a lower error certificate.** On the simultaneous event of Proposition 66,

$$\boxed{\frac{1}{t}\sum_{s=1}^{t} \|P_{s,j} - Q_{s,j}\|_{\mathrm{TV}} \ge \frac{(|R_{t,j}| - b_{t,j})_+}{t}.}$$

Thus crossing the boundary rejects zero conditional drift for that rule, with the registered false-rejection budget. Failing to cross supplies no general upper bound on total variation, no rank certificate and no guarantee for untested actions.

**Proof.** The triangle inequality gives $|D_{t,j}| \ge (|R_{t,j}| - b_{t,j})_+$. Also $|D_{t,j}| \le \sum_s |m_{s,j} - p_{s,j}| \le \sum_s \|P_{s,j} - Q_{s,j}\|_{\mathrm{TV}}$. Combine these inequalities. □

### 76.1 Why one aggregate residual can miss conditional errors

The reference path has binary observations. Its true conditional probability of one is $4/5$ after a one and $1/5$ after a zero. A deliberately inadequate model predicts $1/2$ at every step. Its long-run aggregate frequency can be near one half even though its history-conditioned predictions are wrong.

Three tests are registered before the path is generated. One tests the raw observation. A second tests it only after the preceding observation was one. The third applies the same gate but uses the correct conditional predictor. Gates are functions of past data, not selected after seeing the current measurement.

For the supplied 4,096-step seeded path, the aggregate mean residual is approximately 0.00732. The gated mean residual is approximately 0.14917, while its time-uniform mean boundary is approximately 0.05114. The lower certificate for average conditional total variation is approximately 0.09803. The correctly predicted gated residual is approximately $-0.00303$.

These are illustrative path calculations, not estimated coverage rates. The theorem supplies the probability guarantee. The example supplies an explicit reason to test conditional structure rather than relying only on cancellation-prone aggregate residuals.

The exact reference suite also checks a dependent likelihood-ratio construction on every binary path through depth eight. Its expectation under the declared null is exactly one. This is an additional finite martingale calculation, not a replacement for the general argument or a claim that every possible test has been run.

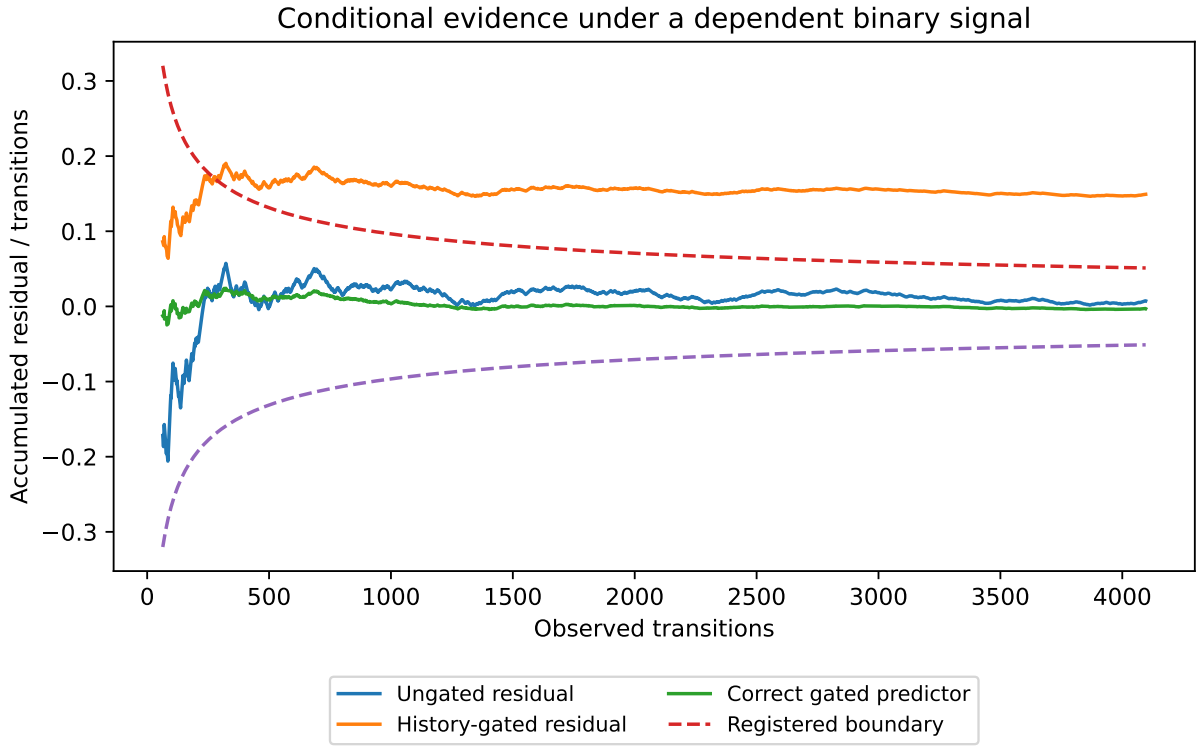



**Figure 6.2:** An ungated average can look compatible with a constant predictor while a prospectively chosen history-gated test reveals conditional mismatch. The figure uses the saved seeded path and its registered time-uniform boundary. It is one illustrative dependent path, not an empirical coverage study. Rejection can certify a lower discrepancy. Nonrejection is not an upper accuracy guarantee.

## 77. Comparing a model with a process outside its representation

The earlier operator-norm law bounds compare two models on one common probability cone. A misspecified generating process may not have a representation in the fitted cone. It is useful to state the more basic comparison at the level of conditional observation laws.

Let $P$ be any causal process on a standard Borel observation space with finite actions. Let $Q$ be a proposed process. Both use the same measurable policy. Write $P_t(\cdot \mid h, a)$ and $Q_t(\cdot \mid h, a)$ for their next-observation kernels, and define

$$d_t(h, a) = \|P_t(\cdot \mid h, a) - Q_t(\cdot \mid h, a)\|_{\mathrm{TV}}.$$

The proposed kernel must be specified on every history needed for the comparison, including histories impossible under $Q$ but possible under $P$. A declared fallback on a model-null history completes a causal kernel without changing $Q$'s own law. An implementation that rejects such a history must report that rejection rather than silently claim to have filtered it.

**Proposition 69. Observation-law error without a shared state space.** With the same initial history, the length-$T$ joint action-and-observation laws satisfy

$$\boxed{\|P_T - Q_T\|_{\mathrm{TV}} \le \min\left(1, \sum_{t=0}^{T-1} \mathbb{E}_{P,\pi}[d_t(H_t, A_t)]\right).}$$

An initial-history discrepancy adds its total variation. In particular, uniform bounds $d_t(h, a) \le \epsilon_t$ give the upper bound $\min(1, \sum_{t<T} \epsilon_t)$.

**Proof.** Introduce hybrid laws that follow $P$ for the first $t$ transitions and $Q$ for the rest, with the common policy throughout. Two successive hybrids have the same $P$ prefix distribution up to time $t$ and the same future $Q$ kernel after the next transition. Conditional on that prefix and its policy-chosen action, changing the next kernel costs $d_t(h, a)$ in total variation. A common continuation kernel cannot increase total variation, because integration of a test bounded between zero and one produces another such test. Average over the shared prefix, sum the adjacent-hybrid bounds, and cap by one. The same contraction argument handles a different initial law. □

This theorem does not make the right-hand side observable without further assumptions. Proposition 68 supplies lower evidence about visited conditional discrepancies, not an upper estimate for every future history. Turning a finite collection of residuals into an upper law guarantee requires additional regularity, coverage or a restricted class with a separating test family.

The distinction prevents a common logical jump: a model can pass several tests and still have a large untested discrepancy. Conversely, a statistically significant residual need not mean that its internal probability algebra is invalid. The model can be a valid distribution and an inadequate approximation at the same time.

## 78. An integrated theory of construction and correction

**Proposition 70. Separate structural acceptance, approximation control and sequential rejection.** Consider a fitted polynomial model in a declared predictive coordinate system, and prospectively registered observation tests with total budget $\delta$.

A successfully replayed quadratic certificate from Proposition 62 establishes internal all-horizon validity. A successful operator-only proposal from Proposition 63 is one way to obtain it, not an assumption that all valid models have such a proposal.

For any generating density on a common probability cone with a justified Lipschitz constant, a positive degree-$n$ approximation and certified coefficient budgets satisfy Propositions 64–65, hence a fixed-horizon upper law bound. If a generating process lies outside that representation, a separately justified bound on its conditional kernels instead invokes Proposition 69.

For predictable tests on actual dependent observations, Propositions 66 and 68 give simultaneous lower discrepancy certificates, including at data-selected stopping times. Covered randomized actions admit the contextual evidence of Proposition 67. These conclusions coexist without implying that structural acceptance or statistical nonrejection proves the generating process belongs to the model class.

**Proof.** Each structural statement is a deterministic consequence of its replayed equalities and inequalities. The approximation statements require their declared common-domain, regularity and coefficient conditions, or alternatively the conditional-kernel discrepancy assumptions. The statistical conclusions hold on the simultaneous probability event of Proposition 66. None of these implications uses nonrejection as an upper bound or infers coverage from a finite determinant. Their conjunction therefore retains the separate hypotheses and conclusions exactly as stated. □

The central object is now more than a fitted operator. It includes an admissible domain with a replayable witness, an explicit approximation class, and an evidence rule specifying which observations could contradict its predictions. Geometry discovery proposes a representation. Verification checks its algebra. Approximation theory accounts for information lost by a chosen finite description. Sequential tests expose conditional discrepancies in observed signals.

Those roles are related but not interchangeable. A failed search means that this construction did not provide a witness. A failed witness means its submitted identities or signs do not establish the claimed validity. A detected residual means a registered predictive condition is inconsistent with the observations at the stated error budget. These three outcomes should be reported differently.

## 79. From exact geometry to robustness

Propositions 62–70 distinguish structural acceptance, controlled approximation and sequential evidence. Exact zero-margin preservation does not supply a positive noise radius. Chapter 7 examines that obstruction and gives strict-margin witnesses for selected perturbed families. Statistical nonrejection is not used as an upper accuracy bound.

CHAPTER 7

# Robust curved certificates and noisy geometry

The preceding chapter discovered a curved probability domain from exact operators. Exact preservation is not the same as stability under uncertainty. In the worked continuous model, every pure boundary state can remain on the boundary. A perturbation too small to be conspicuous in a coefficient table can then move a predicted state outside the probability domain.

This chapter separates three operations. A numerical procedure proposes a geometry. An exact checker decides whether the submitted polynomial family has a declared margin on that geometry. A separately stated uncertainty set determines which other families share the guarantee. If a repair changes the operators, its prediction cost is retained rather than called an exact coordinate change.

The resulting construction is broader than the preceding exact conformal search. It accepts nonconformal polynomial families and certifies a full coefficient box around a repaired estimate. The reference example uses a deterministic perturbation, not an empirical confidence region. Statistical coverage is a conditional theorem, not a conclusion of that example.

Lorentz-positive maps, robust convex constraints and Bernstein positivity certificates have classical antecedents [R8, R13, R21, R22]. The following arguments are self-contained specializations connecting them to the signal model's representation and evidence contracts. The nine propositions continue the numbering at 71. Their count is not a novelty claim or a count of Lean proofs.

## 80. A margin that has probability meaning

Let $\mathcal{L}_d = \{(m, v) : m \geq \|v\|_2\}$, with mass row $\lambda = (1, 0, \ldots, 0)$. A mass preserving invertible matrix $T$ transports a proposed predictive domain to this cone. Its normalized base in the original coordinates is

$$K_T = \{T^{-1}(1, v) : \|v\|_2 \leq 1\}.$$

Write $F_a(u) = TB_a(u)T^{-1}$ for the transported operator density. Observations lie in $[-1, 1]$ with reference measure $du/2$. All operator densities in this chapter are polynomial and all equalities involving integrals use that reference measure. For each action define

$$\sigma_a(u) = \min_{\|v\|_2 \leq 1} \left[ (F_a(u)(1, v))_0 - \|(F_a(u)(1, v))_{\mathrm{sp}}\|_2 \right].$$

The subscript sp denotes spatial coordinates, not a separate measured signal. This margin is expressed before normalization. It simultaneously controls mass orientation and distance from the curved boundary. It is coordinate dependent, so its uncertainty budget must use the same $T$.

**Proposition 71. Robust oriented margin.** Suppose $\sigma_a(u) \geq \kappa_a(u)$ and an uncertain family has transported errors $\Delta_a(u)$ satisfying, for every normalized input,

$$|[\Delta_a(u)(1,v)]_0| \leq e_{a,0}(u),$$
$$\|[\Delta_a(u)(1,v)]_{\mathrm{sp}}\|_2 \leq e_{a,1}(u).$$

If $\kappa_a(u) \geq e_{a,0}(u) + e_{a,1}(u)$, every uncertain branch preserves $\mathcal{L}_d$. If the uncertain family has exact integrated mass and its initial state belongs to $K_T$, it defines a valid probability model at every finite horizon. With a strict inequality, the residual slack is at least $\kappa_a - e_{a,0} - e_{a,1}$.

**Proof.** For $w = F_a(u)(1,v)$ and $d = \Delta_a(u)(1,v)$,

$$(w_0 + d_0) - \|w_{\mathrm{sp}} + d_{\mathrm{sp}}\|_2 \geq w_0 - \|w_{\mathrm{sp}}\|_2 - |d_0| - \|d_{\mathrm{sp}}\|_2.$$

The right side is nonnegative under the hypothesis. Homogeneity extends the assertion from normalized inputs to the entire cone. The only zero-mass cone element is zero. Integrated mass conservation and successive cone preservation give nonnegative normalized observation laws by the same induction as in Proposition 52. A positive-density branch may be normalized. A zero-density branch is not assigned a posterior. □

This is a uniform deterministic guarantee. The inequalities must cover the whole observation interval and the whole normalized domain. A noise scale estimated from a few residuals does not automatically provide those inequalities. For continuous observations the displayed margin is a density margin relative to the declared measure, not a probability assigned to a single observation value.

## 81. Why exact boundary preservation can have zero noise radius

An exact certificate can be correct while having no tolerance to unrestricted perturbations. This is a geometrical fact, not a failure of the certificate checker.

**Proposition 72. Boundary fragility under mass preserving perturbations.** Assume $d \geq 1$. Let $F$ preserve $\mathcal{L}_d$. Suppose a normalized input $z = (1,v)$ has a boundary image $Fz = (m,w)$ with $m = \|w\|_2$. For every $\epsilon > 0$, there is a linear perturbation $\Delta$ with $\lambda\Delta = 0$ and

$$\sup_{\|v'\|_2 \leq 1} \|\Delta(1,v')\|_{\mathcal{L}_d} = \epsilon$$

such that $(F + \Delta)z \notin \mathcal{L}_d$. Here the base norm is $\|(m,w)\|_{\mathcal{L}_d} = \max(|m|, \|w\|_2)$.

**Proof.** If $w \neq 0$, put $e = w/\|w\|_2$. If $w = 0$, choose any spatial unit vector $e$. Define $\Delta = \epsilon(0,e)\lambda$. Its mass row vanishes and its value on every normalized input is $(0, \epsilon e)$, of base norm $\epsilon$. If $w \neq 0$,

the new spatial norm is $m + \epsilon$. If $w = 0$, the original output mass is zero and the new spatial norm is $\epsilon$. Either image lies outside the cone. □

The positive spatial dimension is essential. When $d = 0$, the state is mass alone and $\lambda$ is the identity on $\mathbb{R}$. The condition $\lambda\Delta = 0$ forces $\Delta = 0$, so a perturbation of prescribed norm $\epsilon > 0$ cannot exist. The worked examples in this volume have two spatial coordinates.

For a continuous operator family, a constant spatial perturbation of this form preserves every integrated mass row and produces the same obstruction at an observation with a boundary image. It need not preserve the algebraic restrictions of a narrower uncertainty class. Perturbing a positive congruence factor and rebuilding its congruence, for example, is different from perturbing an arbitrary predictive matrix.

In the preceding exact continuous example, the direct operator at $u = 0$ is $I/4$. Every normalized boundary state has a boundary image. There is therefore no positive unrestricted matrix-error radius on that domain, even though the original family is exactly valid. A robust theorem that silently treats its zero slack as positive would be false.

## 82. An explicit repair and the price of its slack

In Lorentz coordinates let

$$N_a(u) = g_a(u)e_0\lambda, \qquad g_a(u) \geq g_{\min} > 0, \qquad \int g_a(u)\,du/2 = 1.$$

This reference instrument reports a state-independent density and resets to the center. It is a mathematical model modification, not a new observation. Given an estimate $\widehat{F}$, define

$$\widetilde{F}_a(u) = (1-\eta)\widehat{F}_a(u) + \eta N_a(u), \qquad 0 \leq \eta \leq 1.$$

**Proposition 73. Interior repair with an explicit error budget.** Suppose a reference family $F^*$ is valid on $\mathcal{L}_d$, and the combined mass and spatial error of $\widehat{F} - F^*$ is at most $d_a(u)$ on normalized inputs. Then the margin of $\widetilde{F}_a(u)$ is at least

$$\boxed{\eta g_a(u) - (1-\eta)d_a(u).}$$

An additional uncertainty family with combined output error at most $\nu_a(u)$ is certified whenever

$$\eta g_a(u) \geq (1-\eta)d_a(u) + \nu_a(u).$$

Assume exact integrated mass for the estimate. If its integrated base-norm discrepancy from $F^*$ is at most $\varepsilon$, the corrected discrepancy is at most

$$\boxed{(1-\eta)\varepsilon + 2\eta.}$$

When the corrected model is valid on the same cone, with the same initial state, its length-$T$ law under any common policy differs from the reference law by at most

$$\min\left(1, \frac{T}{2}[(1-\eta)\varepsilon + 2\eta]\right).$$

**Proof.** The margin of $\widehat{F}$ is at least $-d_a$ by Proposition 71's inequality applied to $F^*$. Adding the reference changes only output mass, by $\eta g_a$, while scaling the remaining output by $1-\eta$. This proves the first bound and the robust implication. Integrated mass is preserved by mixing.

For the error, write

$$\widetilde{F} - F^* = (1-\eta)(\widehat{F} - F^*) + \eta(N - F^*).$$

On a normalized input, both $N$ and $F^*$ distribute total cone mass one. Their integrated difference has base norm at most two. Integrate the triangle inequality to obtain the stated budget. The final implication is the continuous observation-law bound of Proposition 59. It requires validity of both comparison models on the same cone. □

For constant budgets, a sufficient repair weight is $(d+\nu)/(g_{\min}+d)$ when this number is at most one. This is not an optimality claim. The reference example instead fixes a repair weight and verifies the result independently. If an exact model already has a robust structured certificate, adding a reset may be unnecessary and would introduce avoidable bias.

The term $2\eta$ is deliberately retained. A larger repair creates more room for uncertainty but can change future predictions more. The worst-case law bound can become uninformative at long horizons. Neither positivity nor mixing alone proves long-horizon predictive accuracy.

## 83. Noisy invariant forms are proposals, not certificates

The exact identity $A^{\mathsf{T}}HA = H$ need not survive estimation error. A useful proposal can still be obtained from a least-squares invariant form, but the residual must not be interpreted as a positive-cone guarantee.

Let $\mathcal{M}_A$ map symmetric matrices into a Euclidean product space by

$$\mathcal{M}_A(H) = (A^{\mathsf{T}}HA - H, \operatorname{tr} H).$$

Use the Frobenius norm on symmetric matrices and the induced product norm. Assume $\mathcal{M}_{A^*}$ is injective and $\mathcal{M}_{A^*}(H^*) = (0,1) = b$. This fixes the scale of a one-dimensional invariant-form space.

**Proposition 74. A posteriori proposal error.** Given $\widehat{H}$, let $r = \mathcal{M}_{\widehat{A}}(\widehat{H}) - b$. Suppose

$$\|\mathcal{M}_{\widehat{A}} - \mathcal{M}_{A^*}\| \le e_M, \qquad \sigma_{\min}(\mathcal{M}_{\widehat{A}}) > e_M.$$

Then

$$\boxed{\|\widehat{H} - H^*\|_F \le \frac{\|r\| + e_M\|\widehat{H}\|_F}{\sigma_{\min}(\mathcal{M}_{\widehat{A}}) - e_M}.}$$

One may take $e_M$ from the bound

$$\|\mathcal{M}_{\widehat{A}} - \mathcal{M}_{A^*}\| \le (\|\widehat{A}\|_2 + \|A^*\|_2)\|\widehat{A} - A^*\|_2.$$

For fixed centers $c^* = (I - A^*)^{-1}b^*$ and $\widehat{c} = (I - \widehat{A})^{-1}\widehat{b}$, with $\|\widehat{A} - A^*\|_2 \le e_A$ and $\|\widehat{b} - b^*\|_2 \le e_b$, one has

$$\|\widehat{c} - c^*\|_2 \le \frac{e_b + e_A\|\widehat{c}\|_2}{\sigma_{\min}(I - \widehat{A}) - e_A}$$

when the denominator is positive.

**Proof.** The invariant-form difference satisfies

$$\mathcal{M}_{A^*}(\widehat{H} - H^*) = r - (\mathcal{M}_{\widehat{A}} - \mathcal{M}_{A^*})\widehat{H}.$$

The least singular value of $\mathcal{M}_{A^*}$ is at least that of $\mathcal{M}_{\widehat{A}}$ minus $e_M$. Apply this lower bound and the triangle inequality. To bound the map perturbation, insert and subtract $\widehat{A}^{\mathsf{T}}HA^*$ and apply the Frobenius product inequality. The trace component is unchanged. Finally,

$$(I - A^*)(\widehat{c} - c^*) = (\widehat{b} - b^*) + (\widehat{A} - A^*)\widehat{c}.$$

The same singular-value argument proves the center bound. □

The theorem assumes an actual perturbation budget and a corresponding exact reference problem. The example reports numerical residuals but does not claim to have statistically certified these budgets. It uses a trace-normalized least-squares shape, a nearly rank-one image for scale, and rationalized center and factor. It accepts none of those computations as a proof of validity. The next checker uses only the submitted rational geometry and the complete polynomial operators.

A rationalized invertible factor $L$ specifies $H = L^{\mathsf{T}}L$ exactly. Positive definiteness is then an algebraic fact even when the factor was suggested by floating-point calculations. It still does not establish operator invariance.

## 84. Finite exact checking on a continuum

For three predictive coordinates, the spatial unit circle admits two rational charts. For $s \in \{1, -1\}$ and $t \in [-1, 1]$, put

$$d(t) = 1 + t^2, \qquad r_s(t) = \frac{s}{d(t)}(1 - t^2, 2t).$$

The charts cover the two semicircles, including their endpoints. For an operator $F_a(u)$ and required margin $\kappa \geq 0$, form polynomial vectors and scalars

$$\begin{aligned} w_{a,s}(u,t) &= F_a(u)(d(t), s(1-t^2), 2st)^{\mathsf{T}}, \\ f_{a,s}(u,t) &= w_{a,s,0}(u,t) - \kappa d(t), \\ h_{a,s}(u,t) &= f_{a,s}(u,t)^2 - w_{a,s,1}(u,t)^2 - w_{a,s,2}(u,t)^2. \end{aligned}$$

No square root occurs in these polynomials.

**Proposition 75. Rational semicircle and Bernstein certificate.** If $f_{a,s}$ and $h_{a,s}$ are nonnegative on $[-1, 1]^2$ for both charts and every action, then $\sigma_a(u) \geq \kappa$ for every observation. For rational operators and geometry, nonnegative tensor-Bernstein coefficients on a complete rational subdivision of that square give a finite exact witness. If the actual margin is uniformly strictly greater than $\kappa$, sufficiently fine dyadic subdivision succeeds.

**Proof.** The first polynomial inequality establishes output orientation after subtracting $\kappa d$. The second then gives

$$w_{a,s,0} - \|(w_{a,s,1}, w_{a,s,2})\|_2 \geq \kappa d.$$

Divide by $d > 0$ to obtain the required margin on the circle. For fixed $u$, the function consisting of affine output mass minus the norm of affine spatial output is concave in the input. Every point in the closed disk is a convex combination of circle points. Concavity therefore extends the lower bound to the disk.

On any rectangle mapped affinely to $[0, 1]^2$, the tensor-Bernstein basis functions are nonnegative and sum to one. A polynomial with nonnegative coefficients in that basis is nonnegative throughout the rectangle. If the rectangles cover the square, their inequalities prove the global statement. The reference verifier checks coverage through a complete subdivision tree, not by accepting an arbitrary list of rectangles.

For termination under strict margin, both $f$ and $h$ are strictly positive on the compact square. Their minima are positive. For a fixed-degree polynomial on a small cell of side at most $\delta$, its Taylor expansion at one corner has constant term equal to its value there and nonconstant coefficients bounded uniformly by $O(\delta)$. The linear change from the fixed-degree power basis to the Bernstein basis is fixed and bounded. Thus every Bernstein coefficient differs from that constant value by $O(\delta)$, uniformly over cells. At sufficiently small $\delta$ all coefficients are positive. A finite dyadic subdivision attains that size. □

This proof is a specialized use of classical Bernstein range bounds [R22]. It does not claim that a negative Bernstein coefficient proves the polynomial negative. Such a coefficient can merely mean that further subdivision is needed. A negative evaluated corner, by contrast, is an actual counterexample. At a zero-margin boundary, finite termination of this particular witness language is not promised. The exact conformal and matrix certificates of the preceding chapters remain useful there.

The two tests are both necessary to this sufficient certificate. The squared inequality alone also admits reversed mass orientation. Both charts are required. A verifier that omits one chart has not covered the circle.

## 85. A finite uncertainty box and the role of convexity

A robust witness should identify its uncertainty family rather than supply an unexplained tolerance. Suppose the transformed error has polynomial coefficients

$$\Delta_a(u) = \sum_{k=0}^{D} u^k E_{a,k}, \qquad |(E_{a,k})_{ij}| \le r_k.$$

All bounds are in the submitted Lorentz coordinates. For three coordinates, normalized inputs have each coordinate of magnitude at most one.

**Proposition 76. Coefficient boxes and convex uncertainty.** The coefficient box satisfies the output bounds of Proposition 71 with

$$\boxed{e_0 = 3\sum_{k=0}^{D} r_k, \qquad e_1 = 6\sum_{k=0}^{D} r_k.}$$

Only coefficient arrays obeying $\lambda \int \Delta_a(u)\, du/2 = 0$ are included in the normalized uncertainty family. A checked margin at least $e_0 + e_1$ certifies every such array.

Alternatively, if finitely many complete polynomial operator families satisfy a common margin $\kappa$ on the same domain, every convex combination of those families satisfies that margin. Initial states may also range over a certified convex hull.

For a fixed geometry, the constraints $\sigma_a(u) \ge \kappa$, exact mass, and membership in a convex data-consistency set are convex constraints on the operator coefficients and initial state. Joint optimization of geometry and operators is not asserted to be convex.

**Proof.** Each output scalar is bounded by $3r_k$ on a normalized input. Bound the spatial Euclidean norm by the sum of the absolute values of its two coordinates, obtaining $6r_k$. Sum over coefficients using $|u| \le 1$. Mass conservation is a separate linear constraint and follows for every family satisfying it.

For convex combinations, output mass is affine and the spatial norm is convex. Hence the margin of the combined output is at least the same combination of margins. A convex cone contains convex combinations of admissible initial states. For fixed geometry and input, the inequality

affine mass minus a norm at least $\kappa$ is convex in the coefficients. Intersect over all inputs and observations, then add the linear mass constraints and the convex data set. □

The squared polynomial witness in Proposition 75 is a convenient verification language. Its coefficient constraints are not automatically a convex optimization formulation in unknown operators. Convexity of the underlying admissibility set must not be confused with convexity of every certificate parameterization.

Likewise, a free uncertainty box in all entries contains families that violate integrated mass. The affine restriction is essential. Removing physically valid candidates from a statistical confidence set for some other reason requires its own coverage argument.

## 86. What a search can guarantee

The numerical proposal is one practical route. There is also a precise conditional search theorem showing why a broad class of strict curved certificates can be found without an exact conformal identity.

**Proposition 77. Conditional search for strict rational curved certificates.** Consider finitely many rational polynomial operator families in three coordinates, each with exact mass, and finitely many rational normalized initial states. Suppose there exists a real mass preserving invertible $T_*$ such that every transformed initial state lies strictly inside the Lorentz cone and every transformed operator family has a common positive margin on the observation interval. Then a rational mass preserving invertible $T$, a positive rational margin, and finite rational Bernstein witnesses exist. A dovetailed enumeration of rational transforms, positive rational margins and subdivision depths eventually finds such a witness. Whenever it returns, the witness is sound for the convex hull of the supplied families.

**Proof.** Inversion and matrix multiplication are continuous near $T_*$. Initial membership is strict, and the operator margin is continuous jointly in the transform, the observation, and the normalized input. The latter two range over compact sets and there are finitely many families. Thus the strict inequalities persist in a neighborhood of $T_*$. Rational mass preserving matrices are dense in that affine matrix space, and invertibility persists in a neighborhood. Choose a rational transform there and a rational positive margin smaller than the remaining minimum. Proposition 75 supplies finite subdivisions because there is still a strict gap above the chosen margin. Enumerating all finite rational descriptions and depths eventually reaches this finite witness. Proposition 76 proves the convex-hull conclusion. □

This is a conditional semidecision result, not a polynomial-time complexity theorem. It does not cover every valid model. Initial states on an essential boundary and families with zero margin may violate its assumptions. The implemented procedure uses a finite rationalized proposal and bounded subdivision, not this unbounded enumeration. Failure of that bounded procedure is a withheld certificate rather than a proof that no other domain works.

The theorem's uncertainty families are supplied in fixed predictive coordinates. It does not establish that those families cover an unknown generating process. It demonstrates a route from a strict robust model sector to finite machine-checkable witnesses.

## 87. Selection from data does not remove the uncertainty quantifier

A domain can be selected after examining fitted operators without a separate statistical penalty for each geometry tried. That statement holds only when the final verification ranges over an already justified confidence set.

**Proposition 78. Data-selected geometry with honest coverage.** Suppose a random set $\mathcal{U}$ of operators and initial states, in a declared fixed predictive coordinate system, contains the true model with probability at least $1 - \delta$. A data-dependent procedure may propose any transform $T$ and any domain. If its exact verifier proves domain membership and invariance for every member of the same set $\mathcal{U}$, then on that coverage event the true model has the verified domain. No union bound over geometry proposals is needed for this deterministic implication.

If original-coordinate perturbations satisfy $\|\Delta\|_2 \le e$, their combined mass and spatial output error after a known transform $T$ is at most

$$\boxed{2\|T\|_2\|T^{-1}\|_2 e.}$$

The corresponding transformed uncertainty budget must be used even when $T$ was chosen from the data.

**Proof.** On the coverage event the true model is a member of the set quantified by the verifier. Its domain properties follow by direct inclusion, regardless of how $T$ was selected. For the error, a normalized Lorentz input has Euclidean norm at most $\sqrt{2}$. Thus the Euclidean norm of $T\Delta T^{-1}z$ is at most $\sqrt{2}\|T\|_2\|T^{-1}\|_2 e$. The sum of its mass magnitude and spatial norm is at most $\sqrt{2}$ times its Euclidean norm. Multiply the two bounds. □

The set must not be shrunk around a convenient candidate without preserving coverage. Moving the center by a reset mixture changes the family being certified. A tube around the repaired family need not contain the original true family. Repeatedly estimating new confidence sets still requires the stage accounting of Proposition 48. Statistical rank coverage and observation identifiability remain assumptions of the upstream procedure.

The final reference certificate explicitly says that no statistical coverage is claimed. Its uncertainty box is a deterministic, exactly described set around the repaired operator family. This supports a reproducible robustness statement without pretending that the example has estimated an unknown noise distribution.

## 88. A complete perturbed-operator construction

The input to the new proposer contains only two perturbed polynomial operators and the initial vector. A separate fixture generator creates those inputs from the preceding continuous model by adding $\pm\delta E(u)$, with opposite signs for the two actions and

$$\delta = \frac{1}{2000}, \qquad E(u) = E_0 + uE_1 + u^2E_2,$$

where

$$E_0 = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 2 & 0 \\ 0 & 1 & -1 \end{pmatrix}, \quad E_1 = \begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

The perturbation preserves integrated mass. It is not a statistical sample. The proposer does not receive the predecessor metric or a hidden generator.

A noisy operator ratio supplies a center and a least-squares shape. A finite observation-anchor list supplies a scale proposal. Rationalizing the center and a triangular metric factor with denominator limit 48 gives

$$c = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}, \qquad L = \begin{pmatrix} 123/37 & -2 \\ 0 & 109/41 \end{pmatrix},$$

and hence

$$\boxed{H = L^{\mathsf{T}} L = \begin{pmatrix} 15129/1369 & -246/37 \\ -246/37 & 18605/1681 \end{pmatrix}.}$$

This is not the exact predecessor metric. The transformed family is not assumed conformal. The raw fitted family actually fails the proposed domain: for the direct action at $u = -1$ and Lorentz input $(1, 1, 0)$, its output mass is $106321/82000 > 0$, while the output quadratic form is

$$-\frac{12009932651317769}{557059220496000000} < 0.$$

This is a genuine rejected input state, not a negative Bernstein coefficient of uncertain meaning. The normalized output lies outside the ellipse. No positivity certificate is claimed for the raw fit.

The construction next mixes each operator with the center-reset reference using $\eta = 1/8$. The exact checker proves a margin of at least $1/16$ for the resulting family. It verifies eight polynomial inequalities across both actions and both semicircle charts. The witness contains 116 subdivision leaves and 1,780 exactly recomputed nonnegative Bernstein coefficients. These are arithmetic certificate sizes, not separate theorems.

The exported family admits mass and spatial perturbation budgets each equal to $1/128$, leaving certified slack $3/64$. In particular, it covers every degree-two transformed coefficient perturbation with each entry bounded by $1/4096$, subject to the exact integrated mass constraints. Proposition 76 verifies that the whole coefficient box lies inside the declared output budgets. No uncertainty vertices or future trajectories are sampled to establish that inclusion.

The certificate is replayed independently from its JSON record. The record contains rational operators, initial state, transform, margin, uncertainty definition and complete subdivision trees. Floating-point proposal diagnostics are stored elsewhere and are not trusted by the replay routine.

## 89. The prediction cost needs a common comparison domain

A successful discovered domain does not imply that the known reference model has that same domain. To invoke a common-cone law bound, the example separately checks the corrected family on the predecessor's exact ellipse. This is a second certificate used only for comparison. It is not supplied to geometry discovery.

In that comparison domain, the coefficient-wise upper bounds on the raw fitting discrepancy are

$$\frac{11}{5000}, \qquad \frac{137}{30000}, \qquad \frac{37}{8000}.$$

Using $\int_{-1}^{1} |u|^k du/2 = 1/(k+1)$ gives the integrated upper budget $241/40000$. The same center-reset reference is valid there. Both models have the same initial state.

**Proposition 79. Certified repair with a retained comparison budget.** For the exact construction of Section 88, independent replay establishes internal validity of the repaired model and its declared coefficient uncertainty family. Independent comparison-domain replay and the coefficient bounds above give

$$\varepsilon_{\text{repaired}} \le \frac{7}{8}\frac{241}{40000} + \frac{1}{4} = \frac{81687}{320000}.$$

Consequently, for any common policy and any finite horizon $T$,

$$\boxed{\|P_T - \widetilde{P}_T\|_{\text{TV}} \le \min\left(1, \frac{81687T}{640000}\right).}$$

This compares the unmodified reference process and the nominal repaired process. It is not a guarantee that the raw noisy model was valid or that the repaired uncertainty family contains the original reference.

**Proof.** The polynomial witnesses and exact mass identities imply the nominal and robust validity statements by Propositions 71, 75 and 76. The separate comparison witness establishes the required common domain for the nominal repaired and reference operators. The absolute coefficient bounds dominate the integrated raw error by the triangle inequality. Proposition 73 then gives the displayed repair budget and law bound. All hypotheses are either exact identities in the supplied witnesses or direct norm bounds from the declared rational coefficients. □

The bound is conservative. It becomes one at horizon eight, so it makes no nontrivial long-horizon accuracy claim. The example's point is the separation of a verified all-horizon probability model from a finite-horizon approximation guarantee. A probability-valid model is not necessarily a close model.

The noisy proposal, the failed raw fit, the repair and the comparison bound are all retained in the evidence records. None is replaced by a more favorable trajectory score. The coefficient box is around the repaired family in the specified coordinates, not an inferred distribution of sensor noise.

## 90. From deterministic robustness to data confidence

Chapter 7 gives exact continuum certificates for deterministic uncertainty families. The worked repaired model is internally valid at every finite horizon, but its original comparison bound becomes trivial at horizon eight. Chapter 8 reduces that fixed repair and separately constructs a statistically covered domain for a supplied observable family. Its confidence statement is not transferred to the earlier repaired operators.

CHAPTER 8

# Data-supported domains and smaller repairs

The preceding chapter certified a perturbed operator family after mixing it with a center reset. Its all-horizon validity guarantee was exact, but its error bound became trivial at horizon eight. It also kept deterministic coefficient boxes separate from statistical confidence sets. This chapter addresses both limitations in declared sectors.

First, the same fitted family and previously proposed geometry admit a smaller, independently checked repair. Second, a self-normalized martingale argument produces operator uncertainty sets from a dependent trajectory when the predictive state is an accessible, fixed observable. A nonlinear three-coordinate example connects those sets to a curved domain without a reset correction. Its geometry and feature dictionary are specified in advance. This is not a solution to discovering an unknown hidden state from the same data.

The sequential-regression result is a classical self-normalized concentration theorem [R23], proved here in the form used by the signal model. The contraction argument uses standard Markov-kernel and total-variation ideas [R24]. The connection to the realization calculus consists of an explicit sequence of assumptions, finite witnesses, structured uncertainty, observation-measure distinctions and executable examples. Proposition counts are not historical-novelty counts.

## 91. A smaller repair can be bracketed rather than guessed

Keep the fitted operators, proposed transform and initial state from Section 88 unchanged. In Lorentz coordinates use the center-reset operator $N = e_0\lambda$, of constant density one under $du/2$. Write

$$F_\eta = (1-\eta)\widehat{F} + \eta N, \qquad 0 \le \eta \le 1.$$

Fix a required margin $0 \le \kappa \le 1$. Feasibility means that every action, observation and normalized input has margin at least $\kappa$. Integrated mass and initial membership remain separate requirements.

**Proposition 80. Convex repair bracketing.** The feasible repair weights form a closed interval $[\eta_*, 1]$. If a cone-positive functional $\ell = (1, -v^{\mathsf{T}})$, with $\|v\|_2 \le 1$, and a normalized input $z$ give $c = \ell \widehat{F}_a(u) z < \kappa$, then every feasible weight satisfies

$$\boxed{\eta \ge \frac{\kappa - c}{1 - c}.}$$

A complete certificate at a weight $\eta_U$ gives an upper bound on $\eta_*$. A negative Bernstein coefficient without a negative evaluated point does not give a lower bound.

**Proof.** For a fixed input and observation, output mass minus spatial norm is a concave continuous function of the operator, hence of $\eta$. Its superlevel set is closed and convex. Intersecting these sets preserves those properties. The reset has margin one, so the intersection contains one and is an interval ending there. If $F_\eta$ has margin at least $\kappa$, every such functional satisfies $\ell F_\eta z \ge \kappa$. Since $\ell N z = 1$, this requires $(1 - \eta)c + \eta \ge \kappa$. Here $1 - c > 0$, and solving gives the bound. A complete positive witness proves feasibility. A negative coefficient only says that a particular sufficient polynomial bound failed. □

### 91.1 The unchanged predecessor example

Use the previous transform with center $(1/2, 1/2)$ and triangular factor

$$L = \begin{pmatrix} 123/37 & -2 \\ 0 & 109/41 \end{pmatrix}.$$

At the direct action, observation $u = -1$, normalized input $(1, 1, 0)$ and functional $\ell = (1, 7/25, -24/25)$, exact evaluation gives

$$c = -\frac{17590411}{2332387500}.$$

For the required margin $\kappa = 1/64$, this proves

$$\frac{864543451}{37599646576} \le \eta_* \le \frac{1}{32}.$$

The lower endpoint is approximately 0.02299. The upper endpoint is not a sampled estimate. A complete exact Bernstein certificate proves that $\eta = 1/32$ works. The weight $1/64$ fails the requested $1/64$ margin at the displayed separating functional. This does not by itself prove that its nominal model is probability-invalid.

The new certificate uses 194 leaves and 3,058 recomputed coefficients. A separate comparison-domain certificate uses 185 leaves and 2,833 coefficients. Both retain the same degree-two coefficient box of radius $1/4096$ in their respective declared Lorentz coordinates, subject to exact integrated mass. The more general mass and spatial output budgets remain $1/128$ each. Their sum equals the certified margin, so that larger output-error family has a nonnegative lower margin, not a claimed strictly positive residual margin.

The reset weight has fallen from $1/8$ to $1/32$, while the deterministic uncertainty box is unchanged. No global optimization over geometry or repair operators is claimed. The current bracket concerns this fixed geometry, this reset path and this required margin.

The comparison domain is checked again before invoking the law theorem. Keeping the same raw discrepancy budget $241/40000$ from Section 89 gives

$$\varepsilon_{\mathrm{repair}} \le \frac{31}{32}\frac{241}{40000} + \frac{1}{16} = \frac{87471}{1280000}.$$

Consequently, under a common policy and identical initial states,

$$\boxed{\|P_T - \widetilde{P}_T\|_{\mathrm{TV}} \le \min\left(1, \frac{87471\,T}{2560000}\right).}$$

The first integer horizon where this upper bound reaches one is now 30 rather than eight. The bound remains conservative. The repaired uncertainty tube is still not a confidence set for the unmodified generating process. This improvement is deterministic certificate refinement, not a new statistical experiment.

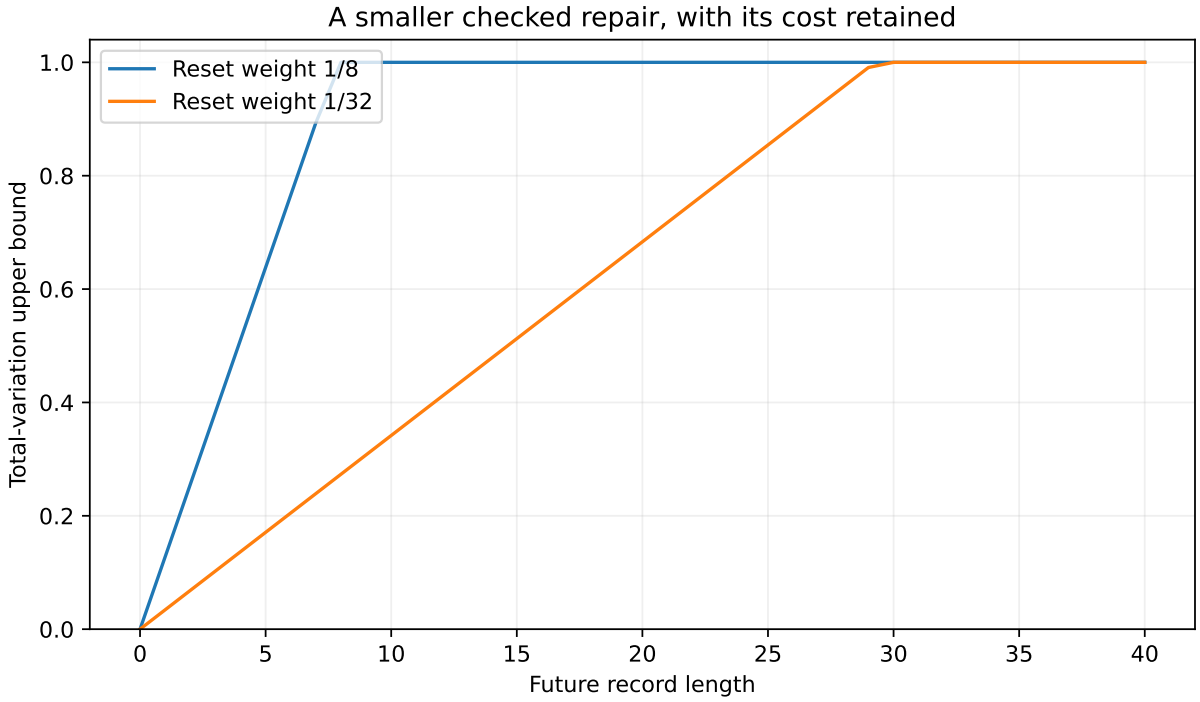



**Figure 8.1:** The same proposed geometry and perturbed operator family are certified with reset weight $1/32$ instead of $1/8$. The retained uncertainty radius is unchanged. The smaller repair's conservative full-record bound first becomes trivial at horizon 30 rather than eight. These lines are upper bounds, not observed model errors. The smaller repair is bracketed for one fixed geometry and reset path, not globally optimized.

## 92. Confidence from dependent observations

Let $\mathcal{F}_{t-1}$ contain all information available before observation $t$, including its already selected action. Let $\phi_t \in \mathbb{R}^d$ be measurable with respect to that information. Assume

$$Y_t = \theta_*^{\mathsf{T}} \phi_t + \xi_t, \qquad \mathbb{E}[e^{h\xi_t} \mid \mathcal{F}_{t-1}] \le e^{h^2\sigma^2/2}$$

for every real $h$, with $\sigma > 0$, and a declared bound $\|\theta_*\|_2 \le S$. With ridge $\gamma > 0$, set

$$V_t = \gamma I + \sum_{s \le t} \phi_s \phi_s^{\mathsf{T}}, \qquad \widehat{\theta}_t = V_t^{-1} \sum_{s \le t} \phi_s Y_s.$$

**Proposition 81. Time-uniform regression confidence.** For $0 < \delta < 1$, with probability at least $1 - \delta$, simultaneously for every finite $t$,

$$\boxed{\|\widehat{\theta}_t - \theta_*\|_{V_t} \le \sigma \sqrt{\log \frac{\det V_t}{\det(\gamma I)} + 2 \log \frac{1}{\delta}} + \sqrt{\gamma} S.}$$

No independence of successive observations or of the regressors is required. The conditional mean, conditional noise and predictability hypotheses are required. The confidence set remains valid at a stopping time selected from the data.

**Proof.** Put $S_t = \sum_{s \le t} \phi_s \xi_s$. For every fixed $q$, the conditional moment bound makes

$$\exp\left(q^{\mathsf{T}} S_t/\sigma - \tfrac{1}{2} q^{\mathsf{T}} (V_t - \gamma I) q\right)$$

a nonnegative supermartingale. Integrate against the fixed Gaussian law $q \sim N(0, (\gamma I)^{-1})$. Conditional Tonelli and completing the square give the nonnegative supermartingale

$$M_t = \left(\frac{\det(\gamma I)}{\det V_t}\right)^{1/2} \exp\left(\frac{\|S_t\|^2_{V_t^{-1}}}{2\sigma^2}\right), \qquad M_0 = 1.$$

For completeness, stop at the first crossing of $1/\delta$, then at any deterministic finite horizon. The supermartingale expectation is at most one, so the crossing probability up to that horizon is at most $\delta$. Increasing the horizon proves the same bound for any finite crossing. Off this event, taking logarithms bounds $\|S_t\|_{V_t^{-1}}$ by the square-root term. Finally,

$$\widehat{\theta}_t - \theta_* = V_t^{-1}(S_t - \gamma\theta_*).$$

The triangle inequality and $V_t \succeq \gamma I$ bound the ridge term by $\sqrt{\gamma} S$. The simultaneous event already covers every selected finite stopping time. □

This is the self-normalized mixture argument in [R23], not a new concentration principle. A bounded response in an interval of length $b - a$ has conditionally sub-Gaussian centered noise

with $\sigma = (b-a)/2$. One direct proof differentiates its tilted log moment-generating function twice. The second derivative is a variance bounded by $(b-a)^2/4$. Integrating from zero gives the stated moment bound.

The determinant term accounts for the observed design, but does not create excitation. If $\phi_t$ never probes a direction, the set need not shrink in that direction. Regressors extracted after inspecting the same response are not predictable. An estimated hidden state cannot be substituted for $\phi_t$ while silently retaining the theorem's true conditional-mean equation.

## 93. From observable moment regression to operator sets

There is a direct connection to the unnormalized instrument. Assume that the normalized predictive state $Z_t$ is an exact, fixed, measurable function of the available observation history. For action $a$, suppose

$$P(U_t \in du \mid \mathcal{F}_{t-1}) = \lambda B_a(u) Z_t \, \mu(du), \qquad Z_{t+1} = \frac{B_a(U_t) Z_t}{\lambda B_a(U_t) Z_t}$$

on realized positive-density branches. Crucially, $Z_{t+1}$ must also be available from the observations using the specified state map. This is an accessibility assumption, not a consequence of successful regression.

Let $B_a(u) = \sum_{k=0}^{D} \psi_k(u) C_{a,k}$. Choose functions $g_j$ so that the known matrix

$$W_{jk} = \int g_j(u) \psi_k(u) \, \mu(du)$$

is invertible.

**Proposition 82. Sequential observable-moment operator sets.**
Define $M_{a,j} = \int g_j(u) B_a(u) \, \mu(du)$. Then

$$\boxed{\mathbb{E}[g_j(U_t) Z_{t+1} \mid \mathcal{F}_{t-1}] = M_{a,j} Z_t, \qquad C_{a,k} = \sum_j (W^{-1})_{kj} M_{a,j}.}$$

Under the rowwise noise and parameter-bound assumptions of Proposition 81, simultaneous confidence ellipsoids for the rows of $M_{a,j}$ give a confidence set for every $C_{a,k}$ by this known linear map. If row $i$ of $\widehat{M}_{a,j} - M_{a,j}$ has $V_{a,t}$-norm at most $\beta_{a,j,i}$, then for any direction $x$,

$$\left| e_i^{\mathsf{T}} (\widehat{C}_{a,k} - C_{a,k}) x \right| \le \sum_j |(W^{-1})_{kj}| \, \beta_{a,j,i} \, \|x\|_{V_{a,t}^{-1}}.$$

**Proof.** In the conditional integral, multiply the normalized state by the observation density. The denominator cancels, giving $\int g_j(u) B_a(u) Z_t \, \mu(du)$. The polynomial expansion gives $M_{a,j} = \sum_k W_{jk} C_{a,k}$ and hence the inverse formula. Apply Proposition 81 to each row, using the predictable regressor $Z_t$ on visits to action $a$, or its action-gated version. A union bound over the predeclared finite rows and moments gives simultaneous coverage. Independence between these

response coordinates is unnecessary. Weighted Cauchy–Schwarz and the triangle inequality give the displayed directional bound. □

Exact mass constraints can be imposed by intersecting the confidence set with the known mass affine subspace. This preserves coverage because the true model is assumed to obey that law. Arbitrarily shrinking a set around a convenient corrected operator does not have the same justification. Data-selected geometry can be checked against the whole resulting set using Proposition 78.

The theorem is conditional on a correct finite observation basis, accessible predictive state and parameter bounds. It does not obtain a hidden predictive state from a single dependent path without additional assumptions. Unknown dictionary discovery, state-estimation error and polynomial misspecification remain distinct problems. The following example makes the state accessible by construction rather than concealing that requirement.

## 94. A nonlinear observable state with a curved moment domain

For $u \in [-1, 1]$, put

$$v(u) = (1, u, u^2)^{\mathsf{T}}, \qquad L_2(u) = \frac{3u^2 - 1}{2}.$$

The normalized states are moments of a scalar signal:

$$K = \{(1, x, y) : x^2 \le y \le 1\}.$$

Their cone has the finite curved description

$$C = \left\{ (m, x, y) : \begin{pmatrix} m & x \\ x & y \end{pmatrix} \succeq 0, \quad m - y \ge 0 \right\}, \qquad \lambda = (1, 0, 0).$$

Assume that the declared probability measure $\mu$ on $[-1, 1]$ has moments zero at orders one and three, $1/3$ at order two and $1/5$ at order four. Define a two-parameter degree-four operator density

$$\boxed{B_{a,b}(u) = v(u)\,(1, au, bL_2(u)).}$$

Here $a, b$ are parameters, not action labels. This example is autonomous. Its dependence on the observed signal is nonlinear because the state includes its square.

**Proposition 83. Structured curved-domain certificate.** The set $K$ is the convex hull of all $v(u)$. If $|a| + |b| \le 1 - \rho$ for $0 \le \rho \le 1$, then $B_{a,b}(u)$ preserves $C$ and has exact integrated mass. On a normalized state, its density is

$$p_{a,b}(u \mid x, y) = 1 + axu + byL_2(u) \ge \rho.$$

Every positive-density posterior is exactly $v(u)$. In particular, starting from $v(s)$ produces the observable Markov law

$$p_{a,b}(u \mid s) = 1 + asu + bs^2 L_2(u).$$

Its next conditional moments satisfy

$$\boxed{\mathbb{E}[U_t \mid U_{t-1} = s] = \frac{a}{3}s, \qquad \mathbb{E}[L_2(U_t) \mid U_{t-1} = s] = \frac{b}{5}s^2.}$$

**Proof.** Every distribution on $[-1,1]$ has $x^2 \le y \le 1$. Conversely, fix $|x| < 1$. A point mass at $x$ gives $(x, x^2)$, while the endpoint mixture with mean $x$ gives $(x, 1)$. Their convex combinations give every intermediate $y$. The cases $x = \pm 1$ are immediate. Homogenizing yields the matrix and scalar inequalities, which also force the only mass-zero cone element to be zero.

On $K$, $|x| \le 1, 0 \le y \le 1$, $|u| \le 1$ and $|L_2(u)| \le 1$. Thus the density is at least $1 - |a| - |b|$. Each operator output is that nonnegative scalar times $v(u)$, so lies in $C$. The first and second reference moments imply $\lambda \int B_{a,b}(u)\, \mu(du) = \lambda$. Normalization cancels the scalar density, yielding $v(u)$. Symmetry through order three gives $\int u L_2(u) d\mu = 0$. The second and fourth moments give $\int L_2(u)^2 d\mu = 1/5$. Integrating $u$ and $L_2$ against the density proves the last two formulas. □

The posterior ray lies on a curved boundary. Arbitrary perturbations of its output coordinates can destroy validity, as in Proposition 72. Parameter uncertainty inside the declared family behaves differently: it changes the likelihood but not the posterior ray. A certified rectangle in $(a, b)$ can therefore be robust even though a free matrix-entry ball has no positive validity radius.

This is an example of uncertainty that preserves the model's algebraic structure. It is not a claim that every real system obeys the two-parameter family. The feature dictionary $1, u, u^2$, its geometry, the reference measure and the parameter model are declared inputs to learning.

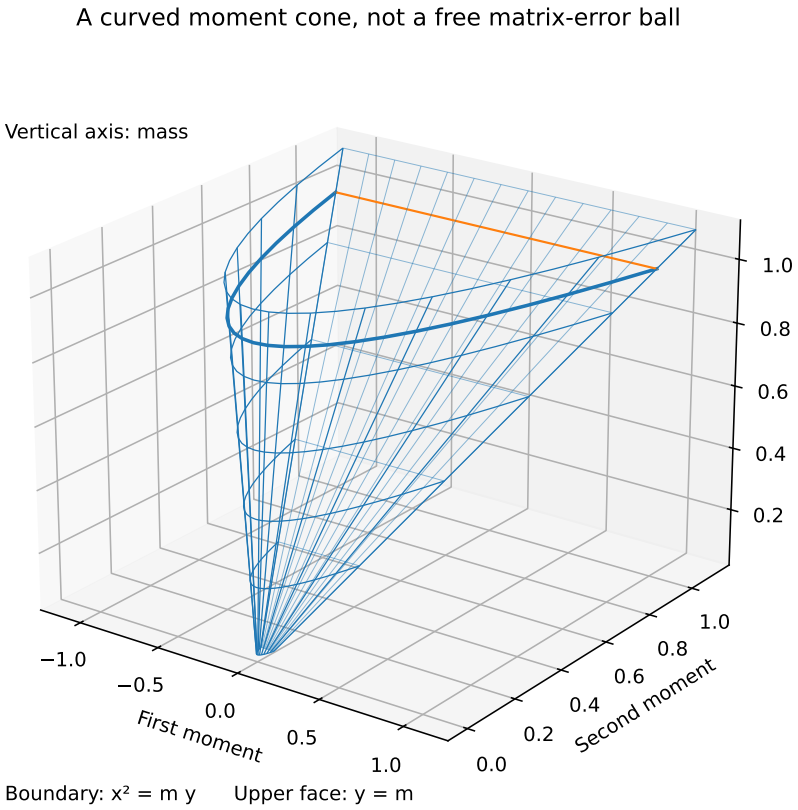



**Figure 8.2:** The moment cone is defined by a positive-semidefinite two-by-two moment matrix and an upper second-moment constraint. Parameter uncertainty changes the observation density but leaves the posterior state $v(u) = (1, u, u^2)$ unchanged. This structured parameter family is not a free matrix-entry uncertainty ball. The geometry and feature dictionary are declared, not discovered by this example.

## 95. A fully rational dependent-data reference experiment

Both the continuous uniform measure $du/2$ and the following five-node measure satisfy the moment assumptions:

| $u$ | $-1$ | $-1/2$ | $0$ | $1/2$ | $1$ |
|---|---|---|---|---|---|
| $\mu\{u\}$ | $7/90$ | $16/45$ | $2/15$ | $16/45$ | $7/90.$ |

The reference experiment uses the **five-node measure**, not rounded observations from the continuous model. At each step, the conditional probability of a node is its reference weight times $p_{a,b}(u \mid s)$. Exact rational probabilities are sampled by integer draws. The learner receives only the node sequence, the measure and the declared model. The true parameters are present only in the separate generator and verification code.

Matching four moments does not identify the two observation laws. The continuous model gives the set of five nodes probability zero, while the discrete model gives it probability one. Their single-observation total variation is therefore one. The continuous theorems and the exact discrete reference calculation are deliberately distinguished.

For a path of $n$ transitions, use two scalar regressions,

$$\begin{aligned} \phi_{a,t} &= U_{t-1}/3, \quad Y_{a,t} = U_t, & \sigma_a &= 1, \\ \phi_{b,t} &= U_{t-1}^2/5, \quad Y_{b,t} = L_2(U_t), & \sigma_b &= 3/4. \end{aligned}$$

The noise scales follow from response ranges $[-1,1]$ and $[-1/2,1]$. The prior parameter bounds are $|a|,|b| \le 1$. Each regression uses ridge one and failure probability $1/40$, giving a total failure budget $1/20$ for both parameters at every time.

**Proposition 84. Exact rational outer confidence intervals.** Write $V = 1 + \sum \phi_t^2$ and $\widehat{\theta} = \sum \phi_t Y_t / V$. Let $k = \lceil \log_2 V \rceil$ and $h = \lceil \log_2(1/\delta) \rceil$. A rational upper enclosure of

$$\sigma \sqrt{\frac{(7/10)(k+2h)}{V}} + \sqrt{\frac{1}{V}}$$

is an upper bound on Proposition 81's scalar radius. Rounding the lower interval endpoint downward and the upper endpoint upward, then intersecting with $[-1,1]$, preserves its conditional coverage. Both intervals cover their true parameters simultaneously at every time with probability at least $19/20$.

**Proof.** The positive exponential series gives $e^{7/10} > 1 + 7/10 + (7/10)^2/2 + (7/10)^3/6 + (7/10)^4/24 > 2$. Hence $\log 2 < 7/10$, so $\log V + 2\log(1/\delta) \le (7/10)(k+2h)$. Divide the scalar regression bound by $\sqrt{V}$. For a nonnegative rational $x$, the smallest integer $j$ with $(j/10^{12})^2 \ge x$ supplies a rational upper square-root enclosure. Integer square root and an exact final comparison compute it. Outward rounding only enlarges the interval. Intersecting with a true prior constraint preserves coverage. Apply the two-regression union bound on their time-uniform events. □

All counts, sums, radius enclosures and interval endpoints in this experiment are rational. No floating-point logarithm is trusted in the confidence arithmetic. An exact replay verifies the arithmetic and the declared protocol. It does not authenticate the observations, prove that a physical process satisfies the model, or measure the procedure's coverage from one pseudorandom run.

### 95.1 Executed result

The seeded reference path uses $a = 2/5$, $b = 1/5$ and initial observation zero. It was first run to 65,536 transitions and then extended along the same seeded path to 262,144. Time-uniform coverage permits that extension without spending another time-selection budget. This is not a new independent coverage experiment.

At the final checkpoint, outward-rounded intervals are

$$\boxed{\frac{89}{250} \le a \le \frac{231}{500}, \qquad \frac{29}{250} \le b \le \frac{29}{100}.}$$

Every member of this rectangle has density at least

$$\rho = 1 - \frac{231}{500} - \frac{29}{100} = \frac{31}{125}.$$

The rounded fitted center is $(409/1000, 203/1000)$. It lies inside the rectangle. No reset or other coefficient correction is applied.

At earlier recorded checkpoints the sufficient density-floor test withheld a whole-box certificate. Its first successful recorded checkpoint was 32,768 transitions, with floor $1/100$. This is not a claim that 32,768 is the smallest possible sample size, or even the first passing individual step. Only the declared checkpoints were evaluated. A withheld whole-box certificate does not declare the nominal fit or true process invalid.

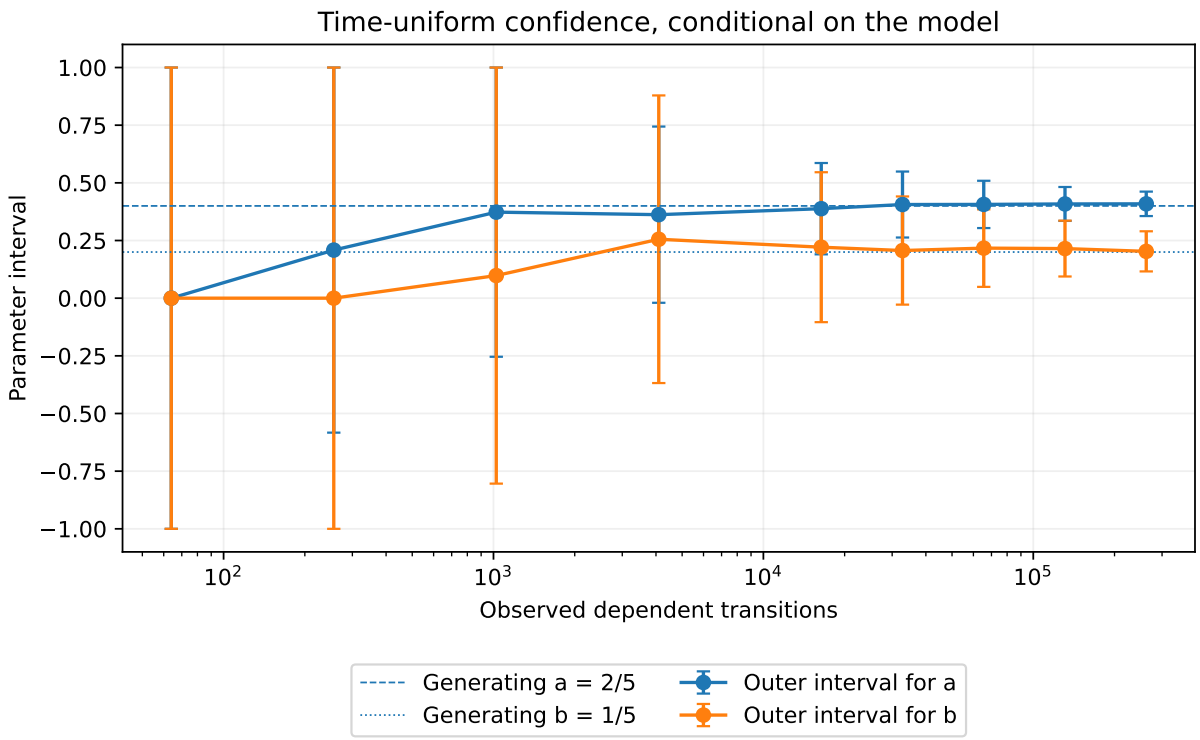



**Figure 8.3:** The intervals are from the saved five-node dependent path. The state map, two-parameter family, reference measure and predictable protocol are supplied. The learner estimates two parameters and rounds its intervals outward. The confidence guarantee is conditional and simultaneous over time. One saved path does not validate coverage empirically or authenticate the model assumptions.

## 96. Structured uncertainty has an observable error bound

Consider two parameters in a certified family. For normalized $z = (1, x, y)$, their density difference is

$$\Delta p(u) = \Delta a\, xu + \Delta b\, yL_2(u).$$

The state-update vector is the same $v(u)$ in both models after observing $u$. This does not make the models equivalent. Their likelihoods and hence their predicted observation laws can differ.

**Proposition 85. Parameter-to-law error without a state-estimation term.** Let

$$c_1 = \int |u|\, d\mu, \qquad c_2 = \int |L_2(u)|\, d\mu.$$

If $|\Delta a| \le d_a$ and $|\Delta b| \le d_b$, the conditional observation laws on any common normalized state obey

$$\boxed{\|P(\cdot \mid z) - \widehat{P}(\cdot \mid z)\|_{\mathrm{TV}} \le \Delta := \frac{c_1 d_a + c_2 d_b}{2}.}$$

The same bound holds on every realized common observation history. With the same initial observation, complete future records of length $T$ have total variation at most $\min(1, T\Delta)$. For the five-node measure,

$$c_1 = \frac{23}{45}, \qquad c_2 = \frac{14}{45}.$$

For $du/2$, the respective constants are $1/2$ and $2/(3\sqrt{3})$.

**Proof.** The first claim follows by integrating $|\Delta p|/2$, using $|x| \le 1$ and $y \le 1$. After the same observed value both models have exactly $v(u)$, so the conditional bound applies again without accumulated coordinate-state error. Telescoping conditional kernels gives the record bound, equivalently the common-history coupling argument of Proposition 69. Direct summation gives the discrete constants. For the continuous constant $c_2$, split the integral at $u = \pm 1/\sqrt{3}$ and integrate the quadratic polynomial with its respective signs. □

The use of $\Delta$ here is an upper bound proved from a confidence set and a model class. It is not the lower discrepancy bound from a rejected residual test in Chapter 6. A test that does not reject cannot supply the displayed upper bound by itself.

For the final reference rectangle and fitted center,

$$d_a = \frac{53}{1000}, \qquad d_b = \frac{87}{1000}, \qquad \boxed{\Delta = \frac{2437}{90000}.}$$

These bounds compare the fitted model with every member of the rectangle. They apply to the true process on the conditional coverage event. Even when that event fails, the rectangle's internal probability certificate remains an exact statement about its members.

## 97. Long-horizon marginal control is not full-record control

The previous record bound grows with $T$. A different question can have a uniform answer: how far apart are the distributions of the signal at a single future time? A lower density bound permits this distinction to be made exactly.

Let $P, Q$ be Markov kernels on a common state space. Suppose $P(s,\cdot) \geq \rho\mu(\cdot)$ for every $s$, where $\mu$ is one probability measure and $0 < \rho \leq 1$. Suppose also $\sup_s \|P(s,\cdot) - Q(s,\cdot)\|_{\mathrm{TV}} \leq \Delta$.

**Proposition 86. Uniform marginal and rolling-window stability.** For initial laws at distance $e_0$, the one-time marginal laws satisfy

$$\boxed{e_t \leq \min\left(1, (1-\rho)^t e_0 + \frac{\Delta[1-(1-\rho)^t]}{\rho}\right).}$$

For $k$ consecutive observations starting at time $t$, their joint laws are at distance at most $\min(1, e_t + (k-1)\Delta)$. A uniform small marginal bound does not in general bound the total variation of arbitrarily long complete records away from one.

**Proof.** For $\rho < 1$, write $P(s,\cdot) = \rho\mu + (1-\rho)R(s,\cdot)$, where $R$ is a Markov kernel. Applying a Markov kernel cannot increase total variation, by integration of bounded test functions. Thus $\|\nu P - \widetilde{\nu}P\|_{\mathrm{TV}} \leq (1-\rho)\|\nu - \widetilde{\nu}\|_{\mathrm{TV}}$. Adding and subtracting $\widetilde{\nu}P$ yields $e_{t+1} \leq (1-\rho)e_t + \Delta$. Iterate this inequality. The case $\rho = 1$ follows directly. Applying the finite-record telescoping argument from the initial time $t$ gives the rolling-window statement.

For the final assertion, take two distinct independent Bernoulli laws with parameters strictly between zero and one. Each is a Markov kernel with a positive common minorization. At any time their marginal distance is the fixed parameter difference. For a full record, test whether its empirical mean lies above the midpoint between the parameters. Under each law, Chebyshev's inequality makes the error probability tend to zero as the record length grows. The total variation of the record laws therefore tends to one. □

This is the standard minorization/contraction mechanism [R24], with its error accounting made explicit. It is not a stability theorem for arbitrary feedback policies that remember the full history. The reference model is autonomous and its current scalar observation is the Markov state. It also does not bound long products of unnormalized likelihood ratios.

With $e_0 = 0$, $\rho = 31/125$ and $\Delta = 2437/90000$, the reference confidence rectangle gives

$$\boxed{\sup_{t\geq 0} e_t \leq \frac{2437}{22320} \approx 0.10919.}$$

This is a nontrivial all-time bound on **one-time marginal observation distributions**. The complete-record bound is still $\min(1, 2437T/90000)$, which becomes trivial at integer horizon 37. The model does not exchange those two different claims. In live filtering after a shared observed value, its state is exactly known and the next-observation law has the separate uniform upper bound $2437/90000$.

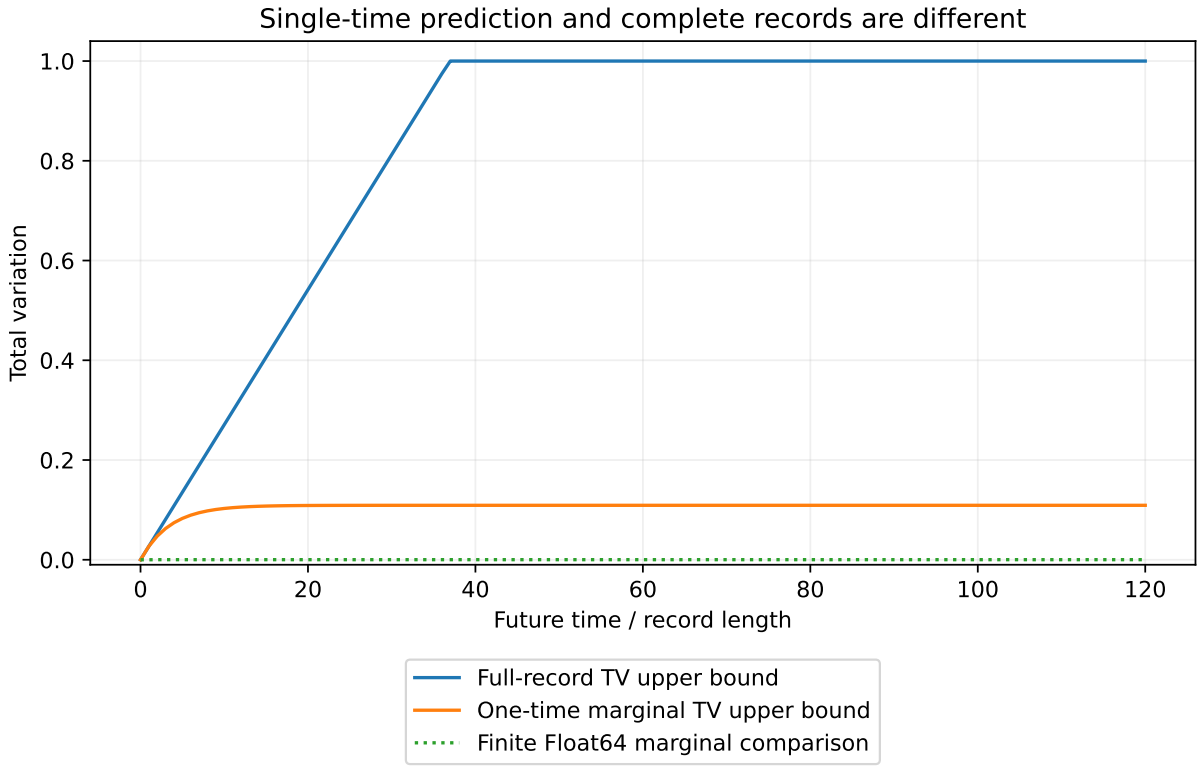



**Figure 8.4:** The full-record bound grows with record length. The one-time marginal bound remains uniformly informative only because the declared Markov family has a common minorization. The figure also shows a finite numerical marginal comparison, separately from either theorem. An all-time single-marginal bound is not an all-time complete-record bound. Common reference measure, model class, initial state and policy remain part of the comparison.

## 98. A conditional data-to-domain theorem

**Proposition 87. Observable structured learning with a covered domain.** Suppose the scalar process obeys the conditional law of Proposition 83 with the declared reference measure and fixed parameters satisfying $|a|, |b| \leq 1$. Use the predictable regressions, fixed dictionary and confidence protocol of Proposition 84. At a data-selected finite time, form the full outward-rounded rectangle. If its density floor $\rho$ is nonnegative, that rectangle has an exact all-horizon probability certificate on $C$. If $\rho > 0$, every chosen center inside it has the marginal and record bounds of Propositions 85–86 relative to every member.

With probability at least $19/20$ over the specified stochastic observation process, these implications cover its true parameter at every checked time, including a stopping time. Internal validity

of a verified rectangle is deterministic. Statistical coverage remains conditional on the declared process, moment, parameter and observation assumptions.

**Proof.** Proposition 84 provides one simultaneous event on which both true parameters belong to all returned intervals. Proposition 83 supplies the domain certificate for the entire rectangle whenever its sufficient density floor is nonnegative. The selected center is one of its members, and the maximum distances to the endpoints bound each parameter difference. Proposition 85 gives conditional and record error. If the floor is positive, all true conditional rows have the same lower reference component, so Proposition 86 applies to the scalar Markov process. Selection of the stopping time does not alter the already simultaneous event. None of these implications verifies the assumptions from a single sample path. □

No claim of empirical coverage follows from the saved run. It is one seeded finite realization of the exactly declared discrete analogue. The general confidence result is a theorem, with assumptions recorded in the certificate. Its independent checker rederives the sufficient statistics, rational radii, outer endpoints, density floor and error budgets from the saved path. It rejects a changed reference measure, a narrower claimed interval, an altered path checksum or an unsupported claim that the assumptions were independently verified.

This is the point where probability-domain certification and data uncertainty become connected in an executed dependent-signal example. The family is more restrictive than an arbitrary learned operator. That restriction is visible and does useful work: it preserves the nonlinear posterior curve while uncertainty changes its likelihoods.

## 99. Two distinct completed reference sectors

The smaller-repair construction retains the earlier operator family and geometry. Its coefficient box is deterministic, and its all-horizon validity is distinct from its finite-horizon comparison bound. The dependent-data construction learns two parameters of a different, supplied nonlinear observable family. It certifies a whole confidence rectangle without reset correction and distinguishes next-step, record and marginal bounds. Neither example establishes unrestricted representation discovery.

## 100. Scope and research boundary

The finite theory contains exact realization and transfer, closure and memory, conditional observable identification, controlled probability instruments, finite polyhedral and curved certificates, explicit approximation budgets, and a restricted dependent-data learning sector. The proof and evidence register records the assumptions, dependencies and finite code counterparts. The formal companion records the checked statements and their limits. General nonlinear dictionary discovery, unknown hidden-state learning, empirical sensor modeling and broad deployment remain research directions.

APPENDIX A

# Proof and evidence register

Every numbered proposition has a written proof in the main text. The records below summarize its local inputs and strongest claim boundary. They do not replace the statement or the definitions around it. Code references name a restricted implementation or finite illustration. Formal coverage is a separate column of evidence.

The dependency register is curated rather than extracted from Lean proof terms. The source package retains the full register, statement hashes, code locations and current coverage map. Hashes establish identity rather than mathematical truth.

## Realization and predictive closure

### P1. Exact realization and consistency

**Location and role.** Section 2. Identity. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Finite Hilbert spaces; a declared analysis/synthesis pair with a left inverse.

**Boundary.** Reconstruction is not predictive sufficiency.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P2. Coherent transport and operator realization

**Location and role.** Section 3. Identity. Principal dependencies: P1.

**Requires.** Consistent coefficient spaces; one pairwise coherent family of chart maps.

**Boundary.** Agreement across charts does not establish correctness.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P3. Complete refinement preserves reconstruction

**Location and role.** Section 4. Identity. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Orthogonal projector tree; a complete antichain.

**Boundary.** Omitted children or energy-only storage lose the reconstruction contract.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P4. Transfer reconstruction and refinement

**Location and role.** Section 4. Identity. Principal dependencies: P3.

**Requires.** Linear maps with complete input, intermediate and output projector resolutions.

**Boundary.** Nonlinear maps require interaction terms.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P5. Polynomial packet interactions

**Location and role.** Section 4. Identity. Principal dependencies: P4.

**Requires.** A specified finite multilinear polynomial expansion and complete packets.

**Boundary.** No automatic general nonlinear dictionary discovery.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P6. Exact quotient criterion

**Location and role.** Section 5. Criterion. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** A summary onto its image; deterministic transitions on the stated domain.

**Boundary.** A quotient must be constant on every summary fiber.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P7. Closure defect and compositional defect

**Location and role.** Section 5. Criterion. Principal dependencies: P6.

**Requires.** Orthogonal retained projector; linear dynamics on the declared whole state space.

**Boundary.** Small omitted energy alone does not bound future influence.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P8. Minimal linear predictive completion

**Location and role.** Section 6. Construction. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Known finite family of linear operators and initial observables.

**Boundary.** The constructed observables need not be measured or low dimensional.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P9. Coarsest joint finite closure

**Location and role.** Section 7. Construction. Principal dependencies: P12.

**Requires.** Known finite stochastic transitions; positive likelihoods; initial partition with at least two blocks.

**Boundary.** The one-block Bayesian necessity exception must be retained.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P10. Exact projected memory and a conditional tail bound

**Location and role.** Section 8. Bound. Principal dependencies: P7.

**Requires.** Linear retained/unresolved split; contraction and bounded history for the tail estimate.

**Boundary.** One-step tail error is not recursive rollout error.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P11. Ideal statistical benefit of refinement

**Location and role.** Section 9. Identity. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Square-integrable target; nested sigma algebras; exact conditional expectations.

**Boundary.** An estimated finite-data predictor can worsen after refinement.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P12. Bayesian closure and its exact defect

**Location and role.** Section 10. Criterion. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** At least two coarse blocks; strictly positive likelihoods for universal necessity.

**Boundary.** Fixed-prior defect permits nonnegative likelihoods only with positive evidence.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P13. Strong fixed-horizon refinement limit

**Location and role.** Section 12. Limit. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Bounded operator; orthogonal projections converging strongly to the identity.

**Boundary.** Fixed horizon and strong convergence, not norm or time-uniform convergence.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

## Learning and stability

### P14. Exact causal delay closure

**Location and role.** Section 15. Criterion. Principal dependencies: P6, P8.

**Requires.** Autonomous finite linear process; positive delay length; true observation-map rank plateau.

**Boundary.** An empirical rank plateau alone is not predictive-space coverage.

**Code link.** `checks/observable_reference.py`, `learn_delay`. Exact finite delay learner under supplied closure and span assumptions.

**Formal coverage.** Complete in the mapped scope.

### P15. Data-span identification and its boundary

**Location and role.** Section 16. Criterion. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Paired exact columns; input-kernel inclusion in successor kernel.

**Boundary.** Interpolation identifies the sampled span, not unseen directions.

**Code link.** `checks/observable_reference.py`, `fit_closed_span`. Exact sampled-span interpolation and kernel compatibility.

**Formal coverage.** Complete in the mapped scope.

### P16. Fixed-space identification bound

**Location and role.** Section 16. Bound. Principal dependencies: P15.

**Requires.** Fixed full-row-rank coordinates; input error below the smallest singular value.

**Boundary.** Does not learn a coordinate space or certify its rank.

**Code link.** `checks/stochastic_reference.py`, `noisy_core_bound`. Related finite fixed-core perturbation bound, not a general noisy-coordinate learner.

**Formal coverage.** Complete in the mapped scope.

### P17. Finite-prefix indistinguishability

**Location and role.** Section 17. Obstruction. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Any fixed finite observation horizon; no supplied dimension bound.

**Boundary.** A finite prefix cannot rule out a later hidden response.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P18. Compositional defect propagation

**Location and role.** Section 19. Bound. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Lipschitz maps on domains containing both compared trajectories; declared local defects.

**Boundary.** Different policies need a separate action-mismatch bound.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P19. Uniform prediction bound for a stable finite-memory truncation

**Location and role.** Section 20. Bound. Principal dependencies: P10, P18.

**Requires.** Finite-memory gain sum below one; uniformly bounded true-history defect.

**Boundary.** Spectral radius alone is not the stated induced-norm condition.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P20. Packet-interaction truncation bound

**Location and role.** Section 21. Bound. Principal dependencies: P5.

**Requires.** Declared polynomial terms; restricted multilinear norms; bounded domain.

**Boundary.** Bounds omitted terms, not the quality of an undiscovered basis.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P21. Coherent reconciliation and coordinate-stable transport

**Location and role.** Section 22. Construction. Principal dependencies: P1, P2.

**Requires.** Fixed exact atlases; normalized nonnegative weights; linear chart predictors.

**Boundary.** Consistent reconciliation need not improve accuracy.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P22. The upward-lift obstruction

**Location and role.** Section 23. Criterion. Principal dependencies: P6.

**Requires.** Coarse and proposed finer state maps; a norm for the lower bound.

**Boundary.** A genuinely finer state generally cannot be initialized from the coarse state alone.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P23. Exact accounting for an enumerated grammar

**Location and role.** Section 24. Identity. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Countable fixed grammar and prior; finite positive full and active evidence; common likelihood measure.

**Boundary.** Continuous likelihood densities need not be bounded by one.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P24. Prior-preserving program refinement

**Location and role.** Section 24. Identity. Principal dependencies: P23.

**Requires.** Prior-mass-preserving split; a whole-record likelihood mixture identity.

**Boundary.** A fixed child mixture is not independent switching at each step.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P25. Observationally grounded finite linear realization

**Location and role.** Section 25. Integration. Principal dependencies: P1, P2, P3, P4, P14, P15, P16, P18.

**Requires.** Autonomous linear truth; sufficient delay; observed windows spanning the true predictive space.

**Boundary.** The learner does not infer its model-class and coverage assumptions from a fit.

**Code link.** `checks/observable_reference.py`, `learn_delay`. Autonomous finite example only.

**Formal coverage.** Complete in the mapped scope.

## Controlled stochastic realization

### P26. Finite instrument closure

**Location and role.** Section 27. Construction. Principal dependencies: P8.

**Requires.** Known nonnegative finite instrument; all initial beliefs; exact mass row.

**Boundary.** A known-model completion is not an observable-data learner.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P27. Minimal finite behavior realization

**Location and role.** Section 28. Construction. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Causal nonnegative finite-alphabet behavior; finite full history/test rank.

**Boundary.** Minimality is within linear realizations of the declared behavior.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P28. Observable table reconstruction

**Location and role.** Section 29. Construction. Principal dependencies: P27.

**Requires.** Known rank upper bound; invertible matching core; exact shifted tables and initial tests.

**Boundary.** Finite invertibility alone supplies only a rank lower bound.

**Code link.** `checks/stochastic_reference.py`, `fit_core`. Population-table reconstruction with declared full-rank coverage.

**Formal coverage.** Complete in the mapped scope.

### P29. Canonical probability cone

**Location and role.** Section 30. Existence. Principal dependencies: P27, P28.

**Requires.** Canonical finite behavior space and all future tests.

**Boundary.** An all-test cone is not automatically finitely describable.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P30. Positive packet decomposition and its boundary

**Location and role.** Section 31. Criterion. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Complete cone-preserving packet maps; strictly positive mass functional.

**Boundary.** Signed Fourier or Haar details need not be probability components.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P31. Missing-intervention nonidentifiability

**Location and role.** Section 32. Obstruction. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** At least two possible interventions with a missing action.

**Boundary.** No amount of data restricted to one action identifies an unobserved action.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P32. Finite-table rank separation and subspace stability

**Location and role.** Section 33. Bound. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** True finite-table rank; known perturbation budget; singular gap above twice that budget.

**Boundary.** Does not exclude true directions absent from the selected table.

**Code link.** `checks/stochastic_reference.py`, `finite_subspace_report`. Finite matrix singular-subspace report.

**Formal coverage.** Complete in the mapped scope.

### P33. Noisy core-block realization

**Location and role.** Section 34. Bound. Principal dependencies: P28.

**Requires.** Fixed square invertible core and stated operator/table perturbation budgets.

**Boundary.** Good core inversion is not global positivity certification.

**Code link.** `checks/stochastic_reference.py`, `noisy_core_bound`. Fixed-core perturbation calculation.

**Formal coverage.** Complete in the mapped scope.

### P34. Normalization-aware filtering stability

**Location and role.** Section 35. Bound. Principal dependencies: P18.

**Requires.** Common declared norm; bounded true states; positive likelihood margin.

**Boundary.** Rare-event conditioning can amplify a small unnormalized error.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P35. Reset-experiment probability certificate

**Location and role.** Section 36. Confidence. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** A prospectively fixed list of experiments; independent resets per experiment; known reset law.

**Boundary.** Not a guarantee for one dependent stream or post-selected experiments.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P36. Exact normalized closure defect

**Location and role.** Section 37. Identity. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Mass-preserving summary; nonnegative instrument; two positive branch likelihoods.

**Boundary.** Posterior agreement can hide incorrect likelihoods.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P37. Conditional observation-based stochastic completion

**Location and role.** Section 39. Integration. Principal dependencies: P27, P28, P29, P30, P32, P33, P34, P35.

**Requires.** Full rank coverage; exact population core; separate perturbation and likelihood assumptions.

**Boundary.** No global fitted cone follows from the spectral estimate alone.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

## Probability-domain certificates

### P38. Test-cone hierarchy and a finite stopping certificate

**Location and role.** Section 42. Criterion. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Exact mass law; finite symbol family; all words through each depth.

**Boundary.** A finite stopping equality is sufficient, not guaranteed to occur.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P39. Finite positive-lift certificate

**Location and role.** Section 43. Certificate. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Rational/full-rank normalized generators; nonnegative lifts; initial weights; exact intertwining and mass.

**Boundary.** Certifies the supplied model, not the generating process.

**Code link.** `checks/cone_reference.py`, `verify_certificate`. Exact supplied finite-generator witness replay.

**Formal coverage.** Complete in the mapped scope.

### P40. Robust finite-domain certificate

**Location and role.** Section 44. Certificate. Principal dependencies: P39.

**Requires.** One common domain for every complete uncertainty-family vertex; exact mass constraints.

**Boundary.** Independent entry boxes can discard essential uncertainty correlations.

**Code link.** `checks/cone_reference.py`, `verify_certificate`. Vertex witnesses replayed separately in the related suite.

**Formal coverage.** Complete in the mapped scope.

### P41. Certified correction and its error budget

**Location and role.** Section 45. Bound. Principal dependencies: P39.

**Requires.** Fixed generator matrix; feasible correction; separately justified error budgets.

**Boundary.** Existence of one valid fit does not show the true model preserves that domain.

**Code link.** `checks/cone_reference.py`, `minimum_mixture_repair`. Explicit simplicial mixture sector, not general convex fitting.

**Formal coverage.** Complete in the mapped scope.

### P42. Instrument contraction before normalization

**Location and role.** Section 46. Bound. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Proper cone; strictly positive mass; cone-preserving normalized instrument.

**Boundary.** Base norm is not an arbitrary Euclidean coefficient norm.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P43. Observation-tree total-variation bound

**Location and role.** Section 47. Bound. Principal dependencies: P42.

**Requires.** Two models valid on one domain; common policy; summed instrument-error budgets.

**Boundary.** Full-law bound grows with horizon and need not remain informative.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P44. Sharp event-probability dependence of conditioning

**Location and role.** Section 48. Bound. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Two probability laws; positive evidence in both; unconditional TV bound.

**Boundary.** Conditioning costs inverse event probability.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P45. Rational finite-certificate search

**Location and role.** Section 49. Search. Principal dependencies: P39.

**Requires.** Rational inputs and existence of a rational finite-generator certificate.

**Boundary.** Semidecision guarantee is not an efficient complete implemented search.

**Code link.** `checks/cone_reference.py`, `search_interval`. Bounded two-coordinate interval search, not the general enumerator.

**Formal coverage.** Complete in the mapped scope.

### P46. A rational rank-three model with no polyhedral probability-domain certificate

**Location and role.** Section 51. Obstruction. Principal dependencies: P38.

**Requires.** Declared rational rotation/probe model; closed convex invariant domain containing its initial state.

**Boundary.** Rules out polyhedral domains here, not all finite certificates.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P47. Polyhedral approximation of the rotation model

**Location and role.** Section 52. Bound. Principal dependencies: P39, P43, P46.

**Requires.** Attenuated rotation and sufficient polygon size; common Lorentz comparison domain.

**Boundary.** Damping changes the model and its observation law.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P48. Sequential confidence accounting for certified proposals

**Location and role.** Section 53. Confidence. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Conditionally covered successive sets with a summable failure budget.

**Boundary.** Does not establish resets, rank or dictionary validity.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P49. Conditional learning with an all-horizon probability certificate

**Location and role.** Section 54. Integration. Principal dependencies: P37, P39, P40, P43, P44.

**Requires.** Declared identification model; a covering uncertainty set; robust witness for all members.

**Boundary.** A nominal witness alone establishes only internal validity.

**Code link.** `checks/cone_reference.py`, `verify_certificate`. Deterministic witness portion; statistical assumptions are external.

**Formal coverage.** Complete in the mapped scope.

## Curved domains and continuous observations

### P50. Matrix-embedding probability certificate

**Location and role.** Section 56. Certificate. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Injective symmetric-matrix embedding; strict interior witness; positive congruences; exact mass.

**Boundary.** Matrix embedding and positive lift are different constructions.

**Code link.** `checks/continuous_reference.py`, `verify_certificate`. Supplied symmetric-matrix/congruence witness sector.

**Formal coverage.** Complete in the mapped scope.

### P51. Exact rational matrix witness for the rotation model

**Location and role.** Section 57. Certificate. Principal dependencies: P46, P50.

**Requires.** The exact rational disk model and its explicit matrix-unit witnesses.

**Boundary.** No attenuation or quantum-physical interpretation is involved.

**Code link.** `checks/continuous_reference.py`, `verify_certificate`. Exact two-by-two matrix embedding example.

**Formal coverage.** Complete in the mapped scope.

### P52. Continuous-observation instrument semantics

**Location and role.** Section 58. Semantics. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Standard Borel observations; common probability reference measure; integrable positive operator density.

**Boundary.** Point densities are version dependent; zero branches have no model-determined posterior.

**Code link.** `checks/continuous_reference.py`, `VerifiedInstrument`. Finite polynomial reference class only.

**Formal coverage.** Complete in the mapped scope.

### P53. Polynomial Gram certificate for interval observations

**Location and role.** Section 59. Certificate. Principal dependencies: P50, P52.

**Requires.** Fixed embedding; positive polynomial Gram witness; exact coefficients and integrated mass.

**Boundary.** A sufficient certificate, not completeness for all positive maps.

**Code link.** `checks/continuous_reference.py`, `verify_certificate`. Sufficient polynomial positivity certificate replay.

**Formal coverage.** Complete in the mapped scope.

### P54. Conditional finite-moment identification

**Location and role.** Section 60. Construction. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Correct finite predictive dimension and degree; invertible complete event core; exact moment tables.

**Boundary.** Rank and degree are supplied assumptions, not inferred from an inverse.

**Code link.** `checks/continuous_reference.py`, `fit_population_moments`. Exact event/moment tables with supplied rank and degree.

**Formal coverage.** Complete in the mapped scope.

### P55. A valid continuous counterexample to finite-moment identification

**Location and role.** Section 61. Obstruction. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Uniform reference measure; declared two polynomial densities with matching low moments.

**Boundary.** Moment agreement and validity do not identify the entire density.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P56. Exact bin coarsening and its posterior

**Location and role.** Section 62. Identity. Principal dependencies: P52.

**Requires.** Exact cell integration; a measurable partition; bin-measurable policy.

**Boundary.** Does not preserve a policy responding to within-bin values.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P57. Exact covariance correction for an averaged congruence

**Location and role.** Section 63. Identity. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Square-integrable matrix function; a probability averaging measure.

**Boundary.** Average congruence differs from congruence of the average.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P58. Positive cell-average approximation

**Location and role.** Section 64. Bound. Principal dependencies: P52, P56.

**Requires.** Finite positive-measure cells; Lipschitz density in the base norm.

**Boundary.** This is a declared density approximation, not exact continuous-law equality.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P59. Continuous-record total-variation bound

**Location and role.** Section 65. Bound. Principal dependencies: P42, P52.

**Requires.** Common cone, mass and reference measure; common policy; integrated error budgets.

**Boundary.** No uniform posterior contraction is supplied.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P60. Exact recognition of a fixed certified template

**Location and role.** Section 67. Construction. Principal dependencies: P53.

**Requires.** A fixed certified template; coefficient intertwining; invertible alignment.

**Boundary.** Template recognition is not domain discovery.

**Code link.** `checks/continuous_reference.py`, `recognize_template`. Fixed-template recognition, not geometry discovery.

**Formal coverage.** Complete in the mapped scope.

### P61. Conditional continuous realization with a finite domain certificate

**Location and role.** Section 68. Integration. Principal dependencies: P52, P53, P54, P56, P58, P59, P60.

**Requires.** Correct continuous class and event core; an independent exact domain witness.

**Boundary.** No unmodeled measurement error or unknown-degree inference is covered.

**Code link.** `checks/continuous_reference.py`, `fit_population_moments`. Worked continuous polynomial sector only.

**Formal coverage.** Complete in the mapped scope.

## Discovery, approximation and evidence

### P62. Oriented quadratic probability certificate

**Location and role.** Section 70. Certificate. Principal dependencies: P52.

**Requires.** Positive definite ellipsoidal metric; oriented quadratic and polynomial positivity witnesses.

**Boundary.** Preserving a double cone without mass orientation is insufficient.

**Code link.** `checks/quadratic_reference.py`, `verify_quadratic`. Exact oriented quadratic witness.

**Formal coverage.** Complete in the mapped scope.

### P63. Conditional operator-only ellipsoid discovery

**Location and role.** Section 71. Search. Principal dependencies: P62.

**Requires.** Suitable mass-preserving ratio; unique invariant-form direction; nonzero rank-one image scale.

**Boundary.** Every geometry proposal still requires full-family verification.

**Code link.** `checks/quadratic_reference.py`, `discover_quadratic`. Restricted invariant-form/rank-one-image discovery.

**Formal coverage.** Complete in the mapped scope.

### P64. Valid positive polynomial approximation

**Location and role.** Section 72. Bound. Principal dependencies: P59.

**Requires.** A valid Lipschitz density on a common cone; doubly stochastic positive kernel.

**Boundary.** Avoids an exact-degree hypothesis but still assumes regularity and a domain.

**Code link.** `checks/sequential_reference.py`, `positive_coefficients`. Positive approximation in a finite symbolic example.

**Formal coverage.** Complete in the mapped scope.

### P65. Certified coefficient error and approximation bias

**Location and role.** Section 73. Bound. Principal dependencies: P64.

**Requires.** Positive estimated Bernstein coefficients; exact integrated mass; covered coefficient errors.

**Boundary.** Increasing polynomial degree does not automatically reduce estimation error.

**Code link.** `checks/sequential_reference.py`, `reconstruct_bernstein`. Finite coefficient reconstruction; no unknown-basis learner.

**Formal coverage.** Complete in the mapped scope.

### P66. Time-uniform predictable-residual certificate

**Location and role.** Section 74. Confidence. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Prospectively registered predictable bounded tests and a summable budget.

**Boundary.** Nonrejection is not an upper error or correctness guarantee.

**Code link.** `checks/sequential_reference.py`, `anytime_radius`. Declared elementary time-uniform radius.

**Formal coverage.** Complete in the mapped scope.

### P67. Covered adaptive-action evidence

**Location and role.** Section 75. Confidence. Principal dependencies: P66.

**Requires.** Known predictable randomized actions with positive coverage.

**Boundary.** Estimates contextual action means, not unobserved-policy long-run distributions.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P68. Sequential evidence gives a lower error certificate

**Location and role.** Section 76. Bound. Principal dependencies: P66.

**Requires.** Actual common-history observation laws and registered tests.

**Boundary.** A lower discrepancy certificate cannot be used as an upper accuracy bound.

**Code link.** `checks/sequential_reference.py`, `dependent_example`. One seeded conditional-mismatch illustration.

**Formal coverage.** Complete in the mapped scope.

### P69. Observation-law error without a shared state space

**Location and role.** Section 77. Bound. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Common observation/action space and policy; conditional TV discrepancies on true histories.

**Boundary.** Allows different internal state spaces but still needs true discrepancy control.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P70. Separate structural acceptance, approximation control and sequential rejection

**Location and role.** Section 78. Integration. Principal dependencies: P62, P63, P64, P65, P66, P67, P68, P69.

**Requires.** Separate structural, approximation and sequential-evidence assumptions.

**Boundary.** Acceptance, accuracy and statistical nonrejection are not equivalent.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

## Robust curved certificates

### P71. Robust oriented margin

**Location and role.** Section 80. Certificate. Principal dependencies: P52.

**Requires.** Common transformed domain; uniform oriented margin and output-error bounds; exact mass.

**Boundary.** Margins and uncertainty must be expressed in the same coordinates.

**Code link.** `checks/robust_reference.py`, `replay_certificate`. Uniform margin certificate and uncertainty-budget check.

**Formal coverage.** Complete in the mapped scope.

### P72. Boundary fragility under mass preserving perturbations

**Location and role.** Section 81. Obstruction. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** At least one spatial coordinate; a cone-positive map with a boundary image.

**Boundary.** Mass-only dimension zero is excluded.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P73. Interior repair with an explicit error budget

**Location and role.** Section 82. Bound. Principal dependencies: P59, P71.

**Requires.** A known valid reference; center-reset mixture; stated discrepancy and mass conditions.

**Boundary.** The repair changes the model; its uncertainty tube need not contain the original truth.

**Code link.** `checks/robust_reference.py`, `make_certificate`. Explicit center-reset repair witness construction.

**Formal coverage.** Complete in the mapped scope.

### P74. A posteriori proposal error

**Location and role.** Section 83. Bound. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** A true normalized invariant metric; a separated linear system; covered proposal perturbations.

**Boundary.** A small numerical residual alone is not acceptance or coverage.

**Code link.** `checks/robust_reference.py`, `propose_geometry`. Numerical proposal only; not the theorem's unknown true-metric assumptions.

**Formal coverage.** Complete in the mapped scope.

### P75. Rational semicircle and Bernstein certificate

**Location and role.** Section 84. Certificate. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Three-coordinate rational polynomial family; both boundary charts; complete subdivision.

**Boundary.** A negative Bernstein coefficient alone is not a counterexample.

**Code link.** `checks/robust_reference.py`, `replay_polynomial`. Exact complete tensor-Bernstein subdivision replay.

**Formal coverage.** Complete in the mapped scope.

### P76. Coefficient boxes and convex uncertainty

**Location and role.** Section 85. Certificate. Principal dependencies: P71.

**Requires.** Coefficient boxes in declared Lorentz coordinates; exact mass affine constraints.

**Boundary.** Jointly optimizing unknown geometry and operators is not asserted convex.

**Code link.** `checks/robust_reference.py`, `replay_certificate`. Coefficient-box inclusion in declared coordinates.

**Formal coverage.** Complete in the mapped scope.

### P77. Conditional search for strict rational curved certificates

**Location and role.** Section 86. Search. Principal dependencies: P75, P76.

**Requires.** A real transform with strictly interior initial states and a common strict margin.

**Boundary.** Conditional existence of rational witnesses is not an implemented general optimizer.

**Code link.** `checks/robust_reference.py`, `make_certificate`. Finite candidate construction; no complete general search implemented.

**Formal coverage.** Complete in the mapped scope.

### P78. Data-selected geometry with honest coverage

**Location and role.** Section 87. Confidence. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** One covered fixed-coordinate operator/initial-state set; verifier quantifies over its entire set.

**Boundary.** Data-selected geometry must transport uncertainty, not just the fitted center.

**Code link.** `checks/robust_reference.py`, `replay_certificate`. Deterministic set certificate only, not upstream statistical coverage.

**Formal coverage.** Complete in the mapped scope.

### P79. Certified repair with a retained comparison budget

**Location and role.** Section 89. Integration. Principal dependencies: P59, P71, P73, P75, P76.

**Requires.** Worked deterministic perturbed family; nominal and comparison-domain witnesses.

**Boundary.** The original repair's bound is trivial at horizon eight.

**Code link.** `checks/robust_reference.py`, `replay_certificate`. Nominal and comparison-domain witnesses for one deterministic example.

**Formal coverage.** Complete in the mapped scope.

## Data-supported domains

### P80. Convex repair bracketing

**Location and role.** Section 91. Bound. Principal dependencies: P75, P79.

**Requires.** Fixed geometry, reset path and requested margin; exact separator and upper witness.

**Boundary.** Brackets this repair path, not optimal geometry or all repair methods.

**Code link.** `checks/robust_reference.py`, `replay_certificate`. Upper certificate; exact lower separator is in the related suite.

**Formal coverage.** Complete in the mapped scope.

### P81. Time-uniform regression confidence

**Location and role.** Section 92. Confidence. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Predictable regressors; correct fixed conditional mean; positive sub-Gaussian scale; parameter bound.

**Boundary.** Estimated hidden regressors and response-dependent features need new analysis.

**Code link.** `checks/data_certificate_reference.py`, `radius`. Scalar specialization only, not general vector online regression.

**Formal coverage.** Complete in the mapped scope.

### P82. Sequential observable-moment operator sets

**Location and role.** Section 93. Confidence. Principal dependencies: P81.

**Requires.** Exact accessible predictive state; correct finite basis; known invertible moment matrix.

**Boundary.** General multirow theorem is not a hidden-state discovery algorithm.

**Code link.** `checks/data_certificate_reference.py`, `sufficient_statistics`. Observable scalar specialization; general matrix learner not implemented.

**Formal coverage.** Complete in the mapped scope.

### P83. Structured curved-domain certificate

**Location and role.** Section 94. Certificate. Principal dependencies: No principal internal prerequisite recorded.

**Requires.** Specified moment feature curve; known reference moments; two-parameter conditional family.

**Boundary.** Geometry, dictionary and family are supplied, while parameters can be learned.

**Code link.** `checks/data_certificate_reference.py`, `probability_row`. Fixed two-parameter family and five-node measure.

**Formal coverage.** Complete in the mapped scope.

### P84. Exact rational outer confidence intervals

**Location and role.** Section 95. Confidence. Principal dependencies: P81.

**Requires.** Exact observed node path; declared measure and prior bounds; rational outward rounding.

**Boundary.** Replay cannot authenticate data or prove model assumptions from one path.

**Code link.** `checks/data_certificate_reference.py`, `confidence_box`. Exact rational scalar intervals and outward rounding.

**Formal coverage.** Complete in the mapped scope.

### P85. Parameter-to-law error without a state-estimation term

**Location and role.** Section 96. Bound. Principal dependencies: P69, P83.

**Requires.** Covered parameter distances; common measure and common observed history.

**Boundary.** Likelihood error remains even when normalized posterior states agree.

**Code link.** `checks/data_certificate_reference.py`, `make_certificate`. Finite parameter rectangle and declared measure constants.

**Formal coverage.** Complete in the mapped scope.

### P86. Uniform marginal and rolling-window stability

**Location and role.** Section 97. Bound. Principal dependencies: P69.

**Requires.** Markov kernels; positive common minorization; uniform one-step discrepancy.

**Boundary.** Uniform marginal control is not uniform complete-record control.

**Evidence.** Written proof. The related suite contains finite examples for the related group of results; no dedicated general algorithm is claimed for this result.

**Formal coverage.** Complete in the mapped scope.

### P87. Observable structured learning with a covered domain

**Location and role.** Section 98. Integration. Principal dependencies: P81, P83, P84, P85, P86.

**Requires.** Correct observable family; predictable protocol; full confidence rectangle and valid chosen center.

**Boundary.** Does not transfer statistical coverage to the unrelated repaired-operator example.

**Code link.** `checks/data_certificate_reference.py`, `replay_certificate`. Data fingerprint, statistics, radii and whole-rectangle validity; assumptions not inferred.

**Formal coverage.** Complete in the mapped scope.

APPENDIX B

# Formal companion

## What the archive checks

The formal companion contains 254 proof modules and 255 compilation targets, including its root. The inventory records 4,454 named declarations and 2,887 named theorems: 2,556 supporting results and 331 example results. These counts describe formal declarations, separately from the 87 numbered manuscript propositions.

Fresh verification passed 697 stages: 255 compilation targets, one exact example execution, one full axiom audit, 254 module kernel replays, 183 isolated rejection controls and three Python checks. The 694 Lean stages admit only propext, Classical.choice and Quot.sound. Pinned upstream libraries were not exhaustively rebuilt or replayed.

The current verification records distinguish fresh kernel checks from inherited evidence. The source package retains both. The manuscript correspondence and internal proof review are separate from kernel acceptance, and no independent external referee or second proof assistant is claimed.

## Coverage at a glance

| Evidence unit | Current scope |
|---|---|
| Written manuscript results | 87 propositions and proofs |
| Complete formal mappings | 87 propositions |
| Partial formal mappings | 0 propositions |
| Without formal mappings | 0 propositions |
| Named Lean theorems | 2,887 supporting and example theorems |
| Rejection controls | 183 isolated checks |

All 87 numbered proposition cores have complete reviewed formal mappings. No proposition core remains partial or unformalized. Surrounding explanatory prose and additional unnumbered claims are not thereby separately formalized. The final scope notes state the principal assumptions and boundaries.

The following scope notes are reproduced from the coverage map. They specify which claims and adjacent discussions are included. Definitions implemented using classical choice are exact

mathematical constructions, and may be noncomputable. They should not be read as a deployed numerical learner.

### P1. Exact realization and consistency

Algebraic realization works on arbitrary types. Norm estimates use continuous linear maps on normed spaces. The divided lower bound explicitly assumes the observation space is nontrivial, so the synthesis norm is positive. No orthogonality is inferred.

**Source modules.** `Realization`, `Normed`.

### P2. Coherent transport and operator realization

Total operations on consistent coefficient subtypes. A reference chart is explicitly supplied, making the family nonempty. Partial-map admissible-domain extensions mentioned in surrounding prose are not covered.

**Source modules.** `Realization`.

### P3. Complete refinement preserves reconstruction

Finite rooted structural trees with nonempty finite child families, arbitrary mixed-depth selections, and the actual condition of exactly one selected node on each full root-to-leaf path. Local parent=sum(children) gives global operator/signal reconstruction. Local sibling-range orthogonality gives the squared-norm identity over real or complex inner-product spaces. A self-adjoint zero-product lemma supplies that orthogonality from projector assumptions. Antichain completeness and energy conservation are conclusions, not assumed aggregate equations. No general best-basis optimization is included.

**Source modules.** `PacketTree`, `TreeEnergy`.

### P4. Transfer reconstruction and refinement

Actual linear maps between possibly different spaces, with explicit finite completeness and intermediate idempotence. The algebra works under weaker assumptions than orthogonality.

**Source modules.** `Transfer`.

### P5. Polynomial packet interactions

All ordered multilinear interactions, including repeated packet indices. The finite polynomial sum retains every cross term. No general nonlinear dictionary discovery is claimed.

**Source modules.** `Polynomial`.

### P6. Exact quotient criterion

Reduced-map existence is constructed from fiber constancy, not assumed. Uniqueness is on the actual image, and action-indexed closure is included.

**Source modules.** `PredictiveClosure`.

### P7. Closure defect and compositional defect

Actual linear maps and continuous-linear-map operator norm. Idempotence is explicit, and orthogonality is not needed for the algebraic identity. Closure is not assumed in order to prove its criterion.

**Source modules.** `LinearClosure`, `Normed`.

### P8. Minimal linear predictive completion

Known linear dynamics on a finite-dimensional vector space over a field. Iteration adds actual dual-map images of observable row subspaces. Stabilization is proved by strict dimension growth within dim(X)-dim(W0), with persistence after equality and least common invariant extension. Evaluation and arbitrary-basis state maps, linearly independent rows, linear readout H and action transitions B are constructed. Any competing linear summary retaining the initial output and closing under the dynamics contains this completed row space. The subspace/basis construction is noncomputable Lean mathematics, not an implemented numerical row-reduction learner or a proof that new observables are measurable from the original data. The surrounding single-operator Cayley-Hamilton discussion is not an additional coverage claim.

**Source modules.** `InvariantCompletion`, `PredictiveCompletion`.

### P9. Coarsest joint finite closure

Finite state space with actual equivalence-relation partitions, known row-stochastic action kernels and positive observation likelihoods. The mandatory initial quotient is nontrivial for semantic minimality. The construction first splits by likelihood vectors, then by aggregate transition probabilities into current blocks. Strict refinements increase quotient cardinality, giving stabilization within card(states)-card(initial likelihood-split quotient) rounds; equality persists. Actual quotient kernels and positive likelihoods are constructed. Their prediction equation holds on all real vectors and their Bayesian equation on probability priors. Semantic JointlyClosed means existence of such stochastic quotient data, not an assumed structural condition. Its equivalence to lumpability plus likelihood constancy uses P12 and yields the unique coarsest admissible refinement. The exact real/setoid construction is noncomputable and does not implement a finite-precision partition learner or identify unknown dynamics.

**Source modules.** `FinitePartitions`, `JointRefinement`, `FiniteStochastic`.

### P10. Exact projected memory and a conditional tail bound

Exact recurrence elimination, lag reindexing, actual projected blocks, continuous-map bridge, same-history geometric truncation bound and initial-term bound. Nonnegative radius/rho and rho<1 are explicit for the bound. H<=t for full-minus-truncated history; the omitted Ico tail also handles an empty interval. No recursively predicted rollout stability is inferred.

**Source modules.** `Memory`, `MemoryBounds`.

### P12. Bayesian closure and its exact defect

Finite real probability vectors and a surjective map onto the actual coarse blocks. With at least two coarse blocks and strictly positive fine and declared coarse likelihoods, commutation for every normalized

nonnegative prior is equivalent to a common positive likelihood scale. Necessity is proved using two-point priors and a cross-block bridge, not assumed as fiber constancy. The one-block exception is proved separately. The effective-likelihood and matrix-defect identities include zero-prior-mass blocks. The fixed-prior algebra assumes nonnegative prior entries and nonzero fine/coarse evidences, so in particular includes nonnegative likelihoods with positive evidences. Totalized real division at zero is not interpreted as a valid posterior. Matrix null direction for a common likelihood scale is also proved.

**Source modules.** `BayesianClosure`, `BayesianDefect`.

### P13. Strong fixed-horizon refinement limit

Strong convergence is proved for actual continuous linear maps with a uniform operator bound. The compression estimate and fixed-power induction are derived. Actual orthogonal subspace projections provide the contractivity premise, for real or complex inner-product spaces. Each horizon is fixed; neither operator-norm convergence nor uniform-in-time accuracy is asserted. The fitted-operator extension retains its independent uniform-bound and strong-error assumptions.

**Source modules.** `StrongRefinement`.

### P14. Exact causal delay closure

Positive delay length. Actual observation maps, not empirical ranks. Kernel factorization, range-valued transition, uniqueness, rank-nullity equivalence, dimension-bound stabilization and warm-up-indexed future output are derived over finite-dimensional field vector spaces. The zero-dimensional case does not license a last block for h=0.

**Source modules.** `DelayClosure`.

### P15. Data-span identification and its boundary

Actual sample and successor linear maps. Kernel compatibility is equivalent to ambient extension, all fits agree on the sampled span, and span invariance is equivalent to successor-image inclusion. The orthogonal Moore-Penrose inverse is constructed using quotient factorization and orthogonal projections, with all four Penrose conditions proved. A proper subspace admits distinct extensions. Classical choices support a symbolic real operator, not an executable floating-point algorithm.

**Source modules.** `DataSpan`, `Pseudoinverse`.

### P16. Fixed-space identification bound

Finite-dimensional real Hilbert spaces, with the spectral operator norm. The smallest row singular value is represented by the infimum of adjoint norm ratios, proved equal to the infimum on unit vectors. Strict input error below this value derives surjectivity and the norm of the constructed orthogonal pseudoinverse. The full residual identity and divided bound are derived. Coordinates and dimension are fixed. No fitted-basis or statistical error is absorbed without a bound.

**Source modules.** `IdentificationBound`.

## P17. Finite-prefix indistinguishability

For every natural horizon T, two explicit rational systems on Fin (T+2) have identical outputs at all times through T and different output at T+1. Their transition powers at T+2 vanish. No bound on the unknown hidden dimension is assumed.

**Source modules.** `FinitePrefix`.

## P18. Compositional defect propagation

The theorem permits genuinely time-indexed normed state spaces and representations. It defines local residuals from the true transition and approximate step, derives the one-step error, and proves the manuscript suffix-product/sum bound. Lipschitz constants are nonnegative; both true and predicted inputs must lie in the declared domain. Output decoding, constant contraction and exact closure-versus-fit decomposition are separate theorems. An arbitrary changed policy is not automatically covered.

**Source modules.** `DefectPropagation`.

## P19. Uniform prediction bound for a stable finite-memory truncation

The predictor uses its own recursively generated history, with integer time indices and an explicit initialization window. A strict norm-sum gain below one yields the actual supremum error bound. Coefficient, geometric memory-tail and initialization-forcing budgets are derived and inserted into the recursive theorem. Zero prehistory handles H=0 and t<H without natural-number subtraction artifacts. Retaining the exact forcing removes its contribution. The condition is sufficient, not necessary; failure of this certificate is not proof of instability.

**Source modules.** `MemoryStability`, `MemoryApproximation`.

## P20. Packet-interaction truncation bound

Continuous multilinear maps retain ordered interaction tuples and their multiplicities. Restricted product-of-packet-range norm bounds yield the exact omitted sum and its amplitude majorant. Telescoping yields the retained Lipschitz bound on the declared ball. Contractivity is explicitly derived for actual orthogonal subspace projectors. The finite-dimensional real Hilbert sector is included in a stronger normed-space algebraic theorem. A separate bound is required for an unexpanded nonpolynomial law and for keeping trajectories inside the domain.

**Source modules.** `PacketApproximation`, `ReconciliationHilbert`.

## P21. Coherent reconciliation and coordinate-stable transport

Linear atlases act on the actual analysis-range subtype. A genuine normed additive group and normed space are constructed by pulling back synthesis; exact transports are isometries and every Lipschitz constant is equivalent to its object-space counterpart. Repaired maps are consistent and intertwine. Weighted variance and unique minimization are proved for the actual squared Frobenius objective, with a constructed orthonormal matrix coordinate equivalence in finite-dimensional real Hilbert space. The target-error inequality uses actual continuous-operator norms and nonnegative weights summing to one.

Agreement is not correctness. The adjacent learned singular-value/pseudoinverse discussion is not claimed as a separate checked theorem.

**Source modules.** `AtlasReconciliation`, `WeightedReconciliation`, `ReconciliationHilbert`.

## P22. The upward-lift obstruction

Existence/uniqueness between actual images, the linear kernel criterion, and the half-distance worst-case norm lower bound. An onto hypothesis in the linear-map construction means that its codomain is the true coarse state space.

**Source modules.** `PredictiveClosure`, `LinearRefinement`, `Normed`.

## P25. Observationally grounded finite linear realization

Conditional finite observable realization. A covering table and true successors recover the unique map by the data-only pseudoinverse formula. Delay-range identification, every iterate, causal future output, lossless atlas recoding and complete packet synthesis are linked. The actual noisy fitted predictor feeds the product-sum error envelope, a contraction radius, and a decoder bound. Coverage and model assumptions remain supplied. No unrestricted hidden-state discovery or executable production-code verification.

**Source modules.** `DelayClosure`, `IdentificationBound`, `LearnedRealization`.

## P26. Finite instrument closure

Known instrument on a finite nonempty hidden state type with finitely many outcomes. Actual nonnegative matrices obey exact per-action column-mass conservation. The P8 iteration from the nonzero mass row stabilizes within card(hidden)-1 steps and is proved equal to the span of all chronological word tests. A finite basis is extended/reindexed to place mass literally in its first coordinate. Its observation map has independent rows, is surjective onto the full coordinate space, has precisely the completed row space, and supplies unique reduced transitions. Word probabilities agree for every initial vector; probability and posterior meaning are restricted to realizableBeliefs and positive evidence. Minimality is against every linear summary with linear readouts of all words on every prior. All constructions using arbitrary real bases are noncomputable. This is not observable-data learning, minimum dimension for one fixed initialization, or a general cone-search theorem.

**Source modules.** `FiniteInstrument`, `InstrumentCompletion`, `InvariantCompletion`, `InstrumentCoordinates`, `InstrumentRealization`.

## P27. Minimal finite behavior realization

The behavior space is the actual real linear span of history/test columns on finite chronological words. Branch operators are constructed by restriction of test shifts, rather than postulated well-defined coordinate maps. Rank means finite dimension of this full column space, not a sampled matrix rank. Every finite linear realization has at least that dimension. Reachable and observable realizations are explicitly isomorphic, intertwining the initial vector, all operators and mass; the intertwiner is unique. Conditional predictions use the true word/history ratio on positive histories. Nonnegativity and causal normalization are separate supplied word-law assumptions. Infinite-record probability measures and the auxiliary

supremum-of-finite-submatrix characterization are not additional claims of this mapping. No model is learned from a finite noisy sample here.

**Source modules.** `BehaviorSpace`, `BehaviorMinimality`, `BehaviorProbability`.

## P28. Observable table reconstruction

Finite core coordinates are actual selected test evaluations. An invertible square population matrix gives surjectivity and a lower rank bound; an explicit global finite behavior-rank upper bound supplies injectivity and full coverage. The formula G_s H^{-1} is an actual matrix expression using the correct inverse side. It yields the full chronological state, mass and probability for every word, a reachable observable model and minimal dimension against every finite linear realization. A distinguished selected empty test is the literal mass coordinate (index zero in Fin r). Normalized updates require positive evidence for probabilistic interpretation. Finding the core, deriving an unknown global rank bound and statistical estimation of table entries are not claimed. Noncomputable inverse/space constructions do not amount to an executed numerical data learner.

**Source modules.** `ObservableCore`, `ObservableRealization`.

## P29. Canonical probability cone

The actual canonical behavior space is the submodule span of history/test functions from P27. Causal consistency and nonnegative behavior are supplied as CausalLaw. The all-test cone is defined by every finite test, not by finitely sampled trajectories. Cone closure under addition and nonnegative scaling, convexity, pointedness, generation by history columns, branch invariance, strict positive mass, the per-action mass identity, normalized posterior closure, and test-coordinate bounds are proved. Closedness and compact normalized-base results use the actual inherited Hausdorff function-space topology and an explicit finite-dimensional behavior-space assumption. Compactness follows through a proved closed embedding into the compact all-test unit cube; boundedness of the desired base is not assumed. The selected empty test is the literal mass-one coordinate. No finite inequality description or learned cone algorithm is claimed for the general canonical space.

**Source modules.** `TestTopology`, `CanonicalCone`.

## P30. Positive packet decomposition and its boundary

The mixture theorem works over real modules, a cone closed under nonnegative scaling, a strictly positive mass functional on nonzero cone elements, and complete linear packet maps preserving that cone. Weight nonnegativity, their unit sum, normalized component membership and mass, vanishing zero-weight components and the positive-weight finite mixture identity are proved separately. The orthant result uses actual finite real matrices: preservation of all nonnegative vectors is equivalent to entrywise nonnegativity; completeness forces zero off-diagonals; idempotence forces diagonal entries 0 or 1; every coordinate has exactly one selected packet. Reconstruction or orthogonality alone does not imply positivity. This theorem does not supply a learned cone or formalize the compactness theorem of P29.

**Source modules.** `PositivePackets`.

## P32. Finite-table rank separation and subspace stability

Full finite-table P32 for real finite-dimensional inner-product maps, hence rectangular real matrices with their Euclidean norms. Actual sorted, zero-padded singular values satisfy the all-index operator perturbation bound. The leading-r left subspace is constructed as images of ordered right Gram modes and proved to be the span of orthonormal left Gram modes; its dimension and exact residual norm are derived. Positive true rank and noise $\le\varepsilon<\sigma_r(X)/2$ imply rank separation and projection gap $\le 2\varepsilon/\sigma_r(X)<1$ without supplied inverse, tail or projection-gap premises. For arbitrary orthonormal bases of the true and learned ranges, the actual cross-matrix minimum singular value is $\ge\mathrm{sqrt}(1-(2\varepsilon/\sigma_r(X))^2)$, estimated coordinates are injective, and the explicit ordinary matrix inverse decoder $U(\hat{U}^{T}U)^{-1}$ has both inverse laws and the stated norm bound. The maximum possible rank convention is proved using zero padding. This does not certify that a finite table captures the full process.

**Source modules.** `SubspaceCoordinates`, `OneSidedSubspace`, `NoisySubspaceResidual`, `SpectralLowerBound`, `SpectralPseudoinverse`, `SpectralSharpness`, `SpectralSubspaces`, `SpectralTail`, `SingularPerturbation`, `LeftSingularModes`, `NoisySpectralSubspace`, `BasisCoordinates`, `BasisDecoder`, `DecoderMatrixFormula`, `FiniteTableP32`.

## P33. Noisy core-block realization

Finite-dimensional real inner-product square core. The least singular value uses the inherited variational rowGain definition (infimum of adjoint norm ratios; inherited unit-sphere equivalence), consistently with P16. The constructed pseudoinverse is shown to be a two-sided inverse under the strict input error gap. Both exact error identities, the true-operator error bound, fitted-operator bound, and checked fitted-gap denominator bound are proved. Exact arithmetic and fixed core only; no data-dependent coverage or floating-point solve guarantee.

**Source modules.** `CoreStability`.

## P34. Normalization-aware filtering stability

Arbitrary real normed state space and continuous mass functional. Prediction error is derived from actual operators and states. True branch has positive mass, true state and posterior obey the supplied radius, and the strict dual-norm error budget guarantees positive fitted mass. Exact normalized residual identity and both denominator bounds are proved. Actual true and fitted recurrences yield the product-sum bound and a uniform contraction estimate. Checked realized fitted denominators do not certify all branches or probability-valid fitted states.

**Source modules.** `FilteringStability`.

## P36. Exact normalized closure defect

The vector identity is on real coordinate functions with an actual linear summary and mass-preservation law; the fine and proposed branch evidences are strictly positive. The separate likelihood identity holds without division. The oblique map is an actual rank-one correction of the identity. The converse requires a nonzero entrywise-nonnegative fine branch and agreement of likelihoods AND normalized posteriors on every strictly positive prior. Explicit uniform and perturbed-uniform priors prove that this interior family spans the fine space. No equality of unnormalized operators is assumed in that converse. A one-state

instrument counterexample proves that posterior agreement alone is insufficient. The idempotent-projector specialization gives the omitted-to-retained block. The statements do not certify a fitted probability domain or control statistical error.

**Source modules.** `NormalizedDefect`, `PositivePriors`, `Block4Examples`.

### P38. Test-cone hierarchy and a finite stopping certificate

The finite-level cones are actual inequalities for all words of length at most h. Their successor/preimage identity, antitonicity, intersection, largest invariant-set property and exact adjacent-level stopping implication are proved. A noncomputable finite set of linear halfspaces is constructed at every level for finite symbol types, establishing the polyhedral description rather than assuming it. An observability row-span hypothesis is connected to injectivity of all-test evaluation. With the explicit per-action mass law, injectivity yields pointedness and strictly positive mass; membership of reachable vectors and their full span yields generation. Finite stabilization is sufficient, not asserted for every family. A one-step acceptance test is not treated as invariance. The cone may be improper without the stated observability/reachability conditions.

**Source modules.** `TestCones`, `FiniteTestRows`.

### P39. Finite positive-lift certificate

The cone is the actual image of nonnegative finite coefficient vectors under V. Normalization is lambda composed with V equals the coefficient-sum functional; surjectivity is the full-row-rank condition when generation is asserted. Cone closedness, convexity, pointedness, strict mass and generation are derived, including for redundant generators. The normalized base is proved equal to the image of the finite probability simplex and to the actual convex hull of the generators, and is compact. Invariance is equivalent to existence of an entrywise nonnegative matrix A with B V = V A, with the necessity witness constructed from the generator images by classical choice. Lifted mass conservation, the actual hidden instrument, intertwining for every chronological word and word probabilities follow from the finite witness and a nonnegative normalized initial coefficient vector. No unique lift is claimed for redundant V. The adjacent Farkas dual-certificate discussion is not claimed as an additional formal theorem.

**Source modules.** `PositiveLift`, `PositiveLiftTopology`.

### P40. Robust finite-domain certificate

The uncertainty set is the actual convex hull of complete operator families with shared convex coefficients across all branches. Nonnegative vertex lifts and exact mass conservation imply every family in that hull preserves the fixed generator cone and mass. Explicit mixed lifts are constructed. History-dependent selection of a family is handled through an actual adaptive word recursion and finite observation-tree mass recursion under any deterministic history-based action policy. Cone membership, nonnegative word likelihood, the unit upper bound and tree normalization are proved for arbitrary finite horizons. A separate theorem uses actual continuous linear operators, their operator norms and a complete facet description of the same cone to prove the stated norm-ball margin implies perturbed invariance and existence of a positive lift. The normed statement specializes to Euclidean coordinates. Exact mass constraints remain separate; independently mixing branches is not inferred to conserve mass. No statistical coverage of the uncertainty hull is asserted.

**Source modules.** `RobustFiniteDomains`, `FacetRobustness`.

## P41. Certified correction and its error budget

A FeasiblePoint records the actual finite nonnegative-lift, intertwining, per-action mass and initial-prior constraints. Internal family validity, normalized initial membership and every finite-word probability are derived from those constraints and normalized generators, not included as assumed conclusions. An actual true member of a confidence set with a fixed-domain witness constructs a feasible point in that set. The two-center bound norm(corrected-truth) <= 2 epsilon is proved for an arbitrary seminormed additive group, hence applies both to continuous linear operators and to initial states in their specified norms. No optimizer, empirical confidence set, or implication from one feasible point to coverage/validity of every point is supplied. Fixed-domain witness construction is classical and noncomputable. The adjacent explicit mixing-repair formula is outside this numbered proposition's mapping.

**Source modules.** `CertifiedCorrection`.

## P42. Instrument contraction before normalization

The base norm is the actual infimum over decompositions x=u-v with u,v in the declared closed, generating cone. Nonnegative mass, existence of decompositions, the infimum bounds, triangle inequality, negation invariance and absolute scalar homogeneity are proved. In finite-dimensional real normed spaces, compactness of the normalized base and a bound by the ambient norm are derived from cone closedness and strict mass; nondegeneracy supplies an actual AddGroupNorm, not an assumed norm. Instrument means cone preservation and exact total mass, not assumed contraction. Summed contraction is proved by the positive-decomposition infimum. For a normalized base equal to a nonempty finite convex hull, the discrepancy supremum is attained at a generator. For two valid instruments it is at most two. The finite positive-lift domains of the previous block instantiate this construction. Invertible synthesis gives the inverse-coordinate l1 formula; redundant synthesis gives only an upper bound for any chosen signed coefficients. The adjacent convex-hull unit-ball identity and general LP minimum-attainment assertion are not separately mapped as additional results. These are noncomputable exact real constructions, not a numerical norm minimizer.

**Source modules.** `BaseNorm`, `BaseNormTopology`, `InstrumentContraction`, `BaseNormModels`.

## P43. Observation-tree total-variation bound

The record space, chronological history update, unnormalized record states, record probabilities and total mass are constructed explicitly for each finite horizon. The summed base-norm error is proved by induction, not assumed as a recurrence conclusion. The actual finite total variation is half the sum of absolute probability differences. Arbitrary shared history-dependent branch families, and shared deterministic or randomized history-dependent action policies over finite alphabets, are covered. Complete per-action instruments are multiplied by the same policy probabilities in the two compared models. Joint action/observation laws are proved normalized; observation-only laws are their actual aggregation and inherit the bound. The P42 all-state discrepancy supremum supplies the pointwise budget. The bound retains the initial-state discrepancy and sum of per-time budgets, with min(1, ...) clipping. No positive lower branch likelihood or action propensity is assumed. The norm, cone, mass row,

observation alphabet and policy are common. No comparison of separately optimized or different policies, infinite-horizon accuracy, continuous observations or learned statistical coverage is claimed.

**Source modules.** `ObservationTree`, `PolicyComparison`, `FiniteTV`.

## P44. Sharp event-probability dependence of conditioning

Finite distributions are actual nonnegative normalized real vectors. Event restriction, event mass and normalized conditional distribution are defined by finite sums. The exact restriction/complement budget and event-mass difference bound are proved. For strictly positive event masses in both laws, conditioning has total variation at most min(1, delta/true_event_mass). If delta is strictly below the true event mass, positivity in the comparison law is derived, not assumed. Aggregating conditional records also obeys the bound. The factor is sharp for every 0<r<=1 and 0<=d<=r, using explicit three-state priors with common event mass r, unconditional distance d and conditional distance d/r. A concrete impossible-event expression is not treated as a prior. No bound independent of event mass or general measure-theoretic conditioning is claimed.

**Source modules.** `FiniteTV`, `Conditioning`, `SharpConditioning`.

## P45. Rational finite-certificate search

Finite rational input matrices and exact per-action mass identities. The input certificate only assumes rational normalized full-row-rank generators; its nonnegative lift and initial coefficients may be real. Finite weak rational Fourier–Motzkin elimination proves rational witnesses exist, including boundary faces. A rational right inverse is derived. Computable natural-number decoding enumerates every finite width and rational witness, exact rational checks are sound, and some finite fuel returns a certificate whenever the stated one exists. Accepted witnesses instantiate the closed, pointed, generating invariant cone and all-word bounds of P39. No strict margin, practical complexity, universal existence, or finite-failure impossibility is claimed.

**Source modules.** `RationalFeasibility`, `RationalCertificateSearch`, `RationalCertificateSoundness`.

## Finite-table spectral stability

Proposition 32 is complete in its stated finite-table scope. The selected subspace is constructed from the leading right Gram modes of the observed map and proved to be the span of orthonormal left singular modes. Its dimension and exact residual norm are derived. The projection bound and ordinary inverse decoder follow from the noise hypotheses, rather than from an assumed desired conclusion.

The actual Mathlib singular values are indexed from zero and padded with zero. Thus the manuscript's $\sigma_r$ corresponds to index $r-1$, and $\sigma_{r+1}$ corresponds to index $r$. The maximum-rank rectangular boundary is proved using this zero convention.

The two complete source modules printed below are endpoints of that development. Their imports lead to the supporting proofs in the source package. The first states the spectral and projection endpoint. The second states basis conditioning, coordinate injectivity and both decoder inverse laws. The decoder formula uses the ordinary nonsingular matrix inverse. Neither module establishes coverage of every mode of an unknown stochastic process.

### Spectral endpoint

```
import KomorebiLSM.SingularPerturbation
import KomorebiLSM.LeftSingularModes

set_option autoImplicit false
namespace Komorebi.LSM.Spectral
variable {E F : Type*} [NormedAddCommGroup E] [InnerProductSpace ℝ E]
variable [NormedAddCommGroup F] [InnerProductSpace ℝ F]
variable [FiniteDimensional ℝ E] [FiniteDimensional ℝ F]

/-- Data-selected left subspace, with the true table's rank as truncation level. -/
noncomputable def learnedLeft (X N : E →L[ℝ] F) : Submodule ℝ F :=
  leadingLeft (X+N).toLinearMap (Module.finrank ℝ X.toLinearMap.range)
    (LinearMap.finrank_range_le X.toLinearMap)

theorem noisy_rank_lower (X N : E →L[ℝ] F)
    (hr : 0 < Module.finrank ℝ X.toLinearMap.range) (ε : ℝ)
    (hnoise : ‖N‖ ≤ ε) (hsmall : ε < leastPositive X.toLinearMap / 2) :
    Module.finrank ℝ X.toLinearMap.range ≤ Module.finrank ℝ (X+N).toLinearMap.range := by
  have hl := noisy_singular_lower X N ε hnoise (Module.finrank ℝ X.toLinearMap.range-1)
  have hp := leastPositive_pos X.toLinearMap hr
  change leastPositive X.toLinearMap - ε ≤ _ at hl
  have hy : 0 < (X+N).toLinearMap.singularValues (Module.finrank ℝ X.toLinearMap.range-1) := by
    linarith
  have hh := (X+N).toLinearMap.singularValues_pos_iff_lt_finrank_range.mp hy
  omega

theorem learnedLeft_finrank (X N : E →L[ℝ] F)
    (hr : 0 < Module.finrank ℝ X.toLinearMap.range) (ε : ℝ)
    (hnoise : ‖N‖ ≤ ε) (hsmall : ε < leastPositive X.toLinearMap / 2) :
    Module.finrank ℝ (learnedLeft X N) = Module.finrank ℝ X.toLinearMap.range :=
  leadingLeft_finrank _ _ _ (noisy_rank_lower X N hr ε hnoise hsmall)
```

```
theorem learnedLeft_observed_tail (X N : E →L[ℝ] F) (ε : ℝ) (hnoise : ‖N‖ ≤ ε) :
    ‖(learnedLeft X N)ᗮ.starProjection.comp (X+N)‖ ≤ ε := by
  change ‖(leadingLeft (X+N).toLinearMap _ _)ᗮ.starProjection.comp (X+N)‖ ≤ ε
  rw [leadingLeft_tail_norm]
  exact noisy_tail_at_rank X N ε hnoise

/-- The spectral and projection conclusions of P32 now follow from the noisy
operators alone. The estimated subspace and its tail are constructed. -/
theorem finite_table_stability (X N : E →L[ℝ] F)
    (hr : 0 < Module.finrank ℝ X.toLinearMap.range) (ε : ℝ)
    (hnoise : ‖N‖ ≤ ε) (hsmall : ε < leastPositive X.toLinearMap / 2) :
    leastPositive X.toLinearMap - ε ≤
      (X+N).toLinearMap.singularValues (Module.finrank ℝ X.toLinearMap.range-1) ∧
    (X+N).toLinearMap.singularValues (Module.finrank ℝ X.toLinearMap.range) ≤ ε ∧
    ‖X.toLinearMap.range.starProjection-(learnedLeft X N).starProjection‖ ≤
      2*ε/leastPositive X.toLinearMap ∧
    2*ε/leastPositive X.toLinearMap < 1 := by
  have hε : 0 ≤ ε := (norm_nonneg N).trans hnoise
  refine ⟨noisy_singular_lower X N ε hnoise _, noisy_tail_at_rank X N ε hnoise, ?_⟩
  exact gap_from_observed_tail X N (learnedLeft X N) hr ε hε hsmall
    (learnedLeft_observed_tail X N ε hnoise) hnoise
    (learnedLeft_finrank X N hr ε hnoise hsmall).symm

/-- The zero-padded spectrum also implements the maximum possible rank convention. -/
theorem singularValue_zero_at_max_rank (A : E →ₗ[ℝ] F) (k : ℕ)
    (hk : min (Module.finrank ℝ E) (Module.finrank ℝ F) ≤ k) :
    A.singularValues k = 0 := by
  apply A.singularValues_eq_zero_iff_le_finrank_range.mpr
  exact (le_min (LinearMap.finrank_range_le A) (Submodule.finrank_le A.range)).trans hk

end Komorebi.LSM.Spectral
```

### Coordinate and decoder endpoint

```
import KomorebiLSM.DecoderMatrixFormula

set_option autoImplicit false
namespace Komorebi.LSM.FiniteTableP32
open Komorebi.LSM.Spectral Komorebi.LSM.BasisCoordinates
variable {E F : Type*} [NormedAddCommGroup E] [InnerProductSpace ℝ E]
variable [NormedAddCommGroup F] [InnerProductSpace ℝ F]
variable [FiniteDimensional ℝ E] [FiniteDimensional ℝ F]

/-- P32's basis conditioning, coordinate injection, exact inverse and inverse norm,
for any orthonormal bases of the true and data-selected ranges. The spectral and
projection assertions are `Spectral.finite_table_stability`; the literal matrix
formula is `BasisCoordinates.decoder_matrix_formula`. -/
theorem coordinates_and_decoder (X N : E →L[ℝ] F)
    (hr : 0 < Module.finrank ℝ X.toLinearMap.range) (ε : ℝ)
    (hnoise : ‖N‖ ≤ ε) (hsmall : ε < leastPositive X.toLinearMap / 2)
    (U : OrthonormalBasis (Fin (Module.finrank ℝ X.toLinearMap.range)) ℝ X.toLinearMap.range)
    (V : OrthonormalBasis (Fin (Module.finrank ℝ X.toLinearMap.range)) ℝ (learnedLeft X N)) :
    Real.sqrt (1-(2*ε/leastPositive X.toLinearMap)^2) ≤
      (cross U V).toEuclideanLin.singularValues (Module.finrank ℝ X.toLinearMap.range-1) ∧
    Function.Injective ((coordinates V).comp X.toLinearMap.range.subtypeL) ∧
    (∀ x : X.toLinearMap.range, decoder U V (coordinates V (x:F)) = (x:F)) ∧
    (∀ y : EuclideanSpace ℝ (Fin (Module.finrank ℝ X.toLinearMap.range)),
      coordinates V (decoder U V y) = y) ∧
    ‖decoder U V‖ ≤ (Real.sqrt (1-(2*ε/leastPositive X.toLinearMap)^2))⁻¹ := by
  obtain ⟨_,_,hg,hs⟩ := finite_table_stability X N hr ε hnoise hsmall
  have hε : 0 ≤ ε := (norm_nonneg N).trans hnoise
  have hδ : 0 ≤ 2*ε/leastPositive X.toLinearMap :=
    div_nonneg (mul_nonneg (by norm_num) hε) (leastPositive_pos X.toLinearMap hr).le
  exact ⟨cross_min_singular U V hr _ hδ hs hg,
    coordinates_injective V _ hδ hs hg,
    decoder_coordinates U V _ hδ hs hg,
    coordinates_decoder U V _ hδ hs hg,
    decoder_norm_bound U V _ hδ hs hg⟩

end Komorebi.LSM.FiniteTableP32
```

## Reproducing the formal evidence

The source package retains the project manifest, dependency pins, complete coverage map, declaration inventory, source hashes, accepted runs, failed development attempts and rejection controls. These records distinguish proof diagnostics from environmental failures. Historical coverage records retain the status they had when written.

Reproduction requires the separately supplied portable runtime and the exact dependency pins recorded with the formal archive. The runtime itself is not included in this publication package. The source archive's verification instructions describe hash checks, clean compilation with warnings as errors, exact example execution, axiom auditing, module replay and isolated rejec-

tion checks. A successful replay verifies those formal statements under that environment. It does not establish the model-class or observation assumptions of an empirical application.

**Additional formal coverage**

**P11, P23, P24. Ideal risk and grammar.** Conditional expectations decompose ideal risk. Summable grammar evidence controls tails. Prior refinement preserves whole record evidence when the corresponding likelihood mixtures and split prior masses agree.

**P31, P35, P37. Stochastic identification.** Passive laws can miss interventions. Independent reset trials give simultaneous bounds. Observable completion retains experiment coverage, spectral separation, state bounds and filtering margins.

**P46 to P51. Curved domains and matrix witnesses.** Rotation witnesses require a curved cone. Positive matrix embeddings certify probability laws. Polygon error, sequential coverage and conditional learning retain domain, initial state and experiment assumptions.

**P52 to P58. Continuous instruments and moments.** Integrable positive densities define consistent laws. Finite Gram witnesses, supplied polynomial degree and covered invertible cores enable reconstruction. Exact bins and positive averages retain policy and resolution assumptions.

**P59 to P65. Realization and approximation.** Common domains and policies give full record budgets. Template matching and ellipsoid proposals require separate verification. Positive polynomial approximation keeps Lipschitz bias, coefficient error and correction costs distinct.

**P66 to P70. Dependent observation evidence.** Actual conditional means give residual bounds with exact countable budgets and interval widths. Covered actions require positive recorded propensities. Causal comparisons and lower evidence retain separate assumptions.

**P71 to P74. Robustness and repair.** Oriented margins control the full cone. Boundary images can have zero unrestricted noise tolerance. Reset repair retains its prediction cost. Proposal error requires the declared singular gap and residual bounds.

**P75 to P80. Finite curved certificates.** Both rational charts and complete Bernstein trees cover the continuum. Strict search, coefficient boxes, selected geometry and repair brackets retain whole family quantifiers. The repaired example keeps its comparison budget.

**P81 to P82. Regression and operator sets.** A derived Gaussian mixture gives time uniform ridge confidence under predictable design and conditional noise assumptions. Observable moment inversion propagates row confidence into operator sets with the actual inverse factors.

**P83 to P87. Covered observable domains.** The declared scalar family links curved domains, exact outer intervals and law error. One simultaneous event covers selected rectangles. Marginal and rolling window stability require positive reference minorization.

All 87 numbered proposition cores are mapped and checked. Each theorem retains its stated hypotheses; surrounding prose claims are not separately certified by this count.

# Glossary

The definitions below summarize usage in this volume. The cited sections supply the precise spaces and hypotheses.

**Admissible domain.** The declared states or inputs on which an operation has its intended meaning. For a probability instrument, this includes a preserved cone and its normalized base. Sections 11 and 41.

**Archive.** An injective representation of an accepted observation with an explicit left inverse. Its exactness does not recover unmeasured variables. Sections 1–2.

**Atlas.** A declared analysis and synthesis pair for one represented object. Consistent coefficients are those in the analysis image. Sections 2–3.

**Base norm.** The norm obtained by decomposing a vector as a difference of positive-cone elements and minimizing their total mass. Section 46.

**Behavior rank.** The dimension of the span of all history/test columns for the declared process, rather than the rank of one sampled finite table. Section 28.

**Certificate.** A finite witness whose checked identities and inequalities imply the stated property. Internal probability validity and statistical coverage require distinct evidence. Sections 41–45.

**Closure defect.** The failure of a transition to commute with the chosen summary. In the retained linear state, the unresolved-to-retained term is $PLQx$. Sections 5 and 37.

**Common policy.** The same action rule on each common observation history in two compared models. Sections 47 and 65.

**Complete antichain.** A selection in a rooted resolution tree meeting each root-to-leaf path exactly once. Section 4.

**Confidence set.** A random parameter or operator set with a stated probability of containing the truth, conditional on its sampling and model assumptions. Sections 53 and 92–95.

**Consistent coefficient space.** The image of an analysis map. Synthesis followed by reanalysis projects ambient coefficients into this space. Section 2.

**Core table.** A selected square history/test probability matrix used to reconstruct observable operators. Invertibility establishes complete coverage only with the stated rank assumptions. Sections 29 and 34.

**Coverage.** Either a span or intervention condition for identification, or the probability guarantee of a confidence procedure. These meanings are distinguished by context. Sections 16, 32 and 53.

**Curved domain.** A probability domain described through a nonlinear boundary, such as a Lorentz cone, ellipsoid or moment cone. Sections 56, 70 and 94.

**Density.** A function relative to a declared reference measure whose integral gives event probabilities. A value at one point is not itself a point probability. Section 58.

**Exact realization.** A representation and transported operations that reconstruct the accepted object and preserve the declared semantic identities. Sections 2–3.

**Gram certificate.** A positive semidefinite matrix witness for polynomial positivity and the relevant operator identities. Section 59.

**Instrument.** A family of observation operators that preserves an admissible cone and conserves total mass after summing or integrating over observations. Sections 27 and 58.

**Internal validity.** The property that a fitted model defines nonnegative, normalized probability laws and admissible positive-evidence updates. It does not establish accuracy. Section 41.

**Intervention.** An externally specified action in the experiment defining a probability, distinguished from conditioning a passively selected action record. Section 27.

**Lorentz cone.** The mass-oriented cone $\mathcal{L}_d = \{(m, v) : m \geq \|v\|_2\}$. Sections 51 and 80.

**Mass functional.** The linear readout $\lambda$ that assigns total unnormalized probability mass. Section 41.

**Memory kernel.** The operators $BD^kC$ describing departure into unresolved variables, evolution there and return to the retained state. Section 8.

**Minorization.** A uniform lower reference component $P(s, \cdot) \geq \rho\mu(\cdot)$, used to contract differences of marginal laws. Section 97.

**Observable state.** A predictive state accessible from the declared observation history, with the stated warm-up or state map. Sections 15 and 93.

**Packet.** A signal component together with its projector, domain, atlas and resolution. Signed packets need not be probability components. Sections 4 and 31.

**Positive lift.** A representation of predictive operators through nonnegative matrices on a finite generating family. Section 43.

**Predictable test.** A bounded observation test selected using information available before the tested observation. Section 74.

**Predictive closure.** The condition that a summary supports well-defined reduced transitions for every declared operation on its admissible domain. Section 5.

**Predictive state.** A summary sufficient for specified future observations, actions and dynamics. Its adequacy is relative to that specification. Sections 1 and 28.

**Probability cone.** A cone preserved by observation operators with a mass functional strictly positive on nonzero admissible states. Sections 30 and 41.

**Reference measure.** The measure relative to which observation densities are defined and integrated. Changing it requires the matching density transformation. Section 58.

**Refinement.** A more detailed resolution or a representation retaining additional information. A strictly finer state cannot generally be recovered from the coarse state alone. Sections 4, 6 and 23.

**Repair.** A declared modification of a fitted model to satisfy an admissibility or robustness condition. Its approximation cost remains in the error budget. Sections 45, 82 and 91.

**Robust certificate.** A witness quantified over every member of a declared uncertainty family, with its mass constraints and coordinates retained. Sections 44, 80 and 85.

**Total variation.** The probability-law distance used to compare event distributions. For finite normalized vectors, it is half their $\ell^1$ distance. Sections 47–48.

**Transfer block.** The component $Q_j L P_i$ of a linear operation between declared input and output packets. Section 4.

**Written proof.** The mathematical argument in the manuscript. Its existence, formal kernel coverage and finite executable evidence are recorded separately. Appendix A.

# References and source lineage

**[K1]** Komorebi P. *Projector-Resolved Large Algebraic Signal Model*. Historical 38-page capstone preserved in `Komorebi_PR_LASM_2.3.0_Source.zip`, `paper/Projector_Resolved_Large_Algebraic_Signal_Model.tex`. In particular, Sections 2–7 and the categorical-atlas distinction. Historical experimental claims are not used here.

**[K2]** Komorebi P. *Projector-Resolved Seed Calculus VI*, corrected presentation in `Paper_VI_With_Visual_Appendix.pdf`, Section 8, Theorem 8.4 and the adjacent coarsening criterion and one-packet boundary. This volume restates and proves the finite result it uses. It does not claim a fresh Lean audit of that paper.

**[K3]** Komorebi P. *PR-LASM 2.3.0 Release Audit*, 9 October 2026, and the accompanying source. The source inspected includes `exact_atlas.py`, `lasm_packetizer.py`, `lasm_model.py`, `docs/RELEASE_GUIDE.md` and `docs/claim-boundary.md`.

**[R1]** Ronald R. Coifman and Mladen Victor Wickerhauser. "Entropy-based algorithms for best basis selection." *IEEE Transactions on Information Theory* 38(2), 713–718, 1992. DOI: 10.1109/18.119732. Source checked: the authors' Washington University research record. The finite additive tree optimization is classical background.

**[R2]** MIT OpenCourseWare. *6.241J Dynamic Systems and Control*, Chapter 24, "Observability," Spring 2011, especially Section 24.5 on the largest invariant subspace contained in the observation kernel. Source checked: the MIT course notes, PDF page 6. Used to identify the classical control-theoretic antecedent of Proposition 8.

**[R3]** Michael Woodward, Yifeng Tian, Arvind Mohan, Yen Ting Lin, Christoph Hader, Hermann Fasel, Misha Chertkov and Daniel Livescu. *Data-Driven Mori-Zwanzig: Approaching a Reduced Order Model for Hypersonic Boundary Layer Transition*, arXiv:2301.07203, version 1, 2023, introduction and description of resolved, memory and unresolved contributions. Cited for the established projection-memory context, not as evidence for PR-LASM's performance or for any application discussed in that paper. Proposition 10 is proved directly by finite linear recurrence expansion here.

**[R4]** Peter Buchholz. "Exact and ordinary lumpability in finite Markov chains." *Journal of Applied Probability* 31(1), 59–75, 1994. DOI: 10.2307/3215235. Source checked: Cambridge University Press publication record. Used for the classical finite Markov aggregation context.

**[R5]** Michael L. Littman, Richard S. Sutton and Satinder Singh. "Predictive Representations of State." *Advances in Neural Information Processing Systems* 14, NIPS 2001. The paper connects state with predictions of future observations conditional on actions. Cited for conceptual lineage, not as a proof of the finite delay theorem or of this implementation. The original paper and proceedings record are the primary sources. The scanned paper carries all three author names.

**[R6]** Steven L. Brunton, Bingni W. Brunton, Joshua L. Proctor and J. Nathan Kutz. "Koopman Invariant Subspaces and Finite Linear Representations of Nonlinear Dynamical Systems for Control." *PLOS ONE* 11(2), e0150171, 2016. DOI: 10.1371/journal.pone.0150171. Used for the established role and limitations of finite invariant observable families in nonlinear modeling. No general nonlinear feature-discovery guarantee is imported.

**[R7]** Daniel Hsu, Sham M. Kakade and Tong Zhang. "A Spectral Algorithm for Learning Hidden Markov Models." *Journal of Computer and System Sciences* 78(5), 1460–1480, 2012. Primary author manuscript inspected: `hmm-jcss.pdf`, hosted by Daniel Hsu at Columbia University, especially its abstract and rank

assumptions. The result illustrates observable-operator learning under explicit spectral hypotheses. Its statistical guarantees are not claimed for PR-LASM.

**[R8]** Stephen Boyd, Laurent El Ghaoui, Eric Feron and Venkataramanan Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. SIAM Studies in Applied Mathematics 15, 1994. The authors' Stanford publication page identifies the book and provides the primary text. Used to identify the classical common quadratic Lyapunov and matrix-inequality setting. The discrete contraction implication used in Section 19 is proved directly there.

**[R9]** Byron Boots, Sajid M. Siddiqi and Geoffrey J. Gordon. "Closing the Learning-Planning Loop with Predictive State Representations." *Robotics: Science and Systems VI*, 2010. DOI: 10.15607/RSS.2010.VI.036. Primary eight-page proceedings paper inspected at the authors' Carnegie Mellon host, especially Section II on interventional tests and Section III on observable matrices. A journal version appeared in *The International Journal of Robotics Research* 30(7), 954–966, 2011. Used for the established PSR lineage, not as an assertion that this volume implements that paper's complete algorithm or planning guarantees. Source: `https://www.cs.cmu.edu/~ggordon/boots-siddiqi-gordon-closing-loop-psrs.pdf`.

**[R10]** Han Zhao and Pascal Poupart. "A Sober Look at Spectral Learning." *ICML 2014 Workshop on Spectral Learning*, 2014. Primary author-hosted manuscript inspected, including its discussion of negative probabilities in spectral estimates. Used only to identify that practical failure mode. The cone and packet statements in this volume are proved independently. Source: `https://hanzhaoml.github.io/papers/ICML2014/spectral-main.pdf`.

**[R11]** Wassily Hoeffding. "Probability Inequalities for Sums of Bounded Random Variables." *Journal of the American Statistical Association* 58(301), 13–30, 1963. DOI: 10.1080/01621459.1963.10500830. Publisher record checked. Proposition 35 supplies a self-contained Bernoulli moment-generating-function proof for the protocol it uses. Source: `https://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10500830`.

**[R7a]** Daniel Hsu, Sham M. Kakade and Tong Zhang. "A Spectral Algorithm for Learning Hidden Markov Models." *COLT*, 2009. Original conference manuscript inspected at the official Learning Theory proceedings site, particularly Sections 2.3–3.1 on rank, sampling and observable representations. The later journal version is [R7]. No statistical guarantee from either version is imported without its assumptions. Source: `https://www.learningtheory.org/colt2009/papers/011.pdf`.

**[R12]** Herbert Jaeger. "Observable operator models for discrete stochastic time series." *Neural Computation* 12(6), 1371–1398, 2000. Author-hosted manuscript with the August 2017 erratum. Section 7 gives the polyhedral-cone connection to finite positive realizations and an irrational-rotation obstruction. The controlled rational example in Section 51 is an independently worked adaptation of that established obstruction, not a claim of discovering it. The corrected manuscript and its geometric argument were inspected. Source: `https://www.ai.rug.nl/minds/uploads/oom_neco00.pdf`.

**[R13]** Zoltán Horváth, Yunfei Song and Tamás Terlaky. *A Novel Unified Approach to Invariance Conditions for a Linear Dynamical System*. arXiv:1405.5167, version 2, 5 July 2016. Primary author manuscript, particularly the polyhedral-invariance discussion and use of Farkas' lemma. Used for the classical invariant-cone and optimization setting. The generator and robust-vertex statements used here are proved directly. Source: `https://arxiv.org/html/1405.5167v2`.

**[R14]** Marco Fanizza, Josep Lumbreras and Andreas Winter. *Quantum theory in finite dimension cannot explain every general process with finite memory*. arXiv:2209.11225, version 2, 5 May 2023. Primary manuscript, especially its discussion of finite linear realizations, positive realizations and polyhedral cones. Cited for the distinction between these representation classes. No quantum implementation or quantum advantage is claimed for PR-LASM. This volume uses a real three-dimensional cone construction proved in Section 51. Source: `https://arxiv.org/html/2209.11225v2`.

**[R15]** Anna Jenčová. *Base norms and discrimination of generalized quantum channels*. arXiv:1308.4030, version 1, 19 August 2013. Primary manuscript, Section 1.2, printed page 5, on finite-dimensional proper cones, normalized bases and base norms. That ordered-vector-space definition is used in Section 46. The classical

instrument contraction and observation-tree bound needed here are derived in the text, without importing quantum-channel assumptions. Source: https://arxiv.org/pdf/1308.4030.

**[R16]** John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. Author-hosted prepublication text, Chapter 2, Theorem 2.22, printed page 82 (PDF page 90), and the trace-preserving characterization in Theorem 2.26. Used for the established Choi and congruence representations of completely positive maps, not to assign quantum physics to the signal model. The relevant theorem page was visually inspected. Source: https://cs.uwaterloo.ca/~watrous/TQI/TQI.pdf. This third-party text is not redistributed in the bundle.

**[R17]** Pablo A. Parrilo. *Algebraic Techniques and Semidefinite Optimization*, MIT 6.972, Lecture 10, 14 March 2006. MIT OpenCourseWare notes, pages 10-1 through 10-3, on polynomial Gram matrices, coefficient constraints and interval positivity. Used for the standard sum-of-squares certificate method. Proposition 53 proves only the sufficient matrix-valued certificate used here and does not import scalar completeness as a matrix theorem. The formula pages were visually inspected. Source: https://ocw.mit.edu/courses/6-972-algebraic-techniques-and-semidefinite-optimization-spring-2006/resources/lecture_10/.

**[R18]** Herbert Jaeger. *Modeling and learning continuous-valued stochastic processes with OOMs*. GMD technical report, 8 June 2000. Author-hosted report, title and abstract inspected, establishing that continuous-valued observable-operator modeling and learning are prior work. The present polynomial, certificate and binning arguments are proved within this volume. No guarantee from the report is assigned to the reference implementation. Source: https://www.ai.rug.nl/minds/uploads/jaeger.00.tr.contoom.pdf.

**[R19]** Ulrich Abel, Ana Maria Acu, Margareta Heilmann and Ioan Raşa. *A representation for the integral kernel of the composition of multivariate Bernstein–Durrmeyer operators*. arXiv:2512.17297v1, 19 December 2025. The authors' primary manuscript, Introduction, gives the classical univariate Bernstein–Durrmeyer kernel used in Section 72. Its composition theorems are not needed or claimed here. Proposition 64 derives the operator-valued positivity and error statements directly. Source: https://arxiv.org/html/2512.17297v1.

**[R20]** Steven R. Howard, Aaditya Ramdas, Jon McAuliffe and Jasjeet Sekhon. *Time-uniform, nonparametric, nonasymptotic confidence sequences*. *Annals of Statistics* 49(2), 1055–1080, 2021. DOI: 10.1214/20-AOS1991. The primary arXiv record and the conditional exponential-supermartingale discussion in the v4 HTML manuscript were inspected. Proposition 66 proves a simpler union-bound confidence sequence, not the paper's sharper rates. Source: https://arxiv.org/abs/1810.08240 and https://arxiv.org/html/1810.08240v4.

**[R21]** Roland Hildebrand. *An LMI description for the cone of Lorentz-positive maps*. Author manuscript, 17 January 2006. Primary manuscript inspected, especially the introduction and definitions on pages 1–2. It supplies classical context for Lorentz-positive maps and the existence of convex descriptions. The present margin and semicircle certificate arguments are proved directly and do not claim to implement the manuscript's full matrix-inequality construction. Source: https://optimization-online.org/wp-content/uploads/2005/12/1260.pdf.

**[R22]** Tareq Hamadneh, Hassan Al-Zoubi, Hamza Alzaareer and Rafael Wisniewski. *Optimization and Positivity Certificates of Rational Functions using Bernstein Form*. arXiv:1906.11037, 2019. Primary manuscript inspected, including its Bernstein range enclosure discussion on printed pages 3–4. It supplies established positivity-certificate context. Proposition 75 gives its own tensor-product rectangle argument and strict-subdivision proof. Source: https://arxiv.org/abs/1906.11037.

**[R23]** Yasin Abbasi-Yadkori, David Pál and Csaba Szepesvári. *Online Least Squares Estimation with Self-Normalized Processes: An Application to Bandit Problems*. arXiv:1102.2670, version 1, 14 February 2011. The primary manuscript's Section 2, Theorem 3, and Section 5 supply the classical martingale-mixture and adaptive-regression context. Proposition 81 includes its own proof, retains conditional noise and design assumptions, and does not claim a new concentration inequality. Source checked: https://arxiv.org/pdf/1102.2670.

**[R24]** David A. Levin and Yuval Peres, with contributions by Elizabeth L. Wilmer. *Markov Chains and Mixing Times*, second edition, 2017. Chapters 4 and 5 develop total variation and coupling. The present minorization argument is proved directly and does not import a finite-state assumption into its general kernel statement. The worked reference chain is finite. Source checked: the author-hosted second-edition text, `https://pages.uoregon.edu/dlevin/MARKOV/mcmt2e.pdf`.