

Quotient Descent: Hidden-to-Visible Structure Across Reduction Mechanisms

J. R. Dunkley

27 April 2026

Abstract

We study a recurring hidden-to-visible reduction pattern across nine explicit instances drawn from large deviations theory, structural identifiability, chemical kinetics, dissipative systems, graphical models, linear systems, network theory, information theory, and probability. In each case a canonical positive or nonnegative upstairs object produces a well-defined visible descendant after observation, elimination, or conditioning, and the hidden contribution is either factored exactly or isolated by a controlled remainder. The note records these instances at a common internal standard of rigour and separates the proved core from the conjectural transfer perimeter. The main compression achieved in the present pass is that the exact classical core is now seen as one induced visible Hilbert structure with two dual faces. On the dual-observable side, orthogonal compression yields the KL, Fisher, covariance, and observability branches. On the primal-state side, minimal-lift quotienting yields quotient-visible precision, Gaussian precision Schur complements, grounded Dirichlet-to-Neumann laws, and the gradient dissipative kernel frequencywise. The canonical minimal lift identifies the first visible response as pullback of the upstairs perturbation and the quartic defect as hidden leakage through the hidden projector. The Donsker-Varadhan bridge also admits an exact local lift into this geometry, though still only extrinsically. The result is not a universal theorem of reduction. It is a precise statement of a recurring structural pattern, together with its proved instances, its sharpest current boundaries, and a disciplined map of where further transfer may be worth testing.

Contents

1	Introduction	3
2	The quotient descent datum and principle	4
3	Instance 0: The Donsker-Varadhan bridge	5
4	Instance A: Structural identifiability and the descended Fisher form	6
5	Instance B: The CME-CLE quadratic shadow	7
6	Instance C: Hidden elimination in dissipative systems	10
7	Instance D: Latent-variable Gaussian graphical models	14
8	Instance E: Observability Gramians in linear systems	16
9	Instance F: The Dirichlet-to-Neumann map on resistive networks	17
10	Instance G: KL divergence chain rule	19
11	Instance H: Law of total covariance	23
12	Common structure	33
13	What this feeds back into <i>Quotient Observation</i>	35
14	Transfer boundaries and candidate domains	36

1 Introduction

1.1 The hidden-to-visible question

A full model may carry structure in many microscopic or latent directions while observation exposes only a quotient of that structure. The same question then appears in different languages. What survives after hidden variables are eliminated? Which object on the visible sector is canonical? What part of the upstairs structure descends exactly, and what part survives only as a controlled remainder?

The aim of this note is to isolate one recurrent answer. Across nine explicit instances, a canonical positive or nonnegative upstairs object produces a visible descendant under observation, elimination, marginalisation, conditioning, or variational contraction. The mechanisms differ on the surface. The surviving architecture is the same often enough to merit a common name. The source domains are classical: Donsker–Varadhan variational formulae and large deviations, Fisher information geometry, Schur complements, Gaussian graphical models, Lyapunov and observability Gramians, electrical network reduction, and relative-entropy sufficiency all provide established contexts in which the same descent pattern already appears in domain-specific language [4–7, 24, 30, 32, 37, 42].

1.2 Main claim and scope

The common structure recorded here is called *quotient descent*. It is not presented as a universal theorem of hidden-to-visible reduction. It is presented as a proved pattern across a sharply delimited family of explicit cases. In each of those cases the descended visible object is uniquely determined by the full object and the observation mechanism, the natural positivity statement survives, and the hidden contribution is either factored exactly or isolated by a controlled obstruction term.

Two further boundaries are important. First, the note keeps the proved core separate from the transfer perimeter. The final section records candidate domains only as candidate domains. Second, the note does not blur together instances whose formal roles differ. In particular, the KL chain rule, the Fisher second-order shadow, the CME-CLE cubic boundary, and the Donsker-Varadhan bridge are structurally related, but they are not stated here as one closed Taylor ladder unless that identification has actually been proved. The main compression achieved in the present pass is more exact and more structural: the classical core now appears as one induced visible Hilbert geometry with dual faces. Proposition 11.5 gives the orthogonal-compression side, while Theorem 11.7 gives the minimal-lift quotient side. The quotient side captures the visible precision map of *Quotient Observation*, the Gaussian Schur structure of Instance D, the grounded boundary energy behind Instance F, and the gradient subcase of Instance C frequencywise. In addition, Instance 0 now admits an exact local lift into the same broad geometry, though not yet an intrinsic probabilistic one.

1.3 Organisation

Section 2 states the quotient descent datum and the quotient descent principle. Sections 3 to 11 record the nine explicit instances. Section 12 extracts the common structure and draws the line between what is proved and what is not claimed. Section 14 then gives the transfer perimeter as a conjectural map rather than as an extension theorem.

2 The quotient descent datum and principle

2.1 Structural pattern

Across several domains of mathematical physics and statistics, a common pattern appears when one passes from a full model to a partially observed or coarse-grained description. The pattern has five components:

- (i) **Positive structure on the full model.** The full model carries a canonical nonnegative bilinear, quadratic, or convex object (Fisher information, fluctuation Hessian, diffusion matrix, positive-real transfer function, positive precision, KL divergence).
- (ii) **Observation quotient.** An observation map defines a structural quotient, partitioning the model into visible and hidden sectors.
- (iii) **Descent.** The nonnegative object descends to a well-defined visible object on the quotient, via a mechanism specific to the instance (chain rule, variational contraction, Taylor truncation, Schur complement of the resolvent, conditional expectation).
- (iv) **Positivity inheritance.** The descended object inherits nonnegativity from the full form.
- (v) **Controlled hidden contribution.** The hidden contribution is isolated by a factorisation or explicit remainder representation (Gram, Schur, Taylor remainder, conditional divergence).

Two additional features appear across many instances but not all:

- (vi) **Refinement monotonicity.** Enlarging the observation class can only sharpen (never coarsen) the quotient structure.
- (vii) **Sharp boundary.** The first non-descended structure is identified at a specific algebraic or analytic order.

2.2 Formal definition

Definition 2.1 (Quotient descent datum). A quotient descent datum is a tuple $(\mathcal{M}, B, \pi, \bar{\mathcal{M}}, \bar{B}, R)$ where:

- (a) \mathcal{M} is a model space (smooth manifold, vector space, function space, or cone);
- (b) B is a canonical bilinear, quadratic, convex, or operator-theoretic object on \mathcal{M} (or on its tangent bundle), equipped with the natural positivity statement of the instance;
- (c) $\pi : \mathcal{M} \rightarrow \bar{\mathcal{M}}$ is a surjective observation map or quotient mechanism;
- (d) \bar{B} is the descended visible object on $\bar{\mathcal{M}}$, satisfying the natural positivity statement of the instance;
- (e) R is a canonically defined hidden contribution or remainder comparing the full and visible structures in the natural way of the instance.

Definition 2.2 (Quotient descent principle). A quotient descent datum satisfies the quotient descent principle if:

1. **Descent:** \bar{B} is uniquely determined by B and π .
2. **Positivity:** positivity of the full object implies the corresponding positivity statement for the descended visible object.
3. **Controlled hidden structure:** the hidden contribution R admits an explicit Gram-type, Schur-type, Taylor-type, or conditional-expectation representation. When a rank notion is available, its rank or rank bound measures the effective hidden dimension.
4. **Compatibility under refinement:** if $\pi_1 = \rho \circ \pi_2$ with π_2 finer than π_1 , then descent through π_1 agrees with descent through π_2 followed by the induced visible reduction ρ . Depending on the instance, this appears as pullback, Loewner monotonicity, data processing, or Schur-complement composition.

3 Instance 0: The Donsker-Varadhan bridge

This is the foundational instance proved in *Finite Observation* [1]. It sits on the classical Donsker–Varadhan variational and large-deviation background, while the present note uses it only through the finite local Hessian statement quoted below [4, 5].

Theorem 3.1 (Donsker-Varadhan bridge [1, Theorem 3.2]). *Let Q be an irreducible generator on n states with stationary distribution π . In reduced Fisher coordinates with $\widehat{K} = \widehat{G} + \widehat{J}$ (symmetric + skew decomposition of the reduced conjugated generator), the Donsker-Varadhan Hessian decomposes as*

$$H_{DV} = H_0 + \Delta_{DV}$$

where $H_0 = -\frac{1}{2}\widehat{G} \succ 0$ is the detailed-balance reference Hessian and

$$\Delta_{DV} = \frac{1}{4}\widehat{J}H_0^{-1}\widehat{J}^T \succeq 0.$$

Quotient descent reading:

- \mathcal{M} : space of empirical measures near stationarity (parametrised by reduced Fisher coordinates $x \in \mathbb{R}^{n-1}$, with optimisation over reduced logarithmic coordinates $y \in \mathbb{R}^{n-1}$).
- B : the DV Hessian H_{DV} , encoding second-order fluctuation cost.
- π : the variational contraction (supremum over y), equivalently the Legendre transform / Schur complement that eliminates the current sector.
- \widehat{B} : the detailed-balance backbone H_0 (what remains when the skew sector is absent).
- $R = \Delta_{DV}$: the nonequilibrium correction, with Gram factorisation $\Delta_{DV} = \frac{1}{4}A_{DV}A_{DV}^T$, $A_{DV} = \widehat{J}H_0^{-1/2}$.
- Rank: $\text{rank}(\Delta_{DV}) = \text{rank}(\widehat{J})$, always even.
- Boundary: $\Delta_{DV} = 0$ iff detailed balance.

Proposition 3.2 (Canonical local Hilbert lift of the DV bridge). *Let $E = \mathbb{R}^{n-1}$ with its Euclidean inner product and define the linear map*

$$T_{DV} : E \rightarrow E \oplus E, \quad T_{DV}x := \left(H_0^{1/2}x, \frac{1}{2}H_0^{-1/2}\widehat{J}^T x \right). \quad (3.1)$$

Let $P : E \oplus E \rightarrow E \oplus \{0\}$ be the orthogonal projection onto the first summand. Then for every $x \in E$,

$$\|T_{DV}x\|^2 = x^T H_{DV}x, \quad (3.2)$$

$$\|PT_{DV}x\|^2 = x^T H_0x, \quad (3.3)$$

$$\|(I - P)T_{DV}x\|^2 = x^T \Delta_{DV}x. \quad (3.4)$$

Hence Instance 0 admits an exact local Hilbert lift in which the detailed-balance backbone is the projected quadratic form and the nonequilibrium correction is the orthogonal remainder.

Proof. By definition,

$$\|T_{DV}x\|^2 = x^T H_0x + \frac{1}{4}x^T \widehat{J}H_0^{-1}\widehat{J}^T x = x^T (H_0 + \Delta_{DV})x = x^T H_{DV}x,$$

which proves (3.2). Since P keeps only the first component of $T_{DV}x$,

$$\|PT_{DV}x\|^2 = \|H_0^{1/2}x\|^2 = x^T H_0x,$$

proving (3.3). Likewise $(I - P)$ keeps only the second component, so

$$\|(I - P)T_{DV}x\|^2 = \left\| \frac{1}{2}H_0^{-1/2}\widehat{J}^T x \right\|^2 = \frac{1}{4}x^T \widehat{J}H_0^{-1}\widehat{J}^T x = x^T \Delta_{DV}x,$$

which is (3.4). The final sentence is just the interpretation of these three identities.

Remark (What this lift does and does not prove). Proposition 3.2 closes part of the previously open Hilbert-lift question for Instance 0. It shows that the bridge does admit an exact *local* orthogonal-splitting representation after canonical doubling of the tangent space. What it does *not* prove is that the Donsker-Varadhan bridge arises from an intrinsic probabilistic conditional-expectation projection of the kind seen in Instances G and H. The lift here is exact but presently extrinsic: it is built from the pair (H_0, \widehat{J}) after the bridge has already been established.

4 Instance A: Structural identifiability and the descended Fisher form

This bridge is proved in [2]. We restate it here to establish the standard. The external background is Fisher information and local identifiability: the quotient formulation below is the geometric version of separating identifiable tangent directions from vertical degeneracies [6–10].

Let Θ be a smooth parameter manifold, let \mathcal{E} be an admissible experiment class, and for each $E \in \mathcal{E}$ let $p_E(z | \theta)$ be the observational law. Define structural equivalence relative to \mathcal{E} by

$$\theta \sim \theta' \iff p_E(\cdot | \theta) = p_E(\cdot | \theta') \text{ for all } E \in \mathcal{E}.$$

Assume the equivalence classes form a smooth quotient manifold $q : \Theta \rightarrow \bar{\Theta} := \Theta / \sim$ near a regular point.

Theorem 4.1 (Structural descent of the Fisher form). *Under the regular quotient hypothesis:*

- (i) *Each experiment $E \in \mathcal{E}$ factors uniquely through the quotient: there exists $\bar{p}_E(z | \bar{\theta})$ with $p_E(z | \theta) = \bar{p}_E(z | q(\theta))$.*
- (ii) *For every vertical vector $v \in \ker dq_\theta$, the score annihilates: $\partial_v \log p_E(z | \theta) = 0$.*
- (iii) *The Fisher form descends uniquely:*

$$I_E(\theta)(u, w) = \bar{I}_E(q(\theta))(dq_\theta u, dq_\theta w).$$

- (iv) *Structural nonidentifiability is vertical degeneracy of q ; practical nonidentifiability is degeneracy or anisotropy of \bar{I}_E .*

Proof. Since $p_E(\cdot | \theta)$ is constant on structural fibres, it depends only on $q(\theta)$, giving the unique reduced model. Differentiating along $v \in \ker dq_\theta$ gives score annihilation. The Fisher identity $I_E(\theta)(v, \cdot) = 0$ then shows I_E depends only on quotient tangent classes and descends uniquely.

Proposition 4.2 (Refinement monotonicity). *Let $\mathcal{E}_1 \subseteq \mathcal{E}_2$ with regular quotients $q_i : \Theta \rightarrow \bar{\Theta}_i$. Then there exists a unique local surjection $\rho : \bar{\Theta}_2 \rightarrow \bar{\Theta}_1$ with $q_1 = \rho \circ q_2$ and $\bar{I}_E^{(2)} = \rho^* \bar{I}_E^{(1)}$ for every $E \in \mathcal{E}_1$.*

Proof. \mathcal{E}_2 -equivalence implies \mathcal{E}_1 -equivalence, so every \sim_2 -class sits inside a \sim_1 -class, inducing ρ . The Fisher pullback follows by functoriality.

Quotient descent reading:

- $\mathcal{M} = \Theta$, $B = I_E$ (Fisher form), $\pi = q$ (structural quotient).
- $\bar{B} = \bar{I}_E$ (descended Fisher form).
- R : the restriction of I_E to vertical directions, identically zero by score annihilation.
- The hidden remainder here is degenerate (zero on vertical directions), which is the special case where the quotient absorbs all information.
- Refinement: Proposition 4.2.

5 Instance B: The CME-CLE quadratic shadow

5.1 Setting

Consider a reaction network on \mathbb{N}^d with R reactions. For each reaction $r \in \{1, \dots, R\}$, let $v_r \in \mathbb{Z}^d$ be the stoichiometric jump vector and $a_r : \mathbb{N}^d \rightarrow [0, \infty)$ the propensity function. This is the standard chemical master equation setting and its diffusion or system-size shadow; the point here is to isolate the exact visible-closure and cubic-obstruction algebra under projection [11–16]. Define the drift and diffusion:

$$b(x) := \sum_{r=1}^R a_r(x) v_r, \quad D(x) := \sum_{r=1}^R a_r(x) v_r v_r^\top. \quad (5.1)$$

Let $\Pi \in \mathbb{R}^{m \times d}$ be a linear observation map. Write $y = \Pi x$ and define the visible objects:

$$b_\Pi(x) := \Pi b(x), \quad D_\Pi(x) := \Pi D(x) \Pi^\top. \quad (5.2)$$

Define the pullback observable algebra:

$$\mathcal{F}_\Pi := \{f : \mathbb{N}^d \rightarrow \mathbb{R} \mid \exists \varphi : \mathbb{R}^m \rightarrow \mathbb{R} \text{ with } f = \varphi \circ \Pi\}. \quad (5.3)$$

5.2 The shadow theorem

Theorem 5.1 (Quadratic visible shadow). *Let L_{CME} be the chemical master equation generator*

$$(L_{\text{CME}}f)(x) = \sum_{r=1}^R a_r(x) [f(x + v_r) - f(x)],$$

and let L_{CLE} be the chemical Langevin generator

$$(L_{\text{CLE}}g)(x) = b(x) \cdot \nabla g(x) + \frac{1}{2} D(x) : \nabla^2 g(x).$$

Then the following hold.

- (i) **Exact quadratic shadow.** *For every $\varphi : \mathbb{R}^m \rightarrow \mathbb{R}$ of polynomial degree at most two, the pullback observable $f = \varphi \circ \Pi$ satisfies*

$$(L_{\text{CME}}f)(x) = b_\Pi(x) \cdot \nabla \varphi(\Pi x) + \frac{1}{2} D_\Pi(x) : \nabla^2 \varphi(\Pi x). \quad (5.4)$$

In particular, $L_{\text{CME}}f = L_{\text{CLE}}f$ on this test class.

- (ii) **Fibre-constancy closure.** *The visible quadratic theory closes on observation space (i.e. the right-hand side of (5.4) depends only on $y = \Pi x$) if and only if both b_Π and D_Π are*

fibre-constant:

$$x \in \Pi^{-1}(y) \implies b_{\Pi}(x) = \bar{b}(y), \quad D_{\Pi}(x) = \bar{D}(y). \quad (5.5)$$

(iii) **Sharp cubic boundary.** For $\varphi \in C^3(\mathbb{R}^m)$, the discrepancy between the two generators on $f = \varphi \circ \Pi$ is

$$(L_{\text{CME}}f)(x) - (L_{\text{CLE}}f)(x) = \sum_{r=1}^R a_r(x) T_3(\varphi, \Pi x, \Pi v_r), \quad (5.6)$$

where $T_3(\varphi, y, \delta)$ is the exact third-order Taylor remainder:

$$T_3(\varphi, y, \delta) = \int_0^1 \frac{(1-t)^2}{2} D^3\varphi(y+t\delta)[\delta, \delta, \delta] dt. \quad (5.7)$$

For polynomial φ of degree exactly three, $D^3\varphi$ is constant and

$$T_3(\varphi, y, \delta) = \frac{1}{6} D^3\varphi(y)[\delta, \delta, \delta].$$

Proof. For $f = \varphi \circ \Pi$ and any reaction r , Taylor-expand φ around Πx :

$$\begin{aligned} \varphi(\Pi(x + v_r)) - \varphi(\Pi x) &= \varphi(\Pi x + \Pi v_r) - \varphi(\Pi x) \\ &= \nabla\varphi(\Pi x) \cdot (\Pi v_r) + \frac{1}{2}(\Pi v_r)^\top \nabla^2\varphi(\Pi x)(\Pi v_r) + T_3(\varphi, \Pi x, \Pi v_r). \end{aligned} \quad (5.8)$$

For $\deg \varphi \leq 2$, all third and higher derivatives vanish, so $T_3 = 0$ and (5.8) is exact. Summing over r with weights $a_r(x)$:

$$\begin{aligned} (L_{\text{CME}}f)(x) &= \sum_r a_r(x) \left[\nabla\varphi(\Pi x) \cdot (\Pi v_r) + \frac{1}{2}(\Pi v_r)^\top \nabla^2\varphi(\Pi x)(\Pi v_r) \right] \\ &= b_{\Pi}(x) \cdot \nabla\varphi + \frac{1}{2} D_{\Pi} : \nabla^2\varphi, \end{aligned}$$

using the definitions (5.1) to (5.2). This proves (i).

For (ii), the right-hand side of (5.4) is $b_{\Pi}(x) \cdot \nabla\varphi(y) + \frac{1}{2} D_{\Pi}(x) : \nabla^2\varphi(y)$ with $y = \Pi x$. The forward direction: if this depends only on y for *all* quadratic φ , then by choosing $\varphi(y) = e_i^\top y$ and $\varphi(y) = y_i y_j$ respectively, we recover the requirement that $b_{\Pi}(x)$ and $D_{\Pi}(x)$ are constant on each fibre $\Pi^{-1}(y)$. The converse is immediate: fibre-constancy of both quantities makes the right-hand side a function of y alone.

For (iii), the integral-form Taylor theorem gives

$$\varphi(y + \delta) = \varphi(y) + \nabla\varphi(y) \cdot \delta + \frac{1}{2} \delta^\top \nabla^2\varphi(y) \delta + \int_0^1 \frac{(1-t)^2}{2} D^3\varphi(y+t\delta)[\delta, \delta, \delta] dt.$$

Applying this with $y = \Pi x$ and $\delta = \Pi v_r$ yields (5.6). Since L_{CLE} uses only first and second derivatives, the discrepancy is exactly $\sum_r a_r(x) T_3(\varphi, \Pi x, \Pi v_r)$.

5.3 Refinement

Proposition 5.2 (Observation refinement for the quadratic shadow). *Let $\Pi_1 = P\Pi_2$ where $P \in \mathbb{R}^{m_1 \times m_2}$ is a surjective linear map ($m_1 < m_2$). Then for every $\varphi_1 : \mathbb{R}^{m_1} \rightarrow \mathbb{R}$ with $\deg \varphi_1 \leq 2$ and every $x \in \mathbb{N}^d$:*

(i) *The pullback $\varphi_1 \circ P$ has degree at most two on \mathbb{R}^{m_2} , and*

$$L_{\text{CME}}(\varphi_1 \circ \Pi_1) = L_{\text{CME}}((\varphi_1 \circ P) \circ \Pi_2).$$

(ii) *The visible drift and diffusion coarsen pointwise:*

$$b_{\Pi_1}(x) = P b_{\Pi_2}(x), \quad D_{\Pi_1}(x) = P D_{\Pi_2}(x) P^\top.$$

(iii) *The closed quadratic shadow for Π_1 is exactly the statement that the coarsened pair*

$$x \mapsto (P b_{\Pi_2}(x), P D_{\Pi_2}(x) P^\top)$$

is constant on the fibres of $\Pi_1 = P\Pi_2$. In particular, if the Π_2 -shadow closes with descended functions $\bar{b}_2(y_2)$ and $\bar{D}_2(y_2)$, then the Π_1 -shadow closes precisely when

$$P\bar{b}_2(y_2), \quad P\bar{D}_2(y_2)P^\top$$

are constant on the P -fibres in $\Pi_2(\mathbb{N}^d)$. Thus the refinement statement is functorial compatibility of the pointwise shadow, not automatic monotonicity of closure.

Proof. (i) follows because $\varphi_1 \circ P$ is a polynomial in m_2 variables of degree at most two, and $(\varphi_1 \circ P) \circ \Pi_2 = \varphi_1 \circ (P\Pi_2) = \varphi_1 \circ \Pi_1$.

(ii): $b_{\Pi_1} = \Pi_1 b = P\Pi_2 b = P b_{\Pi_2}$. Similarly $D_{\Pi_1} = \Pi_1 D \Pi_1^\top = P\Pi_2 D \Pi_2^\top P^\top = P D_{\Pi_2} P^\top$.

(iii): By Theorem 5.1, the Π_1 -quadratic shadow closes exactly when b_{Π_1} and D_{Π_1} are constant on Π_1 -fibres. Substituting the identities in (ii) gives the first claim. If the Π_2 -shadow closes, then $b_{\Pi_2}(x) = \bar{b}_2(\Pi_2 x)$ and $D_{\Pi_2}(x) = \bar{D}_2(\Pi_2 x)$, and the same criterion becomes constancy of $P\bar{b}_2(y_2)$ and $P\bar{D}_2(y_2)P^\top$ whenever $P y_2$ is fixed. This proves the stated compatibility criterion.

Quotient descent reading:

- $\mathcal{M} = \mathbb{N}^d$ (state space), $B = (b, D)$ (drift and diffusion pair from the CME generator).
- $\pi = \Pi$ (linear observation map).
- $\bar{B} = (\bar{b}, \bar{D})$ (visible drift and diffusion, when fibre-constant).
- R : the cubic Taylor remainder, the first structure lost in the descent. It is controlled by the exact integral remainder (5.7), and hence by third derivatives of φ together with the projected jump sizes $|\Pi v_r|$.
- Refinement: Proposition 5.2; pointwise shadows coarsen functorially, while closure under coarsening requires the explicit fibre-constancy criterion stated there.
- Boundary: cubic order. The quadratic shadow is exact; the first discrepancy is third-order in jump size.

5.4 Toy model

Example 5.3 . *Consider the network $\emptyset \xrightarrow{\alpha} X, X \xrightarrow{\mu} \emptyset, X \xrightarrow{\kappa} Y, Y \xrightarrow{\mu} \emptyset$, observed through $\Pi = \begin{pmatrix} 1 & 1 \end{pmatrix}$ so that $n = x + y$. The stoichiometric jumps project as $\Pi v_1 = 1, \Pi v_2 = -1, \Pi v_3 = 0, \Pi v_4 = -1$. The internal conversion $X \rightarrow Y$ has projected jump zero, so it is*

structurally invisible. Computing:

$$b_{\Pi}(x, y) = \alpha - \mu(x + y) = \alpha - \mu n, \quad D_{\Pi}(x, y) = \alpha + \mu(x + y) = \alpha + \mu n.$$

Both are fibre-constant, so the visible quadratic theory closes exactly. The internal conversion is a vertical direction in the quotient.

Remark (Form of the fibre-constancy condition). The closure criterion in Theorem 5.1(ii) has the form of the classical *lumpability* condition for Markov chains [17, 18]: a function of states yields a Markov process precisely when the relevant generator coefficients are constant on the fibres of the projection. Here the same fibre-constancy is required of the projected drift and diffusion rather than of transition rates, but the algebraic role is the same.

Proposition 5.4 (Why Instance B is not Instance G). *Assume there exist x and r with $a_r(x) > 0$ and $\Pi v_r \neq 0$. Then the cubic discrepancy in (5.6) cannot be represented, even locally on the class of cubic test functions, as a canonically nonnegative remainder analogous to the conditional KL term in Instance G.*

Proof. Instance B is a filtered-observable statement, not an information-theoretic chain rule. The map

$$\varphi \mapsto (L_{\text{CME}} - L_{\text{CLE}})(\varphi \circ \Pi)(x)$$

is linear in φ , and on cubic polynomials it is determined by the third derivative through (5.6). Hence replacing φ by $-\varphi$ changes the sign of the discrepancy. If this discrepancy were equal to a canonically nonnegative remainder on cubic observables, then both it and its negative would have to be nonnegative, forcing it to vanish identically on cubic test functions.

But for any nonzero $\delta := \Pi v_r$, one can choose a cubic polynomial, for example $\varphi(y) = \langle u, y \rangle^3$ with $u \cdot \delta \neq 0$, for which

$$T_3(\varphi, \Pi x, \delta) = \frac{1}{6} D^3 \varphi(\Pi x)[\delta, \delta, \delta] = \langle u, \delta \rangle^3 \neq 0.$$

So the cubic discrepancy is a genuine sign-indefinite obstruction term. It identifies the first algebraic boundary of visible closure, but it is not a KL-type hidden divergence.

Remark (Different categories of marginalisation). Both Instance B and Instance G are described loosely as “marginalisation,” but they marginalise different categories of object. Instance B marginalises a filtered-observable algebra: a linear functional space carrying a generator L , with no underlying probability calculus on the projected level. The discrepancy between the full and visible generators is therefore linear in the test function φ , and sign-indefinite by Proposition 5.4. Instance G marginalises a probability measure under disintegration, and the resulting hidden term is the integral of a Kullback–Leibler quantity, hence canonically nonnegative by Gibbs’ inequality. The sign asymmetry is therefore categorical, not coincidental, and is one structural reason the two instances do not collapse onto a single Taylor ladder.

6 Instance C: Hidden elimination in dissipative systems

6.1 Setting

Consider a linear co-located dissipative input-output system:

$$\dot{z} = Az + Gu, \quad y = G^T z, \tag{6.1}$$

with state $z \in \mathbb{R}^n$, input $u \in \mathbb{R}^m$, output $y \in \mathbb{R}^m$, and the dissipation inequality

$$A + A^\top \preceq 0. \quad (6.2)$$

The transfer function on the open right half-plane is

$$H(s) = G^\top (sI - A)^{-1} G, \quad \Re s > 0. \quad (6.3)$$

This instance is grounded in passive systems, positive-real transfer functions, and Schur-complement elimination; the theorem records exactly which part of that classical machinery survives as quotient descent in the stated finite-dimensional setting [19–27].

Split the state into resolved and hidden blocks:

$$z = \begin{pmatrix} r \\ h \end{pmatrix}, \quad A = \begin{pmatrix} A_{rr} & A_{rh} \\ A_{hr} & A_{hh} \end{pmatrix}, \quad G = \begin{pmatrix} G_r \\ G_h \end{pmatrix}. \quad (6.4)$$

6.2 The hidden elimination theorem

Theorem 6.1 (Dissipative hidden elimination). *Assume zero initial state, $\Re s > 0$, and that $sI - A_{hh}$ is invertible. Then:*

(i) **Positive-realness.** *The full transfer function is positive-real:*

$$H(s) + H(s)^* \succeq 0. \quad (6.5)$$

(ii) **Dynamic Schur complement representation.** *Exact elimination of the hidden block gives*

$$H(s) = D_{\text{eff}}(s) + C_{\text{eff}}(s) \Sigma_r(s)^{-1} B_{\text{eff}}(s), \quad (6.6)$$

where the dynamic Schur complement and effective input/output maps are:

$$\Sigma_r(s) = sI_r - A_{rr} - A_{rh}(sI_h - A_{hh})^{-1}A_{hr}, \quad (6.7)$$

$$B_{\text{eff}}(s) = G_r + A_{rh}(sI_h - A_{hh})^{-1}G_h, \quad (6.8)$$

$$C_{\text{eff}}(s) = G_r^\top + G_h^\top (sI_h - A_{hh})^{-1}A_{hr}, \quad (6.9)$$

$$D_{\text{eff}}(s) = G_h^\top (sI_h - A_{hh})^{-1}G_h. \quad (6.10)$$

(iii) **Positivity inheritance.** *The reduced transfer $H(s)$ computed via (6.6) is positive-real. More precisely, for each s with $\Re s > 0$:*

$$H(s) + H(s)^* = 2\Re(s) X(s)^* X(s) - X(s)^* (A + A^\top) X(s) \succeq 0, \quad (6.11)$$

where $X(s) = (sI - A)^{-1}G$.

(iv) **Gradient subcase.** *If $A = -B$ with $B = B^\top \succeq 0$ and the hidden states are not directly forced ($G_h = 0$), then:*

(a) *The transfer is real and positive for real $s > 0$.*

(b) *The dynamic Schur complement simplifies to*

$$\Sigma_r(s) = sI_r + B_{rr} - B_{rh}(sI_h + B_{hh})^{-1}B_{hr}, \quad (6.12)$$

and $H(s) = G_r^\top \Sigma_r(s)^{-1} G_r$. *Consequently, for real $s > 0$,*

$$\Sigma_r(s) \preceq sI_r + B_{rr}, \quad H(s) - H_{\text{bare}}(s) \succeq 0, \quad (6.13)$$

where $H_{\text{bare}}(s) := G_r^\top (sI_r + B_{rr})^{-1} G_r$.

(c) The transfer H is a matrix Stieltjes function: it maps the upper half-plane $\{\Im s > 0\}$ to the lower half-plane $\{M : \Im M \leq 0\}$.

Proof. (i) For any $\xi \in \mathbb{C}^m$, set $X = (sI - A)^{-1} G \xi$. Then $G \xi = (sI - A)X$, so

$$2\Re(\xi^* H(s) \xi) = 2\Re(X^* G \xi) = 2\Re(s) \|X\|^2 - X^* (A + A^\top) X \geq 0,$$

since $\Re s > 0$ and $A + A^\top \leq 0$. This gives (6.5) and (6.11).

(ii) The resolvent equation $(sI - A) \begin{pmatrix} r \\ h \end{pmatrix} = \begin{pmatrix} G_r \\ G_h \end{pmatrix} u$ gives, from the hidden block:

$$h = (sI_h - A_{hh})^{-1} (A_{hr} r + G_h u).$$

Substituting into the resolved block:

$$\Sigma_r(s) r = B_{\text{eff}}(s) u.$$

The output $y = G_r^\top r + G_h^\top h$ then gives (6.6) after substitution.

(iii) Positivity of the reduced transfer follows from (i), since the Schur complement representation is an algebraic identity: it computes the same $H(s)$.

(iv) When $A = -B$, $B = B^\top \succeq 0$, and $G_h = 0$, the full resolvent is $(sI + B)^{-1}$ with block inverse

$$((sI + B)^{-1})_{rr} = \Sigma_r(s)^{-1},$$

where (6.12) is the standard Schur-complement formula. Since $(sI_h + B_{hh})^{-1} \succeq 0$ for real $s > 0$, the correction term

$$B_{rh}(sI_h + B_{hh})^{-1} B_{hr} \succeq 0,$$

so (6.13) follows by Loewner monotonicity and order reversal under inversion on positive definite matrices. Hence $H(s)$ is real and positive for real $s > 0$.

For the Stieltjes property, let z lie in the upper half-plane. Then

$$\Im((zI + B)^{-1}) = \frac{(zI + B)^{-1} - (\bar{z}I + B)^{-1}}{2i} = -\Im(z) (zI + B)^{-1} (\bar{z}I + B)^{-1} \leq 0.$$

Congruence by the fixed matrix $G = (G_r, 0)$ preserves Loewner order, so

$$\Im H(z) = \Im(G^\top (zI + B)^{-1} G) \leq 0.$$

Thus H is a matrix Stieltjes function.

6.3 Refinement

Proposition 6.2 (Resolution refinement for dissipative elimination). Consider two splittings of the state space: $z = (r_1, h_1)$ with $\dim r_1 = k_1$, and $z = (r_2, h_2)$ with $\dim r_2 = k_2 > k_1$, where the resolved block of the coarser splitting is a sub-block of the finer. Let $H^{(i)}(s)$ denote the transfer function obtained after eliminating the corresponding hidden block. Then:

- (i) Both reduced transfers are positive-real.
- (ii) The coarser reduction is recovered from the finer one by further elimination: $H^{(1)}(s)$ is obtained from the effective data of splitting 2 by a second Schur complement over the states that are resolved in splitting 2 but hidden in splitting 1.

(iii) In the gradient subcase, the Stieltjes property is preserved at each elimination stage.

Proof. (i) follows from Theorem 6.1(iii) applied to each splitting. (ii) is the composition identity for Schur complements of block resolvents. (iii) follows because each stage is a compression of $(sI + B)^{-1}$ by a fixed matrix after eliminating a positive block, and the argument in Theorem 6.1(iv) applies verbatim at each stage.

Quotient descent reading:

- \mathcal{M} : the full state space \mathbb{R}^n , carrying the dissipative system.
- $B = H(s)$: the positive-real transfer function (frequency-dependent bilinear form on input/output space).
- π : hidden elimination (projection onto the resolved block, followed by dynamic Schur complement).
- \bar{B} : the exact reduced transfer $H(s)$ represented by the effective resolved operator $\Sigma_r(s)$ together with $B_{\text{eff}}, C_{\text{eff}}, D_{\text{eff}}$.
- R : the hidden memory kernel

$$K_{\text{hid}}(s) := A_{rh}(sI_h - A_{hh})^{-1}A_{hr},$$

which is the explicit hidden contribution entering the reduced dynamics. In the gradient subcase with $G_h = 0$, comparison with the bare resolved model yields the additional PSD correction

$$H(s) - H_{\text{bare}}(s) = G_r^\top [\Sigma_r(s)^{-1} - (sI_r + B_{rr})^{-1}] G_r \geq 0$$

for real $s > 0$, which is the transfer-function analogue of $\Delta_{DV} \geq 0$ in the Finite bridge.

- Refinement: Proposition 6.2.
- Boundary: the hidden memory kernel $A_{rh}(sI_h - A_{hh})^{-1}A_{hr}$ is the first dynamic structure that enters.

6.4 Toy model

Example 6.3. Let $n = 2$, $A = \begin{pmatrix} -a & c \\ c & -d \end{pmatrix}$ with $a, d > 0$ and $|c| \leq \sqrt{ad}$, and $G = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ (hidden state not directly forced). Then:

$$H(s) = \frac{1}{s + a - c^2/(s + d)} = \frac{s + d}{(s + a)(s + d) - c^2},$$

$$H_{\text{bare}}(s) = \frac{1}{s + a}.$$

The hidden contribution is

$$R(s) = H(s) - H_{\text{bare}}(s) = \frac{c^2}{(s + a)[(s + a)(s + d) - c^2]} > 0 \quad \text{for real } s > 0.$$

The Schur correction $c^2/(s + d)$ to the resolved dynamics is the exact analogue of $\frac{1}{4}\widehat{J}H_0^{-1}\widehat{J}^\top$: the off-diagonal coupling c plays the role of \widehat{J} , the hidden stiffness $(s + d)$ plays the role of H_0 , and the quadratic-in-coupling structure $c^2/(s + d)$ parallels the Gram form $\widehat{J}H_0^{-1}\widehat{J}^\top$.

7 Instance D: Latent-variable Gaussian graphical models

7.1 Setting

Let $X = (X_V, X_H)$ be a jointly Gaussian random vector on $\mathbb{R}^{|V|+|H|}$ with precision matrix $\Omega \in \text{SPD}(n)$, $n = |V| + |H|$. The index sets V (visible) and H (hidden/latent) partition the variables. Block-decompose:

$$\Omega = \begin{pmatrix} \Omega_{VV} & \Omega_{VH} \\ \Omega_{HV} & \Omega_{HH} \end{pmatrix}. \quad (7.1)$$

The marginal distribution of X_V is Gaussian with covariance Σ_{VV} (the leading block of $\Sigma = \Omega^{-1}$). This is the Schur-complement form of Gaussian marginalisation, central in multivariate analysis, Gaussian graphical models, and latent-variable graphical model selection [24, 28, 30, 31].

7.2 The marginalisation theorem

Theorem 7.1 (Precision descent under marginalisation). *The marginal precision matrix of X_V is the Schur complement:*

$$\bar{\Omega} := \Sigma_{VV}^{-1} = \Omega_{VV} - \Omega_{VH} \Omega_{HH}^{-1} \Omega_{HV}. \quad (7.2)$$

Moreover:

- (i) **Uniqueness.** $\bar{\Omega}$ is uniquely determined by Ω and the partition $V \cup H$.
- (ii) **Positivity.** $\bar{\Omega} \succ 0$.
- (iii) **Gram factorisation of hidden gap.** Define $R := \Omega_{VH} \Omega_{HH}^{-1} \Omega_{HV}$, and the canonical factor

$$B_{\text{can}} := \Omega_{VH} \Omega_{HH}^{-1/2} \in \mathbb{R}^{|V| \times |H|}.$$

Then

$$R = B_{\text{can}} B_{\text{can}}^\top \succeq 0, \quad \text{rank}(R) = \text{rank}(\Omega_{VH}) \leq |H|. \quad (7.3)$$

If $\tilde{B}, \tilde{B}' \in \mathbb{R}^{|V| \times r}$ are minimal factors with $r = \text{rank}(R)$ and

$$R = \tilde{B} \tilde{B}^\top = \tilde{B}' \tilde{B}'^\top,$$

then $\tilde{B}' = \tilde{B} Q$ for some $Q \in O(r)$.

- (iv) **Graph-sparsity plus low-rank decomposition.** If Ω has the sparsity pattern of a graph $G = (V \cup H, E)$, then $\bar{\Omega}$ decomposes as

$$\bar{\Omega} = S - L \quad (7.4)$$

where $S = \Omega_{VV}$ is sparse (inheriting the V - V edge pattern of the graph G , since $(\Omega_{VV})_{ij} = 0$ whenever there is no edge between $i, j \in V$ in G) and $L = R$ is positive semidefinite with $\text{rank}(L) \leq |H|$. The sparse-plus-low-rank structure is an exact algebraic consequence of the quotient, not an approximation.

Proof. The identity (7.2) is the classical Schur complement formula for marginals of multivariate Gaussians, obtained by integrating out X_H from the joint density.

- (i): The Schur complement is a deterministic algebraic function of Ω and the index partition.
- (ii): The Schur complement of a positive definite matrix is positive definite; this is the positive-definite case of Haynsworth's inertia theorem.

(iii): Since $\Omega_{HH} \succ 0$, the square root $\Omega_{HH}^{1/2}$ exists and is invertible. Setting $B_{\text{can}} = \Omega_{VH}\Omega_{HH}^{-1/2}$ gives $R = B_{\text{can}}B_{\text{can}}^\top$, so $R \succeq 0$. Since $\Omega_{HH}^{-1/2}$ is invertible, $\text{rank}(B_{\text{can}}) = \text{rank}(\Omega_{VH})$, and $\text{rank}(B_{\text{can}}B_{\text{can}}^\top) = \text{rank}(B_{\text{can}})$. For minimal factors \tilde{B}, \tilde{B}' with full column rank r , define $Q := \tilde{B}^+\tilde{B}'$. Then $\tilde{B}Q = \tilde{B}'$ and

$$QQ^\top = \tilde{B}^+R(\tilde{B}^+)^\top = \tilde{B}^+\tilde{B}\tilde{B}^\top(\tilde{B}^+)^\top = I_r,$$

so $Q \in O(r)$.

(iv): $S = \Omega_{VV}$ has the sparsity of the V -induced subgraph, and $L = R$ is PSD of rank at most $|H|$.

Proposition 7.2 (Refinement under latent revelation). *Let $H = H_1 \cup H_2$ be a partition of the hidden set. Define*

$$\tilde{\Omega}_{\text{all hidden}} := \Omega/\Omega_{HH} \quad (\text{marginalise all of } H),$$

$$\tilde{\Omega}_{\text{some hidden}} := \Omega/\Omega_{H_2H_2} \quad (\text{marginalise only } H_2, \text{ reveal } H_1).$$

Then $\tilde{\Omega}_{\text{all hidden}}$ is obtained from $\tilde{\Omega}_{\text{some hidden}}$ by a further Schur complement (marginalising the revealed H_1 from the intermediate). Moreover, on the original visible set V :

$$(\tilde{\Omega}_{\text{some hidden}})_{VV} \succeq \tilde{\Omega}_{\text{all hidden}}. \quad (7.5)$$

That is, revealing latent variables can only increase the effective precision on the original visible set.

Proof. Schur complements compose, so marginalising H_2 and then H_1 gives the same matrix as marginalising $H_1 \cup H_2$ in one step. For the Loewner inequality, write the intermediate precision after marginalising H_2 as

$$\tilde{\Omega} = \begin{pmatrix} \tilde{\Omega}_{VV} & \tilde{\Omega}_{VH_1} \\ \tilde{\Omega}_{H_1V} & \tilde{\Omega}_{H_1H_1} \end{pmatrix} \succ 0.$$

Then

$$\tilde{\Omega}_{\text{all hidden}} = \tilde{\Omega}_{VV} - \tilde{\Omega}_{VH_1}\tilde{\Omega}_{H_1H_1}^{-1}\tilde{\Omega}_{H_1V},$$

while $(\tilde{\Omega}_{\text{some hidden}})_{VV} = \tilde{\Omega}_{VV}$. The correction term is positive semidefinite, so (7.5) follows.

Quotient descent reading:

- $\mathcal{M} = \text{SPD}(n)$ (joint precision matrices), $B = \Omega$.
- π : marginalisation over H (Schur complement).
- $\tilde{B} = \tilde{\Omega}$ (marginal precision).
- $R = \Omega_{VH}\Omega_{HH}^{-1}\Omega_{HV}$: the latent mediation, with Gram form BB^\top , $\text{rank} \leq |H|$.
- Refinement: Proposition 7.2.
- Boundary: the sparsity pattern of $\tilde{\Omega}$ can be denser than Ω_{VV} ; this is the fill-in phenomenon for Schur complements of sparse matrices.

Example 7.3 . *Let $V = \{1, 2\}$, $H = \{3\}$, and $\Omega = \begin{pmatrix} 3 & 0 & 1 \\ 0 & 3 & 1 \\ 1 & 1 & 2 \end{pmatrix}$. Then $\Omega_{VV} = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$ is diagonal (no direct edge between 1 and 2), but $\tilde{\Omega} = \Omega_{VV} - \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \end{pmatrix} = \begin{pmatrix} 2.5 & -0.5 \\ -0.5 & 2.5 \end{pmatrix}$ has a nonzero (1, 2) entry. The single latent variable creates an effective coupling between the two visible nodes. The hidden rank is $\text{rank}(R) = 1$.*

8 Instance E: Observability Gramians in linear systems

8.1 Setting

Consider a stable linear time-invariant system:

$$\dot{x} = Ax, \quad y = Cx, \quad (8.1)$$

with state $x \in \mathbb{R}^n$ and output $y \in \mathbb{R}^p$, where A is Hurwitz (all eigenvalues have strictly negative real part). The *observability Gramian* is

$$W_C := \int_0^\infty e^{A^\top t} C^\top C e^{At} dt, \quad (8.2)$$

the unique positive semidefinite solution of the Lyapunov equation

$$A^\top W_C + W_C A + C^\top C = 0. \quad (8.3)$$

W_C encodes the total information about the initial state $x(0)$ available from the infinite output trajectory $\{y(t)\}_{t \geq 0}$. This is the standard observability-Gramian setting; the quotient split below separates observed output-energy directions from directions constrained only through hidden coupling [32–35].

8.2 The output-refinement theorem

Theorem 8.1 (Observability Gramian descent). *Let A be Hurwitz and let $C_1 \in \mathbb{R}^{p_1 \times n}$, $C_2 \in \mathbb{R}^{p_2 \times n}$ be two output matrices with $p_1 \leq p_2$. Assume C_1 consists of a subset of the rows of C_2 . Then:*

- (i) **Uniqueness.** *Each W_{C_i} is the unique PSD solution of its Lyapunov equation.*
- (ii) **Positivity.** *$W_{C_i} \succeq 0$, with $\ker W_{C_i}$ equal to the unobservable subspace of (A, C_i) .*
- (iii) **Additivity over outputs.** *If $C_2 = \begin{pmatrix} C_1 \\ C_{\text{new}} \end{pmatrix}$, then*

$$W_{C_2} = W_{C_1} + W_{C_{\text{new}}}. \quad (8.4)$$

The hidden gap $R := W_{C_2} - W_{C_1} = W_{C_{\text{new}}} \succeq 0$ is itself the observability Gramian of the missing outputs.

- (iv) **Gram factorisation.** *The Gramian $W_{C_{\text{new}}}$ has the integral Gram form*

$$W_{C_{\text{new}}} = \int_0^\infty F(t) F(t)^\top dt, \quad F(t) := e^{A^\top t} C_{\text{new}}^\top. \quad (8.5)$$

The new output block has channel dimension $p_2 - p_1$, but time propagation through A can spread those channels across many state directions, so one should not expect the rank of $W_{C_{\text{new}}}$ to be bounded by $p_2 - p_1$. The only general a priori bound is $\text{rank}(W_{C_{\text{new}}}) \leq n$.

- (v) **Loewner refinement.** *$W_{C_2} \succeq W_{C_1}$ in the Loewner order. More outputs yield more information.*

Proof. (i): Stability of A guarantees that the Lyapunov equation $A^\top W + WA + C^\top C = 0$ has a unique solution.

(ii): The integral representation (8.2) is manifestly PSD. Its kernel is $\{x : C e^{At} x = 0 \text{ for all } t \geq 0\}$, which is exactly the unobservable subspace.

(iii): By linearity of the Lyapunov equation: if $C_2^\top C_2 = C_1^\top C_1 + C_{\text{new}}^\top C_{\text{new}}$ (which holds when $C_2 = (C_1^\top, C_{\text{new}}^\top)^\top$), then $W_{C_2} = W_{C_1} + W_{C_{\text{new}}}$ by uniqueness of the Lyapunov solution.

(iv): Immediate from (8.2) applied to C_{new} .

(v): Direct from (iii) and $W_{C_{\text{new}}} \succeq 0$.

Quotient descent reading:

- $\mathcal{M} = \mathbb{R}^n$ (state space), $B = W_C$ (observability Gramian as the information form).
- π : restriction of the output (use fewer sensors / measure fewer states).
- $\bar{B} = W_{C_1}$ (Gramian from partial output).
- $R = W_{C_{\text{new}}}$: the missing-output Gramian, with integral Gram form. Its channel dimension is $p_2 - p_1$, while its state-space rank can be as large as n .
- Refinement: adding outputs can only increase the Gramian (Loewner monotonicity).
- Boundary: the unobservable subspace of (A, C_1) is the kernel of W_{C_1} , exactly characterising what cannot be reconstructed from partial output.

Example 8.2. Consider $A = \begin{pmatrix} -1 & 2 \\ 0 & -3 \end{pmatrix}$ (stable, upper-triangular) with full output $C_2 = I_2$ and partial output $C_1 = (1 \ 0)$ (observe only x_1). Despite observing only x_1 , the cross-coupling $A_{12} = 2$ propagates information about x_2 into the x_1 channel, so (A, C_1) is observable and $W_{C_1} \succ 0$. However, $W_{C_2} \succ W_{C_1}$: direct observation of x_2 provides additional information beyond what propagates through A .

9 Instance F: The Dirichlet-to-Neumann map on resistive networks

9.1 Setting

Let $G = (V, E, w)$ be a connected weighted graph with positive conductances $w_{ij} > 0$. This is the classical Kron or Dirichlet-to-Neumann reduction of a resistor network, with the hidden interior eliminated but its energetic contribution kept visible in the quotient reading [36–41]. The graph Laplacian is

$$L_{ij} = \begin{cases} -w_{ij} & i \neq j, \\ \sum_{k \neq i} w_{ik} & i = j. \end{cases} \quad (9.1)$$

$L \succeq 0$ with $\ker L = \text{span}\{\mathbf{1}\}$. Partition $V = B \cup I$ (boundary = visible, interior = hidden). Block-decompose:

$$L = \begin{pmatrix} L_{BB} & L_{BI} \\ L_{IB} & L_{II} \end{pmatrix}. \quad (9.2)$$

Since G is connected and $I \subsetneq V$, the submatrix L_{II} is positive definite. Indeed, for $u \in \mathbb{R}^I$ extend u by zero on B ; then $u^\top L_{II} u$ is the graph energy of this extension, and it vanishes only when the extension is constant on the connected graph, hence only when $u = 0$.

9.2 The effective conductance theorem

Theorem 9.1 (Dirichlet-to-Neumann descent). *The effective Laplacian (Dirichlet-to-Neumann map) on the boundary is*

$$\bar{L} := L_{BB} - L_{BI} L_{II}^{-1} L_{IB}. \quad (9.3)$$

This satisfies:

- (i) **Laplacian structure preserved.** \bar{L} is a valid graph Laplacian on B : it has non-positive off-diagonals, rows summing to zero, and $\bar{L} \succeq 0$ with $\ker \bar{L} = \text{span}\{\mathbf{1}_B\}$.
- (ii) **Uniqueness.** \bar{L} is uniquely determined by L and the partition $B \cup I$.
- (iii) **Electrical interpretation.** For any boundary voltage assignment $v_B \in \mathbb{R}^{|B|}$, the interior voltages that minimise the total dissipated power $v^\top L v$ are $v_I^* = -L_{II}^{-1} L_{IB} v_B$, and the power dissipated at this minimiser is $v_B^\top \bar{L} v_B$. Thus \bar{L} is the effective conductance matrix of the network as seen from the boundary.
- (iv) **Mediated conductance and Gram factorisation.** The Schur correction $R := L_{BI} L_{II}^{-1} L_{IB} \succeq 0$ represents the mediated conductance through interior nodes. Setting $C := -L_{BI} \succeq 0$ (entrywise, since L_{BI} has non-positive entries) and $B_{\text{gram}} := C L_{II}^{-1/2}$:

$$R = B_{\text{gram}} B_{\text{gram}}^\top, \quad \text{rank}(R) \leq |I|. \quad (9.4)$$

Proof. (i): Off-diagonals of \bar{L} : write $\bar{L}_{ij} = (L_{BB})_{ij} - (L_{BI} L_{II}^{-1} L_{IB})_{ij}$ for $i \neq j \in B$. Now $(L_{BB})_{ij} = -w_{ij} \leq 0$, and $L_{II}^{-1} \geq 0$ entrywise (since L_{II} is an M -matrix of a connected graph), so $(L_{BI} L_{II}^{-1} L_{IB})_{ij} = \sum_{k,l \in I} (L_{BI})_{ik} (L_{II}^{-1})_{kl} (L_{IB})_{lj}$. Since $(L_{BI})_{ik} \leq 0$, $(L_{II}^{-1})_{kl} \geq 0$, and $(L_{IB})_{lj} \leq 0$, each term in the sum is the product $(\leq 0)(\geq 0)(\leq 0) \geq 0$. Thus the entire sum is ≥ 0 , and $\bar{L}_{ij} = (L_{BB})_{ij} - (\text{nonneg}) \leq 0$.

Row sums: $L \mathbf{1} = 0$ gives $L_{BB} \mathbf{1}_B + L_{BI} \mathbf{1}_I = 0$ and $L_{IB} \mathbf{1}_B + L_{II} \mathbf{1}_I = 0$. From the second: $\mathbf{1}_I = -L_{II}^{-1} L_{IB} \mathbf{1}_B$. Substituting into the first: $\bar{L} \mathbf{1}_B = L_{BB} \mathbf{1}_B - L_{BI} L_{II}^{-1} L_{IB} \mathbf{1}_B = L_{BB} \mathbf{1}_B + L_{BI} \mathbf{1}_I = 0$.

Positive semidefiniteness with one-dimensional kernel follows from \bar{L} being a Schur complement of the PSD matrix L with kernel $\text{span}\{\mathbf{1}\}$.

(ii): Algebraic function of L and the partition.

(iii): The minimisation $\min_{v_I} v^\top L v$ over interior voltages at fixed boundary voltage v_B gives the quadratic form $v_B^\top \bar{L} v_B$ by completing the square in v_I , identical to the Schur complement construction.

(iv): $L_{II} \succ 0$ admits a square root. $C = -L_{BI}$ has non-negative entries. $R = C L_{II}^{-1} C^\top = B_{\text{gram}} B_{\text{gram}}^\top \succeq 0$ with $\text{rank}(R) = \text{rank}(C) \leq |I|$.

Proposition 9.2 (Composition under staged elimination). *Let $I = I_1 \cup I_2$ and define \bar{L}_I by condensing all of I at once, and \bar{L}_{I_2} by condensing only I_2 so that the intermediate boundary is $B \cup I_1$. Then condensing I_1 from the intermediate network recovers the full condensation:*

$$\bar{L}_I = (\bar{L}_{I_2}) / (\bar{L}_{I_2})_{I_1 I_1}. \quad (9.5)$$

In general there is no entrywise monotonicity of the off-diagonal effective conductances under revealing interior nodes.

Proof. Write the full Laplacian in three blocks corresponding to B , I_1 , and I_2 . Schur complements over disjoint blocks compose, so eliminating I_2 and then I_1 gives the same result as eliminating $I_1 \cup I_2$ in one step. This proves (9.5). The final sentence is a warning rather than an additional claim: for the path $1 - 3 - 2$, condensing node 3 creates an effective

edge of weight $1/2$ between 1 and 2, whereas revealing node 3 removes any direct 1, 2 edge in the enlarged boundary network.

Remark (Connection to the Donsker-Varadhan bridge). For a reversible (detailed-balance) Markov chain with generator Q and stationary distribution π , the conductance matrix $w_{ij} = \pi_i q_{ij}$ defines a weighted graph whose Laplacian is $L_{ij} = -\pi_i q_{ij}$ for $i \neq j$. In Fisher coordinates, the reduced symmetric operator \widehat{G} is the restriction of this weighted Laplacian to the tangent space $\{\sqrt{\pi}\}^\perp$. So the detailed-balance backbone of Instance 0 lives in the same Laplacian algebraic family as Instance F. What is not claimed here is an exact identification of the full DV bridge with a Dirichlet-to-Neumann map: that would require a separate boundary/interior partition and elimination theorem. The robust statement is the shared Laplacian backbone, with the nonequilibrium correction Δ_{DV} appearing only after the skew sector is introduced.

Quotient descent reading:

- \mathcal{M} : weighted graph (V, E, w) , carrying $B = L$ (graph Laplacian).
- π : interior condensation (Schur complement over hidden nodes).
- $\bar{B} = \bar{L}$ (effective boundary Laplacian, itself a valid Laplacian).
- $R = L_{BI} L_{II}^{-1} L_{IB}$: mediated conductance, Gram form, rank $\leq |I|$.
- Refinement: Proposition 9.2 gives staged-elimination compatibility rather than entrywise monotonicity.
- Boundary: \bar{L} can have fill-in (nonzero entries where L_{BB} had zeros), corresponding to effective long-range boundary couplings mediated by interior paths.

Example 9.3. Consider the path graph $1 - 3 - 2$ with unit weights, $B = \{1, 2\}$, $I = \{3\}$. Then $L_{BB} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $L_{BI} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$, $L_{II} = (2)$. The effective Laplacian is $\bar{L} = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix}$: the interior node creates an effective conductance of $1/2$ between the boundary nodes that have no direct edge. The effective resistance between nodes 1 and 2 is 2 (sum of the two unit resistances in series), consistent with $1/\bar{w}_{12} = 1/(1/2) = 2$.

10 Instance G: KL divergence chain rule

10.1 Setting

Let $(\mathcal{X}_V, \mathcal{A}_V)$ and $(\mathcal{X}_H, \mathcal{A}_H)$ be standard Borel spaces, and let $\Omega = \mathcal{X}_V \times \mathcal{X}_H$ with product σ -algebra. Let P, Q be probability measures on Ω with $P \ll Q$. Write P_V, Q_V for the visible marginals and $P_{H|V}, Q_{H|V}$ for regular conditional distributions of the hidden coordinate given the visible coordinate. Then $P_V \ll Q_V$, and the conditional divergence below is understood with the usual extended-value convention on a fixed regular conditional version.

The *Kullback-Leibler divergence* (relative entropy) is

$$D_{\text{KL}}(P||Q) := \int \log \frac{dP}{dQ} dP. \quad (10.1)$$

The chain rule is the measure-theoretic version of information sufficiency and conditional decomposition; the standard-Borel assumption supplies regular conditional distributions [42–45].

10.2 The chain rule theorem

Theorem 10.1 (KL descent under marginalisation). *Under the hypotheses above, the KL divergence decomposes as*

$$D_{\text{KL}}(P\|Q) = D_{\text{KL}}(P_V\|Q_V) + \int D_{\text{KL}}(P_{H|v}\|Q_{H|v}) dP_V(v). \quad (10.2)$$

Moreover:

- (i) **Uniqueness.** $D_{\text{KL}}(P_V\|Q_V)$ is uniquely determined by P , Q , and the marginalisation $\pi : \mathcal{F} \rightarrow \mathcal{F}_V$.
- (ii) **Positivity.** $D_{\text{KL}}(P_V\|Q_V) \geq 0$, with equality if and only if $P_V = Q_V$.
- (iii) **Hidden remainder.** The conditional divergence

$$R := \int D_{\text{KL}}(P_{H|v}\|Q_{H|v}) dP_V(v) \geq 0, \quad (10.3)$$

with equality if and only if $P_{H|v} = Q_{H|v}$ for P_V -a.e. v .

- (iv) **Gaussian specialisation.** If $P = \mathcal{N}(\mu_P, \Sigma_P)$ and $Q = \mathcal{N}(\mu_Q, \Sigma_Q)$ on $\mathbb{R}^{|V|+|H|}$, then

$$R = \frac{1}{2} \left[\log \frac{|\Sigma_{Q,H|V}|}{|\Sigma_{P,H|V}|} - |H| + \text{tr}(\Sigma_{Q,H|V}^{-1} \Sigma_{P,H|V}) + Q \right], \quad (10.4)$$

where $\Sigma_{\cdot,H|V} := \Sigma_{\cdot,HH} - \Sigma_{\cdot,HV} \Sigma_{\cdot,VV}^{-1} \Sigma_{\cdot,VH}$ is the $|H| \times |H|$ conditional covariance (covariance-space Schur complement, cf. Instance H), and

$$Q = \bar{\delta}^\top \Sigma_{Q,H|V}^{-1} \bar{\delta} + \text{tr}(\Sigma_{Q,H|V}^{-1} M \Sigma_{P,VV} M^\top) \geq 0, \quad (10.5)$$

with

$$\bar{\delta} := \mu_{Q,H} - \mu_{P,H} + B_Q(\mu_{P,V} - \mu_{Q,V}),$$

the P_V -mean conditional-mean difference, $M := B_Q - B_P$ the regression coefficient difference, and $B_\cdot := \Sigma_{\cdot,HV} \Sigma_{\cdot,VV}^{-1}$ the regression matrix. Both terms in Q are nonneg. After the regression contraction, the quadratic contribution lives in the $|H|$ -dimensional hidden Mahalanobis geometry, while the visible law enters through $\mu_{P,V}$ and $\Sigma_{P,VV}$.

Proof. By disintegration, $P(dv, dh) = P_V(dv)P_{H|v}(dh)$ and $Q(dv, dh) = Q_V(dv)Q_{H|v}(dh)$. Since $P \ll Q$, one may choose versions so that $P_{H|v} \ll Q_{H|v}$ for P_V -a.e. v , and the Radon-Nikodym derivative satisfies, P -a.e.,

$$\frac{dP}{dQ}(v, h) = \frac{dP_V}{dQ_V}(v) \frac{dP_{H|v}}{dQ_{H|v}}(h).$$

Taking logarithms and integrating against P gives

$$\begin{aligned} D_{\text{KL}}(P\|Q) &= \int \log \frac{dP_V}{dQ_V}(v) dP(v, h) + \int \log \frac{dP_{H|v}}{dQ_{H|v}}(h) dP(v, h) \\ &= D_{\text{KL}}(P_V\|Q_V) + \int \left[\int \log \frac{dP_{H|v}}{dQ_{H|v}}(h) dP_{H|v}(h) \right] dP_V(v), \end{aligned}$$

with equality in $[0, \infty]$. This gives (10.2).

(i) is immediate: P_V and Q_V are uniquely determined by P , Q .

(ii) is Gibbs' inequality: $D_{\text{KL}} \geq 0$ with equality iff the arguments coincide.

(iii) follows by applying Gibbs' inequality to the conditional divergence for P_V -a.e. v : $D_{\text{KL}}(P_{H|v} \| Q_{H|v}) \geq 0$, with equality in the integral iff the conditionals agree for P_V -a.e. v .

(iv) For Gaussian P and Q , the conditional distributions are Gaussian with covariances given by the Schur complements $\Sigma_{P,H|V}$ and $\Sigma_{Q,H|V}$ (which are constant in v) and means that are affine in v . The conditional KL is the Gaussian KL formula applied to these $|H|$ -dimensional Gaussians. The covariance-dependent part is the first three terms of (10.4). The mean-dependent part Q is the expected squared Mahalanobis distance between conditional means:

$$Q = \mathbb{E}_{v \sim P_V} [\delta(v)^\top \Sigma_{Q,H|V}^{-1} \delta(v)],$$

where $\delta(v) = \mu_{Q,H|V}(v) - \mu_{P,H|V}(v)$ is an affine function of v , giving a nonneg quantity computed in the hidden Mahalanobis geometry, with the visible dependence entering through the mean and covariance of P_V .

10.3 Refinement

Proposition 10.2 (Data processing refinement). *Let $\pi_1 = \rho \circ \pi_2$ where π_1, π_2 are observation maps with π_2 finer (more informative) than π_1 . Then*

$$D_{\text{KL}}(P_{\pi_1} \| Q_{\pi_1}) \leq D_{\text{KL}}(P_{\pi_2} \| Q_{\pi_2}). \quad (10.6)$$

That is, finer observation preserves more divergence (more information about the difference between P and Q).

Proof. Apply the chain rule (10.2) twice. With $V_1 \subseteq V_2$ (identifying visible sets):

$$D_{\text{KL}}(P_{V_2} \| Q_{V_2}) = D_{\text{KL}}(P_{V_1} \| Q_{V_1}) + \int D_{\text{KL}}(P_{V_2 \setminus V_1 | v_1} \| Q_{V_2 \setminus V_1 | v_1}) dP_{V_1}(v_1).$$

Since the second term is nonneg, $D_{\text{KL}}(P_{V_2} \| Q_{V_2}) \geq D_{\text{KL}}(P_{V_1} \| Q_{V_1})$. This is the data processing inequality.

Quotient descent reading:

- \mathcal{M} : the space of pairs of probability measures (P, Q) on a standard Borel product space.
- $B = D_{\text{KL}}(P \| Q)$: the full divergence (nonneg functional on pairs).
- π : marginalisation to the visible σ -algebra.
- $\bar{B} = D_{\text{KL}}(P_V \| Q_V)$: the marginal divergence.
- $R = \int D_{\text{KL}}(P_{H|v} \| Q_{H|v}) dP_V(v)$: the conditional divergence. For Gaussians, its covariance part is controlled by the $|H|$ eigenvalues of $\Sigma_{Q,H|V}^{-1} \Sigma_{P,H|V}$, and its mean/regression part is the P_V -average hidden Mahalanobis discrepancy in (10.5).
- Refinement: data processing inequality (Proposition 10.2).
- Boundary: $R = 0$ iff P and Q have identical conditional structure given V .

Proposition 10.3 (Second-order shadow of the KL chain rule). *Let $\{p_\theta(v, h) : \theta \in U\}$ be a C^1 parametric family of strictly positive densities on $\mathcal{X}_V \times \mathcal{X}_H$, with differentiation under the integral justified near $\theta_0 \in U$. Write*

$$p_\theta(v, h) = p_{V,\theta}(v) p_{H|V,\theta}(h | v)$$

and define the full score, marginal score, and conditional score at θ_0 by

$$\begin{aligned} s(v, h) &:= \nabla_{\theta} \log p_{\theta}(v, h) \Big|_{\theta_0}, \\ s_V(v) &:= \nabla_{\theta} \log p_{V, \theta}(v) \Big|_{\theta_0}, \\ s_{H|V}(v, h) &:= \nabla_{\theta} \log p_{H|V, \theta}(h | v) \Big|_{\theta_0}. \end{aligned}$$

Then

$$s = s_V + s_{H|V}, \quad \mathbb{E}_{\theta_0}[s_{H|V} | V] = 0,$$

and the Fisher information decomposes as

$$I(\theta_0) = I_V(\theta_0) + I_{H|V}(\theta_0), \quad (10.7)$$

where

$$I(\theta_0) = \mathbb{E}_{\theta_0}[ss^{\top}], \quad I_V(\theta_0) = \mathbb{E}_{\theta_0}[s_V s_V^{\top}], \quad I_{H|V}(\theta_0) = \mathbb{E}_{\theta_0}[s_{H|V} s_{H|V}^{\top}] \succeq 0.$$

In particular, the score-level decomposition is exact. Under the usual additional C^2 dominated-regularity hypotheses for local asymptotic expansion of KL, it is the second-order term of the KL chain rule and recovers Fisher descent with a positive conditional Fisher remainder.

Proof. Differentiate

$$\log p_{\theta}(v, h) = \log p_{V, \theta}(v) + \log p_{H|V, \theta}(h | v)$$

at θ_0 to obtain $s = s_V + s_{H|V}$. Since $p_{H|V, \theta_0}(\cdot | v)$ is a conditional density,

$$\mathbb{E}_{\theta_0}[s_{H|V} | V = v] = \int \nabla_{\theta} p_{H|V, \theta}(h | v) \Big|_{\theta_0} dh = \nabla_{\theta} \int p_{H|V, \theta}(h | v) dh \Big|_{\theta_0} = 0.$$

Therefore the cross term vanishes:

$$\mathbb{E}_{\theta_0}[s_V s_{H|V}^{\top}] = \mathbb{E}_{\theta_0}[s_V \mathbb{E}_{\theta_0}[s_{H|V}^{\top} | V]] = 0,$$

and similarly for its transpose. Expanding $I(\theta_0) = \mathbb{E}_{\theta_0}[(s_V + s_{H|V})(s_V + s_{H|V})^{\top}]$ gives (10.7). The final sentence is the standard local quadratic expansion of KL under regularity.

Remark (Hierarchy and limits). Proposition 10.3 makes one part of the hierarchy exact:

$$\text{Instance G (KL chain rule)} \xrightarrow{\text{local quadratic expansion}} \text{Instance A (Fisher descent)}.$$

When the parameter splits as $\theta = (\alpha, \beta)$ into visible and nuisance blocks and $I_{\beta\beta}$ is invertible, efficient elimination of the nuisance block produces the Schur complement

$$I_{\alpha\alpha}^{\text{eff}} = I_{\alpha\alpha} - I_{\alpha\beta} I_{\beta\beta}^{-1} I_{\beta\alpha},$$

which is the Fisher-side antecedent of Instance D in Gaussian families.

Two cautions matter. First, Instance 0 is not literally a KL chain rule: the Donsker-Varadhan functional is not, in general, the ordinary relative entropy of the empirical measure against stationarity. The relation to Instance G is therefore structural and local, not an equality of full functionals. Second, Instance B should not presently be placed on the same Taylor ladder. Its

cubic shadow is a generator-level obstruction term, not a proved third-order coefficient of a KL decomposition on path space.

Remark (Quantum boundary of Instance G). For density matrices and a quantum channel Φ , the Umegaki relative entropy satisfies the monotonicity inequality

$$S(\rho\|\sigma) \geq S(\Phi(\rho)\|\Phi(\sigma)).$$

So the monotonicity part of Instance G survives in the quantum setting. What fails in general is the classical exact decomposition

$$D_{\text{KL}}(P\|Q) = D_{\text{KL}}(P_V\|Q_V) + R$$

with a positive conditional remainder built from classical disintegrations. The obstruction is broader than entanglement alone: noncommutativity already prevents a classical conditional-probability calculus, and exact additive remainders survive only under additional commuting or classical-quantum structure. Entanglement is one important source of failure, but not the unique one.

Example 10.4. Let $V = \{1\}$, $H = \{2\}$, $P = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$, $Q = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, I_2\right)$, so P and Q differ only in the correlation ρ between visible and hidden. The marginal KL is $D_{\text{KL}}(P_V\|Q_V) = 0$ (both marginals are $\mathcal{N}(0, 1)$). All the divergence lives in the conditional:

$$R = \frac{1}{2}[-\log(1 - \rho^2) - 1 + (1 - \rho^2) + \rho^2] = -\frac{1}{2}\log(1 - \rho^2),$$

which is positive for $\rho \neq 0$. The hidden-visible correlation is invisible to the marginal but fully captured by the conditional remainder.

11 Instance H: Law of total covariance

11.1 Setting

Let $Y \in L^2(\Omega \rightarrow \mathbb{R}^p)$ be a square-integrable random vector and let $X : \Omega \rightarrow \mathcal{X}$ be a random variable on a common probability space, with \mathcal{X} standard Borel. Write $\mu(X) := \mathbb{E}[Y \mid \sigma(X)]$ for the conditional mean as an L^2 -projection. For a regular conditional version, write $\mu(x) := \mathbb{E}[Y \mid X = x]$ and $\Sigma(x) := \text{Cov}(Y \mid X = x)$. This is the Hilbert-space projection form of the same residual decomposition, expressed probabilistically as total covariance [44, 45].

11.2 The total covariance theorem

Theorem 11.1 (Covariance descent under conditioning). Under the finite second-moment hypothesis $\mathbb{E}[\|Y\|^2] < \infty$:

$$\text{Cov}(Y) = \text{Cov}(\mu(X)) + \mathbb{E}[\Sigma(X)], \quad (11.1)$$

where $\text{Cov}(\mu(X))$ is the between-group covariance (variance of the conditional mean) and $\mathbb{E}[\Sigma(X)]$ is the within-group covariance (expected conditional covariance). Moreover:

- (i) **Uniqueness.** Both terms in (11.1) are uniquely determined by the joint distribution of (X, Y) .
- (ii) **Positivity.** $\text{Cov}(\mu(X)) \succeq 0$ and $\mathbb{E}[\Sigma(X)] \succeq 0$.

(iii) **Gram factorisation.** Setting $f(X) := \mu(X) - \mathbb{E}[Y]$ and $\varepsilon := Y - \mu(X)$:

$$\text{Cov}(\mu(X)) = \mathbb{E}[f(X) f(X)^\top], \quad (11.2)$$

$$\mathbb{E}[\Sigma(X)] = \mathbb{E}[\varepsilon \varepsilon^\top]. \quad (11.3)$$

Both are integral Gram forms (expectations of outer products). For discrete X taking k values, $\text{rank}(\text{Cov}(\mu(X))) \leq \min(k-1, p)$.

(iv) **Gaussian specialisation.** If (X, Y) is jointly Gaussian with $X \in \mathbb{R}^q$, $Y \in \mathbb{R}^p$, block covariance Σ_{YX} , etc., then:

$$\mathbb{E}[\Sigma(X)] = \Sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \quad (\text{Schur complement of covariance}), \quad (11.4)$$

$$\text{Cov}(\mu(X)) = \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} = B_H B_H^\top, \quad B_H := \Sigma_{YX} \Sigma_{XX}^{-1/2}. \quad (11.5)$$

The Gram factor B_H has $\text{rank}(B_H) = \text{rank}(\Sigma_{YX}) \leq \min(p, q)$.

Proof. Write $Y = \mu(X) + \varepsilon$ where $\varepsilon := Y - \mu(X)$ satisfies $\mathbb{E}[\varepsilon | X] = 0$. Then:

$$\begin{aligned} \text{Cov}(Y) &= \text{Cov}(\mu(X) + \varepsilon) \\ &= \text{Cov}(\mu(X)) + \text{Cov}(\varepsilon) + \text{Cov}(\mu(X), \varepsilon) + \text{Cov}(\varepsilon, \mu(X)). \end{aligned}$$

The cross terms vanish: $\text{Cov}(\mu(X), \varepsilon) = \mathbb{E}[\mu(X) \varepsilon^\top] - \mathbb{E}[\mu(X)] \mathbb{E}[\varepsilon]^\top$. By iterated expectation,

$$\mathbb{E}[\mu(X) \varepsilon^\top] = \mathbb{E}[\mu(X) \mathbb{E}[\varepsilon^\top | X]] = 0,$$

and $\mathbb{E}[\varepsilon] = \mathbb{E}[\mathbb{E}[\varepsilon | X]] = 0$. So $\text{Cov}(Y) = \text{Cov}(\mu(X)) + \text{Cov}(\varepsilon)$. Finally, $\text{Cov}(\varepsilon) = \mathbb{E}[\varepsilon \varepsilon^\top] = \mathbb{E}[\mathbb{E}[\varepsilon \varepsilon^\top | X]] = \mathbb{E}[\Sigma(X)]$, giving (11.1).

(i): Both terms depend only on the joint distribution of (X, Y) .

(ii): $\text{Cov}(\mu(X)) = \mathbb{E}[f(X) f(X)^\top] \succeq 0$ as an expectation of PSD matrices. Similarly $\mathbb{E}[\Sigma(X)] \succeq 0$.

(iii): Equations (11.2) to (11.3) are immediate from the proof. For discrete X taking values $\{x_1, \dots, x_k\}$ with probabilities $\{p_1, \dots, p_k\}$, the between-group covariance is $\sum_{i=1}^k p_i (m_i - \bar{m})(m_i - \bar{m})^\top$ where $m_i = \mu(x_i)$ and $\bar{m} = \mathbb{E}[Y]$. This is the covariance of a distribution supported on at most k points in \mathbb{R}^p , so its rank is at most $\min(k-1, p)$ (the centring reduces the effective support by one dimension).

(iv): For jointly Gaussian (X, Y) , $\mu(X) = \mathbb{E}[Y] + \Sigma_{YX} \Sigma_{XX}^{-1} (X - \mathbb{E}[X])$ is affine in X , so $\text{Cov}(\mu(X)) = \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY}$. The conditional covariance $\Sigma(x) = \Sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY}$ is constant in x , so $\mathbb{E}[\Sigma(X)] = \Sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY}$, the covariance-space Schur complement.

11.3 Refinement

Proposition 11.2 (Refinement under additional conditioning). Let Z be an additional random variable. Then:

$$\mathbb{E}[\text{Cov}(Y | X, Z)] \preceq \mathbb{E}[\text{Cov}(Y | X)] \quad (11.6)$$

in the Loewner order. That is, conditioning on more variables can only reduce the expected residual covariance.

Proof. Apply the law of total covariance (11.1) within the X -conditional:

$$\text{Cov}(Y | X) = \text{Cov}(\mathbb{E}[Y | X, Z] | X) + \mathbb{E}[\text{Cov}(Y | X, Z) | X].$$

Taking expectations:

$$\mathbb{E}[\text{Cov}(Y | X)] = \mathbb{E}[\text{Cov}(\mathbb{E}[Y | X, Z] | X)] + \mathbb{E}[\text{Cov}(Y | X, Z)].$$

Since $\mathbb{E}[\text{Cov}(\mathbb{E}[Y | X, Z] | X)] \succeq 0$, we obtain (11.6).

Quotient descent reading:

- \mathcal{M} : the space of jointly distributed random vectors (X, Y) with $Y \in L^2$.
- $B = \text{Cov}(Y)$: the total covariance (PSD matrix).
- π : conditioning on X (the observation).
- $\bar{B} = \mathbb{E}[\text{Cov}(Y | X)]$: the expected conditional covariance (what remains after observation). $\bar{B} \preceq B$.
- $R = \text{Cov}(\mu(X))$: the between-group covariance (variance explained by X), with integral Gram form, rank \leq cardinality/dimension of X .
- Refinement: Proposition 11.2 (more conditioning reduces \bar{B}).
- Boundary: $\bar{B} = 0$ iff Y is X -measurable (no residual uncertainty).

Proposition 11.3 (Exact Gaussian duality of Instances H and D). *Let $X = (X_V, X_H)$ be jointly Gaussian with covariance*

$$\Sigma = \begin{pmatrix} \Sigma_{VV} & \Sigma_{VH} \\ \Sigma_{HV} & \Sigma_{HH} \end{pmatrix}$$

and precision $\Omega = \Sigma^{-1}$, partitioned conformally as

$$\Omega = \begin{pmatrix} \Omega_{VV} & \Omega_{VH} \\ \Omega_{HV} & \Omega_{HH} \end{pmatrix}.$$

Then the covariance-side and precision-side descended objects are exact dual Schur quantities:

$$\Sigma_{V|H} := \Sigma_{VV} - \Sigma_{VH}\Sigma_{HH}^{-1}\Sigma_{HV} = \Omega_{VV}^{-1}, \quad (11.7)$$

$$\bar{\Omega}_V := \Sigma_{VV}^{-1} = \Omega_{VV} - \Omega_{VH}\Omega_{HH}^{-1}\Omega_{HV}. \quad (11.8)$$

So conditioning in covariance space and marginalisation in precision space are dual operations linked by block inversion.

Proof. For jointly Gaussian vectors, the conditional covariance of X_V given X_H is the covariance-space Schur complement

$$\Sigma_{V|H} = \Sigma_{VV} - \Sigma_{VH}\Sigma_{HH}^{-1}\Sigma_{HV}.$$

The block inverse formula for $\Sigma^{-1} = \Omega$ shows that the (V, V) block of Ω is exactly $\Sigma_{V|H}^{-1}$, proving (11.7). The same block inverse formula, applied in the opposite direction, gives

$$\Sigma_{VV}^{-1} = \Omega_{VV} - \Omega_{VH}\Omega_{HH}^{-1}\Omega_{HV},$$

which is (11.8). The final sentence is just the interpretation of these two identities.

Remark (Duality with Instance D and foundation for the Schur complement). Instance D shows that the marginal *precision* of X_V is the Schur complement of the joint *precision* matrix.

Instance H shows that the conditional *covariance* of Y given X is the Schur complement of the joint *covariance* matrix. These are dual operations on dual objects: one acts on the precision (information), the other on the covariance (uncertainty). In the Gaussian case, both reduce to exact Schur identities, but they produce different quantities: Instance D gives a marginal precision, while Instance H gives a conditional covariance.

Crucially, Instance H is *more general* than Instance D: the law of total covariance holds for *all* distributions with finite second moments, not just Gaussians. The Schur complement that drives Instance D is the Gaussian specialisation of the conditional expectation projection that drives Instance H. This shows that the QDP is not fundamentally about the Schur complement; it is about orthogonal decomposition in L^2 , of which the Schur complement is a Gaussian shadow.

Example 11.4 . Let X take values $\{1, 2, 3\}$ with equal probability, and let $Y \in \mathbb{R}^2$ have conditional distributions:

$$Y | X=1 \sim \mathcal{N}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, 0.1I\right), \quad Y | X=2 \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}, 0.1I\right), \quad Y | X=3 \sim \mathcal{N}\left(\begin{pmatrix} -1 \\ -1 \end{pmatrix}, 0.1I\right).$$

The between-group covariance has rank $\min(3-1, 2) = 2$ (the three conditional means are not collinear in \mathbb{R}^2). The within-group covariance is $\mathbb{E}[\Sigma(X)] = 0.1I$ (homoscedastic noise). The total covariance $\text{Cov}(Y) = \text{Cov}(\mu(X)) + 0.1I$ is the sum of cluster structure and isotropic noise, a decomposition familiar from Gaussian mixture models.

11.4 A common Hilbert-space root behind G, A, and H

Proposition 11.5 (Hilbert-space projection root). Let \mathcal{H} be a real Hilbert space, let $V \subseteq \mathcal{H}$ be a closed subspace with orthogonal projection P , and let E be a finite-dimensional real Hilbert space. For a linear map $T : E \rightarrow \mathcal{H}$ define quadratic forms on E by

$$Q(u) := \|Tu\|_{\mathcal{H}}^2, \quad Q_P(u) := \|PTu\|_{\mathcal{H}}^2, \quad Q_{P^\perp}(u) := \|(I - P)Tu\|_{\mathcal{H}}^2. \quad (11.9)$$

Then:

(i) **Exact orthogonal split.**

$$Q = Q_P + Q_{P^\perp}. \quad (11.10)$$

(ii) **Positivity.** Both Q_P and Q_{P^\perp} are positive semidefinite quadratic forms, represented by the positive semidefinite operators

$$T^*PT \quad \text{and} \quad T^*(I - P)T.$$

(iii) **Refinement.** If $V_1 \subseteq V_2 \subseteq \mathcal{H}$ with orthogonal projections P_1, P_2 , then for every $u \in E$,

$$Q_{P_2^\perp}(u) \leq Q_{P_1^\perp}(u), \quad Q_{P_1}(u) \leq Q_{P_2}(u). \quad (11.11)$$

Equivalently, enlarging the visible subspace decreases the orthogonal remainder and increases the projected part in the Loewner order on quadratic forms.

(iv) **Boundary.** $Q_{P^\perp} \equiv 0$ if and only if $\text{Ran}(T) \subseteq V$.

Proof. For every $u \in E$, orthogonality of PTu and $(I - P)Tu$ gives the Pythagorean identity

$$\|Tu\|_{\mathcal{H}}^2 = \|PTu\|_{\mathcal{H}}^2 + \|(I - P)Tu\|_{\mathcal{H}}^2,$$

which is (11.10). Positivity in (ii) is immediate from the norm-square representation. For (iii), $Q_{P_i^\perp}(u)$ is the squared distance from Tu to the closed subspace V_i . Since $V_1 \subseteq V_2$, distance to V_2 is no larger than distance to V_1 , proving the first inequality in (11.11). The second then follows from $Q = Q_{P_i} + Q_{P_i^\perp}$. Finally, $Q_{P^\perp} \equiv 0$ holds if and only if $(I - P)Tu = 0$ for all u , which is equivalent to $Tu \in V$ for all u , that is, $\text{Ran}(T) \subseteq V$.

Corollary 11.6 (Instances A and H as shadows of Proposition 11.5). *The clean classical branches $G \rightarrow A$ and H arise from Proposition 11.5 by two different choices of (\mathcal{H}, V, T) .*

- (i) **Fisher branch.** *In Proposition 10.3, take $\mathcal{H} = L^2(P_{\theta_0})$, let $V = L^2(\sigma(X))$ be the closed subspace of X -measurable square-integrable functions, and define $T(v) = S_v$ for the score in direction $v \in T_{\theta_0}\Theta$. Then*

$$Q(v) = I_{\text{full}}(v, v), \quad Q_P(v) = I_{\text{marg}}(v, v), \quad Q_{P^\perp}(v) = I_{\text{cond}}(v, v),$$

so Proposition 10.3 is exactly the quadratic-form pullback of (11.10).

- (ii) **Covariance branch.** *In Instance H, take $\mathcal{H} = L^2(\Omega)$, let $V = L^2(\sigma(X))$, and define*

$$T(a) = a^\top(Y - \mathbb{E}[Y]), \quad a \in \mathbb{R}^P.$$

Then

$$Q(a) = a^\top \text{Cov}(Y)a,$$

while the projection and orthogonal remainder are

$$Q_P(a) = a^\top \text{Cov}(\mu(X))a, \quad Q_{P^\perp}(a) = a^\top \mathbb{E}[\Sigma(X)]a.$$

So Theorem 11.1 is the covariance shadow of the same Hilbert projection identity. In the quotient-descent convention of Instance H, the descended residual covariance is the orthogonal remainder Q_{P^\perp} , while the projected part Q_P is the explained covariance carried by the observation.

Remark (The exact classical compression). The strongest compression achieved in the present note is now more exact than before. Proposition 11.5 gives the compression side of the classical core, while Theorem 11.7 gives the quotient side. Remark 11.5 identifies them as dual faces of one induced visible Hilbert structure. Instance D then appears as the coordinate Gaussian case of the quotient theorem, Instance F as its grounded semidefinite boundary-energy version, and Proposition 11.10 shows that the gradient subcase of Instance C is the same quotient geometry applied frequency by frequency. Proposition 3.2 shows that Instance 0 also admits an exact local lift into this broad geometry, though at present only after canonical tangent-space doubling. This is a real compression of the note, not a slogan.

11.5 A common Hilbert-quotient root behind *Quotient*, D, F, and gradient C

The constrained quadratic minimisation in (11.14) below is the classical weighted minimum-norm problem, and the formula $L_{C,H} = H^{-1}C^\top(CH^{-1}C^\top)^{-1}$ is its standard generalised-inverse solution [46–48]. The contribution of Theorem 11.7 below is the quotient-descent reading of that classical algebra: the visible quadratic form $\Phi_C(H)$, the canonical minimal lift $L_{C,H}$, and the hidden projector $P_{C,H}$ identified together as a single induced visible structure dual to compression.

Theorem 11.7 (Hilbert quotient root and canonical minimal lift). Let E and Y be finite-dimensional real inner-product spaces, let $H \in \text{Sym}_{++}(E)$, and let $C : E \rightarrow Y$ be a surjective linear map. Define

$$\Phi_C(H) := (CH^{-1}C^\top)^{-1} \quad (11.12)$$

and the H -minimal lift

$$L_{C,H} := H^{-1}C^\top\Phi_C(H). \quad (11.13)$$

Then:

(i) **Unique minimal lift.** For every $y \in Y$, the constrained problem

$$\min\{x^\top Hx : Cx = y\} \quad (11.14)$$

has the unique minimiser $x_* = L_{C,H}y$.

(ii) **Canonical visible subspace.** One has

$$CL_{C,H} = I_Y, \quad (11.15)$$

and

$$\text{Ran}(L_{C,H}) = (\ker C)^{\perp H}. \quad (11.16)$$

So the visible space is identified with the H -orthogonal complement of the hidden fibre.

(iii) **Visible quadratic form.** The visible form is exactly the upstairs form restricted to the canonical lift:

$$\Phi_C(H) = L_{C,H}^\top H L_{C,H}, \quad (11.17)$$

and for every $y \in Y$,

$$y^\top \Phi_C(H)y = \min_{Cx=y} x^\top Hx. \quad (11.18)$$

(iv) **Hidden projector.** If

$$P_{C,H} := I_E - L_{C,H}C, \quad (11.19)$$

then $P_{C,H}^2 = P_{C,H}$, $\text{Ran}(P_{C,H}) = \ker C$, and $\text{Ran}(L_{C,H})$ is H -orthogonal to $\text{Ran}(P_{C,H})$. Thus $P_{C,H}$ is the H -orthogonal projector onto the hidden fibre.

Proof. Let $y \in Y$. The functional $x \mapsto x^\top Hx$ is strictly convex because $H \succ 0$, so any feasible critical point is the unique minimiser. Introduce a Lagrange multiplier $\lambda \in Y$ and set

$$\mathcal{L}(x, \lambda) = x^\top Hx - 2\lambda^\top (Cx - y).$$

Criticality gives $Hx = C^\top \lambda$, hence $x = H^{-1}C^\top \lambda$. Imposing the constraint yields

$$CH^{-1}C^\top \lambda = y,$$

so $\lambda = \Phi_C(H)y$ and therefore $x = L_{C,H}y$. This proves (i). Equation (11.15) is immediate:

$$CL_{C,H} = CH^{-1}C^\top \Phi_C(H) = I_Y.$$

For any $z \in \ker C$ and $y \in Y$,

$$\langle L_{C,H}y, z \rangle_H = y^\top \Phi_C(H)Cz = 0,$$

so $\text{Ran}(L_{C,H}) \subseteq (\ker C)^{\perp H}$. Both spaces have dimension $\dim Y$, because C is surjective, so (11.16) follows.

For (iii),

$$L_{C,H}^\top H L_{C,H} = \Phi_C(H) C H^{-1} H H^{-1} C^\top \Phi_C(H) = \Phi_C(H) (C H^{-1} C^\top) \Phi_C(H) = \Phi_C(H),$$

which is (11.17). Since $x_* = L_{C,H} y$ is the unique minimiser, (11.18) follows.

For (iv),

$$P_{C,H}^2 = (I_E - L_{C,H} C)^2 = I_E - 2L_{C,H} C + L_{C,H} (C L_{C,H}) C = I_E - L_{C,H} C = P_{C,H},$$

using (11.15). Also $C P_{C,H} = C - C L_{C,H} C = 0$, so $\text{Ran}(P_{C,H}) \subseteq \ker C$. Conversely, if $z \in \ker C$, then $P_{C,H} z = z$, hence $\text{Ran}(P_{C,H}) = \ker C$. Finally, by (11.16), $\text{Ran}(L_{C,H}) = (\ker C)^{\perp H} = (\text{Ran}(P_{C,H}))^{\perp H}$.

Proposition 11.8 (First visible response and quartic hidden leakage). *In the setting of Theorem 11.7, let $\Delta \in \text{Sym}(E)$. Then:*

(i) *First visible response.*

$$D\Phi_C(H)[\Delta] = L_{C,H}^\top \Delta L_{C,H}. \quad (11.20)$$

(ii) *Quartic hidden leakage.* Define

$$Q_{C,H}(\Delta) := -\frac{1}{2} D^2 \Phi_C(H)[\Delta, \Delta]. \quad (11.21)$$

Then

$$Q_{C,H}(\Delta) = L_{C,H}^\top \Delta P_{C,H} H^{-1} \Delta L_{C,H}, \quad (11.22)$$

and, equivalently,

$$Q_{C,H}(\Delta) = (H^{-1/2} P_{C,H}^\top \Delta L_{C,H})^\top (H^{-1/2} P_{C,H}^\top \Delta L_{C,H}) \succeq 0. \quad (11.23)$$

In particular,

$$Q_{C,H}(\Delta) = 0 \iff P_{C,H}^\top \Delta L_{C,H} = 0. \quad (11.24)$$

Proof. Write $A := H^{-1}$ and $K(H) := C A C^\top$, so $\Phi_C(H) = K(H)^{-1}$. Along the line $H + t\Delta$ one has

$$\left. \frac{d}{dt} A(t) \right|_{t=0} = -A\Delta A, \quad \left. \frac{d^2}{dt^2} A(t) \right|_{t=0} = 2A\Delta A\Delta A.$$

Hence

$$DK(H)[\Delta] = -C A \Delta A C^\top, \quad D^2 K(H)[\Delta, \Delta] = 2C A \Delta A \Delta A C^\top.$$

Using $D(K^{-1})[\dot{K}] = -K^{-1} \dot{K} K^{-1}$ gives

$$D\Phi_C(H)[\Delta] = \Phi_C(H) C A \Delta A C^\top \Phi_C(H) = L_{C,H}^\top \Delta L_{C,H},$$

which proves (11.20).

For the second derivative,

$$D^2 \Phi_C(H)[\Delta, \Delta] = 2\Phi_C(H) DK(H)[\Delta] \Phi_C(H) DK(H)[\Delta] \Phi_C(H) - \Phi_C(H) D^2 K(H)[\Delta, \Delta] \Phi_C(H).$$

Substituting the formulas above and rearranging yields

$$-\frac{1}{2} D^2 \Phi_C(H)[\Delta, \Delta] = \Phi_C(H) C A \Delta (A - A C^\top \Phi_C(H) C A) \Delta A C^\top \Phi_C(H).$$

Since $L_{C,H} = AC^\top\Phi_C(H)$ and $P_{C,H} = I_E - L_{C,H}C = I_E - AC^\top\Phi_C(H)C$, this is exactly (11.22). Moreover

$$HP_{C,H} = H - C^\top\Phi_C(H)C = P_{C,H}^\top H,$$

so $P_{C,H}H^{-1} = H^{-1}P_{C,H}^\top$. Therefore

$$Q_{C,H}(\Delta) = L_{C,H}^\top\Delta H^{-1}P_{C,H}^\top\Delta L_{C,H} = (H^{-1/2}P_{C,H}^\top\Delta L_{C,H})^\top(H^{-1/2}P_{C,H}^\top\Delta L_{C,H}),$$

which proves (11.23). The vanishing criterion (11.24) is immediate from the Gram form.

Corollary 11.9 (Tower laws for minimal lifts and quartic defects). *Let $C_1 : E \rightarrow Y_1$ and $C_2 : Y_1 \rightarrow Y_2$ be surjective, and write $C := C_2C_1$. Set*

$$H_1 := \Phi_{C_1}(H), \quad L_1 := L_{C_1,H}, \quad L_2 := L_{C_2,H_1}, \quad L := L_{C,H}.$$

Then:

(i) **Minimal lifts compose.**

$$L = L_1L_2. \quad (11.25)$$

(ii) **Visible first response composes.** For every $\Delta \in \text{Sym}(E)$,

$$D\Phi_C(H)[\Delta] = L_2^\top D\Phi_{C_1}(H)[\Delta]L_2. \quad (11.26)$$

Equivalently, if $\Delta_1 := D\Phi_{C_1}(H)[\Delta]$, then

$$D\Phi_C(H)[\Delta] = D\Phi_{C_2}(H_1)[\Delta_1]. \quad (11.27)$$

(iii) **Quartic defect has a tower law.**

$$Q_{C,H}(\Delta) = L_2^\top Q_{C_1,H}(\Delta)L_2 + Q_{C_2,H_1}(\Delta_1). \quad (11.28)$$

Proof. Since $C = C_2C_1$ and $H^{-1} = C_1H^{-1}C_1^\top$,

$$\Phi_C(H) = (C_2H_1^{-1}C_2^\top)^{-1} = \Phi_{C_2}(H_1).$$

Then

$$L_1L_2 = H^{-1}C_1^\top\Phi_{C_1}(H)H_1^{-1}C_2^\top\Phi_{C_2}(H_1) = H^{-1}C_1^\top C_2^\top\Phi_C(H) = H^{-1}C^\top\Phi_C(H) = L,$$

proving (11.25). Equation (11.26) now follows from Proposition 11.8(i):

$$D\Phi_C(H)[\Delta] = L^\top\Delta L = L_2^\top L_1^\top\Delta L_1L_2 = L_2^\top D\Phi_{C_1}(H)[\Delta]L_2.$$

Since Proposition 11.8(i) also gives $D\Phi_{C_2}(H_1)[\Delta_1] = L_2^\top\Delta_1L_2$, this is exactly (11.27).

For the second-order statement, use the composition identity $\Phi_C = \Phi_{C_2} \circ \Phi_{C_1}$ and differentiate twice along $H + t\Delta$. The chain rule gives

$$D^2\Phi_C(H)[\Delta, \Delta] = D\Phi_{C_2}(H_1)[D^2\Phi_{C_1}(H)[\Delta, \Delta]] + D^2\Phi_{C_2}(H_1)[\Delta_1, \Delta_1].$$

Applying Proposition 11.8(i) to the first term and the definition of Q to both terms yields (11.28).

Remark (Induced visible Hilbert structure: compression and quotient). Proposition 11.5 and Theorem 11.7 are two faces of one classical mechanism. Proposition 11.5 describes the compression side: one starts with an ambient Hilbert space and projects onto a visible

subspace. Theorem 11.7 describes the quotient side: one starts with an upstairs precision geometry and minimises over the hidden fibre. The two are dual under Riesz inversion:

$$\Phi_C(H)^{-1} = CH^{-1}C^\top. \quad (11.29)$$

So quotient on the primal-state side is compression on the dual-observable side.

Both faces are Pythagoras in dual inner products. Proposition 11.5 is the Pythagorean identity in the ambient L^2 inner product on \mathcal{H} :

$$\|Tu\|_{\mathcal{H}}^2 = \|PTu\|_{\mathcal{H}}^2 + \|(I - P)Tu\|_{\mathcal{H}}^2.$$

Theorem 11.7(iv) yields Pythagoras in the upstairs precision inner product on E : every $x \in E$ admits the H -orthogonal decomposition $x = L_{C,H}Cx + P_{C,H}x$, hence

$$x^\top Hx = (Cx)^\top \Phi_C(H)(Cx) + (P_{C,H}x)^\top H(P_{C,H}x).$$

The Riesz identity (11.29) links the two: $\Phi_C(H)$ on the quotient side is the inverse of the Gram matrix of C^\top in the dual inner product on the compression side.

This identifies several earlier instances more sharply than before. Instance D is the coordinate Gaussian case of Theorem 11.7. After grounding one node, or equivalently quotienting out the constant mode, Instance F is the semidefinite boundary-energy version of the same theorem. In the gradient subcase of Instance C, the dynamic Schur kernel is the same quotient theorem applied frequency by frequency to $H_s = sI + B$.

In the local Gaussian or exponential-family case, the duality (11.29) has the form of the dual coordinate structure of an Amari–Chentsov dually flat manifold [9, 10]: marginalisation in expectation coordinates corresponds to compression in natural coordinates, and the Schur algebra in either direction may be read as the linear shadow of that information-geometric duality. The present note does not claim a full information-geometric formulation, only that the lens has this candidate home.

Proposition 11.10 (Gradient Instance C as frequencywise Hilbert quotient). *In the gradient subcase of Theorem 6.1(iv), let $C_r(r, h) := r$ be the coordinate projection onto the resolved variables and set*

$$H_s := sI + B, \quad s > 0. \quad (11.30)$$

Then

$$\Sigma_r(s) = \Phi_{C_r}(H_s) = sI_r + B_{rr} - B_{rh}(sI_h + B_{hh})^{-1}B_{hr}, \quad (11.31)$$

and therefore

$$H(s) = G_r^\top \Phi_{C_r}(H_s)^{-1} G_r. \quad (11.32)$$

Thus the gradient hidden-elimination theorem is exact Hilbert quotient descent applied at each frequency.

Proof. Write H_s in the (r, h) block decomposition:

$$H_s = \begin{pmatrix} sI_r + B_{rr} & B_{rh} \\ B_{hr} & sI_h + B_{hh} \end{pmatrix}.$$

The coordinate projection C_r selects the r -block, so Theorem 11.7 gives

$$\Phi_{C_r}(H_s) = H_{s,rr} - H_{s,rh}H_{s,hh}^{-1}H_{s,hr},$$

which is exactly (11.31). The transfer representation (11.32) is then the formula already proved in Theorem 6.1(iv).

11.6 A scalar cross-instance invariant: the determinant clock

Geometric Observation [3] introduces the scalar invariant

$$\tau(\Lambda) := \log \det(I_S + \Lambda), \quad (11.33)$$

attached to the hidden-load operator Λ beneath a fixed visible ceiling on the support S . The clock is additive under sequential composition (the transport law of [3]). Several instances in the present note carry a natural specialisation of this scalar.

- **Instances D and H (Gaussian).** The Schur identity $\bar{\Omega} = \Sigma_{VV}^{-1}$ together with $\det \Omega = \det \bar{\Omega} \cdot \det \Omega_{HH}$ gives the determinant split

$$\log \det \bar{\Omega} = \log \det \Omega - \log \det \Omega_{HH}, \quad \log \det \Sigma_{V|H} = \log \det \Sigma - \log \det \Sigma_{HH}.$$

The Gram-factor rank $\text{rank}(R) = \text{rank}(\Omega_{VH})$ in Theorem 7.1(iii) and $\text{rank}(B_H) = \text{rank}(\Sigma_{YX})$ in Theorem 11.1(iv) have the form of the number of nontrivial canonical correlations between visible and hidden coordinates [28, 29], and the orthogonal indeterminacy $\bar{B}' = \bar{B}Q$ with $Q \in O(r)$ in Theorem 7.1(iii) has the form of the rotation freedom of factor-analysis loadings. The scalar τ then has the form of twice the local Gaussian mutual information between visible and hidden blocks [3].

- **Instance G (Gaussian).** The Gaussian conditional KL formula (10.4) contains the contribution

$$\frac{1}{2} \log(|\Sigma_{Q,H|V}|/|\Sigma_{P,H|V}|),$$

which has the form of a τ -difference between the conditional covariance Schur complements of P and Q .

- **Hilbert-quotient theorem.** For composed surjections $C = C_2C_1$ as in Corollary 11.9, the determinant identity

$$\det \Phi_C(H) = \det \Phi_{C_2}(\Phi_{C_1}(H))$$

gives an additive separation of $\log \det \Phi$ across stages, by the Schur quotient identity [25].

The determinant clock is therefore a candidate scalar cross-instance invariant whose existing instance-level shadows are recorded above. Its full Λ -form requires a declared visible ceiling and is developed in [3]; the present note records only the ceiling-free shadows that survive without that extra datum.

These shadows are the multiplicative dual of the additive conservation property (S5) below in the Gaussian block-PSD instances (D, F, H, and the gradient subcase of C). Wherever the upstairs object $B \succ 0$ admits a block decomposition with hidden block B_{HH} and visible Schur descendant \bar{B} , the standard log-determinant identity

$$\log \det B = \log \det \bar{B} + \log \det B_{HH} \quad (11.34)$$

gives the multiplicative shadow of $B = \bar{B} + R$. The visible scalar $\log \det \bar{B}$ is the local Gaussian form of the determinant clock on those instances.

12 Common structure

Theorem 12.1 (Synthesis). *The nine explicit instances in this document satisfy the quotient descent principle (Definition 2.2). Their common structure is:*

	<i>Full form</i>	<i>Descent</i>	<i>Hidden</i>	<i>Conservation</i>	<i>Boundary</i>
0 (DV)	H_{DV}	<i>Var. contr.</i>	$\frac{1}{4}\widehat{J}H_0^{-1}\widehat{J}^T$ Gram	<i>Additive</i>	$\widehat{J}=0$ (DB)
A (Fisher)	$I_E(\theta)$	<i>Chain rule</i>	<i>Vert. degen.</i>	<i>Add. (degen.)</i>	<i>Exp. class</i>
B (CME)	(b, D)	<i>Taylor trunc.</i>	<i>Cubic rem.</i> (sign-indef.)	<i>Cubic obstr.</i>	<i>3rd order</i>
C (Dissip.)	$H(s)$	<i>Resolvent Schur</i>	<i>Memory kernel</i>	<i>Add. (grad.)</i>	<i>Memory kernel</i>
D (Graph.)	Ω	<i>Marg. Schur</i>	<i>Low-rank Schur</i>	<i>Add. (S-L)</i>	<i>Fill-in</i>
E (Gram.)	W_C	<i>Lyap. add.</i>	$W_{C_{\text{new}}}$ Gram	<i>Additive</i>	<i>Unobs. subsp.</i>
F (Net.)	L	<i>Interior Schur</i>	<i>Bdry-energy</i>	<i>Additive</i>	<i>Fill-in</i>
G (KL)	D_{KL}	<i>Chain rule</i>	<i>Cond. KL int.</i>	<i>Add. (chain)</i>	<i>Cond. agree</i>
H (Tot. cov.)	$\text{Cov}(Y)$	L^2 proj.	$\text{Cov}(\mu(X))$ Gram	<i>Add. (eve)</i>	Y is X -meas.

In each case:

- (S1) *The descended object is uniquely determined by the full form and the observation.*
- (S2) *Positivity of the full form implies positivity of the descended form.*
- (S3) *The hidden contribution admits an explicit factorisation or remainder representation, with rank or rank bound measuring hidden complexity when that notion is available.*
- (S4) *Further observation is compatible with descent, either by pullback, monotonicity, data processing, or staged elimination, depending on the instance.*
- (S5) *In eight of the nine instances the descended object and the hidden contribution combine additively without remainder: full = visible + hidden, equivalently $B = \bar{B} + R$ in the natural sense of the instance. The single exception is Instance B, whose first algebraic discrepancy is the sign-indefinite cubic obstruction of Proposition 5.4.*

Proof. Each property is verified instance by instance. We give the key reference for each.

(S1) *Uniqueness:* Instance 0 by Legendre contraction; A by Theorem 4.1(i); B by Theorem 5.1(i); C by Theorem 6.1(ii); D by Theorem 7.1(i); E by Theorem 8.1(i) (Lyapunov uniqueness); F by Theorem 9.1(ii); G by Theorem 10.1(i); H by Theorem 11.1(i).

(S2) *Positivity:* Instance 0 by $\Delta_{DV} \succeq 0$; A by Fisher nonnegativity; B by $D_{\Pi} \succeq 0$; C by positive-realness; D by Haynsworth inertia (Theorem 7.1(ii)); E by integral representation (Theorem 8.1(ii)); F by Theorem 9.1(i); G by Gibbs' inequality (Theorem 10.1(ii) and (iii)); H by Theorem 11.1(ii).

(S3) *Controlled hidden structure:* Instance 0 by Proposition 3.2 together with $\text{rank}(\Delta_{DV}) = \text{rank}(\widehat{J})$; A by the vertical kernel $\ker dq$; B by the exact cubic remainder; C by the hidden memory kernel and, in the gradient case, Proposition 11.10; D by Theorem 7.1(iii) together with Theorem 11.7; E by the integral Gram form (Theorem 8.1(iv)); F by Theorem 9.1(iv) together with Remark 11.5; G by Theorem 10.1(iv) together with Proposition 10.3; H by Theorem 11.1(iii) (integral Gram form, $\text{rank} \leq \min(k-1, p)$).

(S4) *Compatibility under refinement:* Instance 0 by the converter formula; A by Proposition 4.2; B by Proposition 5.2; C by Proposition 6.2; D by Proposition 7.2; E by Theorem 8.1(v); F by Proposition 9.2; G by Proposition 10.2 (data processing inequality); H by Proposition 11.2 (Loewner monotonicity).

(S5) *Exact additive conservation:* Instance 0 by $H_{DV} = H_0 + \Delta_{DV}$ (Theorem 3.1); A by $I_E = \bar{I}_E$ on quotient tangent classes with vanishing vertical part (Theorem 4.1); C in the

gradient subcase by $H(s) - H_{\text{bare}}(s) \succeq 0$ with explicit Schur summand (Theorem 6.1(iv)); D by $\bar{\Omega} = \Omega_{VV} - R$ (Theorem 7.1); E by $W_{C_2} = W_{C_1} + W_{C_{\text{new}}}$ (Theorem 8.1(iii)); F by $L_{BB} = \bar{L} + R$ (Theorem 9.1); G by the chain rule (10.2); H by the law of total covariance (11.1). Instance B is the exception by Proposition 5.4.

Remark (Why this was not found before). The nine explicit instances span eight communities: large deviations (Instance 0), statistical inference (A), chemical kinetics (B), control/systems theory (C, E), graphical models (D), network science (F), information theory (G), and probability/statistics (H). Each community had its own version of hidden elimination. What was missing was the recognition that they all share a common abstract pattern with the same four structural properties. The reason is that the descent mechanisms are superficially different (variational contraction, chain rule, Taylor truncation, resolvent Schur complement, L^2 projection, marginalisation), so the unity is visible only after one identifies the common role of the nonnegative bilinear form and the observation quotient. The DV bridge theorem of Instance 0 was the key that revealed this structure, because it sits at the intersection of large deviations (the rate function), information geometry (Fisher coordinates), and linear algebra (the Schur complement), and its proof is elementary enough to make the abstract pattern unmistakable.

Remark (The hierarchy of instances). The reliable hierarchy is now sharper than before because the exact classical core has two dual faces.

$$\text{Induced visible Hilbert structure} \rightsquigarrow \begin{cases} \text{Compression side: } G \rightarrow A, H, E, \\ \text{Quotient side: } \Phi, D, F, \text{ gradient } C. \end{cases}$$

On the compression side, Proposition 11.5 and Corollary 11.6 show that the KL, Fisher, covariance, and observability branches arise from orthogonal projection in an ambient Hilbert space. On the quotient side, Theorem 11.7 shows that quotient-visible precision is the canonical minimal-lift form induced by the upstairs precision geometry, Proposition 11.8 identifies its first visible response and quartic hidden leakage, and Corollary 11.9 gives the tower laws. Proposition 11.10 shows that the gradient subcase of Instance C belongs to this quotient sector frequencywise.

The two sides are dual by Riesz inversion through (11.29). So the recurring Schur algebra in the note is not a primitive starting point. It is one coordinate shadow of a more basic induced visible Hilbert structure.

Instance 0 now sits closer to this picture than before. Proposition 3.2 gives an exact *local* Hilbert lift of the Donsker-Varadhan bridge after canonical doubling of the tangent space. What remains open is whether that bridge already comes from an intrinsic upstairs carrier whose quotient-visible descendant is the closed object of *Quotient Observation*. This still does *not* justify collapsing every instance into one Taylor ladder, and in particular it does not identify Instance B with Instance G.

Remark (What is not claimed). This document does not assert one abstract theorem covering all hidden-to-visible reductions. It does not identify a single invariant class containing all nine explicit instances. It does not claim that every reduction problem must fall into one of these forms. What it does provide is a precise characterisation of the common structure, with each instance proved at a uniform standard, and a clear identification of where the boundaries lie.

13 What this feeds back into *Quotient Observation*

Remark (Four exact feedback points for *Quotient Observation*). The present note does feed back into *Quotient Observation*, but in a disciplined way.

First, it clarifies why the positive-cone level is the right global level. Theorem 11.7 shows that quotient-visible precision is the Riesz operator of the quotient norm induced by the upstairs precision geometry. So the passage from tangent bridge data to the intrinsic visible precision is not an ad hoc replacement. It is the quotient-side face of the exact classical Hilbert structure.

Second, it clarifies why Schur mediation is not an isolated matrix trick. Proposition 11.3 and Remark 11.5 show that the Gaussian precision-side Schur object is dual to conditional covariance descent and sits inside one induced visible Hilbert geometry. In that sense the Schur structure that dominates *Quotient Observation* is not a primitive starting point. It is the coordinate shadow of a more basic quotient-versus-compression mechanism.

Third, it sharpens the derivative and quartic defect package. Proposition 11.8 shows that the first visible response is the upstairs perturbation pulled back to the canonical minimal visible lifts, while the quartic defect is the hidden-projector norm of that perturbation. So the sign-definite quartic term in *Quotient Observation* is exactly a hidden-leakage law, not merely a convenient positive correction.

Fourth, it sharpens the global hidden-load interpretation. The determinant clock in *Quotient Observation* is most naturally read as the log-volume loss between the tangent ceiling and the true descended visible law on the active support. The present note does not reprove that theorem, but the minimal-lift and hidden-projector picture explains why such a scalar should sit naturally above the local quartic defect.

What is *not* claimed is that the present note subsumes *Quotient Observation*. The positive-cone transport law, hidden-load cone, determinant clock, and closure analysis of Φ remain stronger structures specific to that paper. Likewise, the graphical and network Schur instances here should not be read as a universal gluing theorem: they establish exact positive descent inside their own SPD and Dirichlet-energy settings, while common-gluing obstructions and Bell-square phenomena remain separate constraints studied in *Quotient Observation*.

13.1 Immediate internal targets

The most important next wins suggested by the present note are now sharper than before.

1. Identify the intrinsic upstairs carrier behind the Donsker-Varadhan bridge. Proposition 3.2 gives an exact local lift of the bridge, but still extrinsically. The right next question is no longer only whether the bridge can be made to look more Hilbert-like. It is whether there exists an intrinsic upstairs positive carrier whose quotient-visible descendant is the closed object of *Quotient Observation* and whose local tangent shadow is the Donsker-Varadhan bridge data.

2. Attack the true non-selfadjoint frontier directly. Proposition 11.10 shows that the gradient subcase of Instance C already belongs to the quotient sector frequencywise, and Instance E already belongs to the compression sector. So the real unresolved frontier is narrower: the non-gradient positive-real or passive sector. The right target is an operator-valued minimal-lift or hidden-projector theorem in a Hardy, Herglotz, or storage-function setting.

3. Relate the global hidden-load theory to local hidden leakage. Proposition 11.8 identifies the quartic defect as hidden leakage through the hidden projector. The next exact bridge back into *Quotient Observation* is to determine whether the determinant clock and hidden-load cone are the nonlinear integrated version of that same geometry.

4. Formalise first-obstruction order as a reusable notion. Instance B and the defect package in *Quotient Observation* suggest a common methodology: after exact descent is identified, define the first algebraic or analytic order at which non-descended structure reappears. A precise formulation of that question now looks more valuable than adding many more borderline examples.

14 Transfer boundaries and candidate domains

This section is deliberately separated from the proved core of the note. The items below are not claimed as instances; they are candidate domains where the same hidden-to-visible pattern may be worth testing. They are ordered by proximity to the proved material.

14.1 Near-transfer domains

Quantum state tomography with restricted measurements. The quantum case should be treated as a boundary test, not as an automatic transfer. Monotonicity of quantum relative entropy under channels is robust, but the classical chain-rule decomposition of Instance G generally fails without extra commuting or classical-quantum structure. The right question is therefore which parts of quotient descent survive under noncommutativity, and which require a genuinely classical conditional calculus [49–51].

Latent-variable graphical models. The work of Chandrasekaran, Parrilo, and Willsky on latent variable selection via convex optimisation is a direct instance of hidden-to-visible Schur complement structure on precision matrices. The proved Gaussian algebra in Instance D suggests testing whether the hidden contribution (the latent variables' effect on the observed precision) admits a low-rank Gram factorisation governed by the number of effective latent variables [30, 31].

Causal inference with hidden confounders. The presence of unmeasured confounders in a causal graph creates a structural quotient: two causal models are observationally equivalent if they induce the same distribution on observed variables. Instance A suggests looking for a descended observational Fisher object that characterises what is identifiable, together with a controlled hidden contribution, while keeping this separate from full graphical identifiability claims [8, 55–57].

14.2 Medium-transfer domains

Renormalisation group flows. The RG flow eliminates UV degrees of freedom, defining a quotient from the full theory to the effective theory. The structural analogy is that effective actions inherit constrained positivity from the full theory, while integrated-out modes enter through a Wilsonian elimination mechanism. Whether that can be cast as a genuine quotient-descent theorem is open here [52–54].

Passive and positive-real model reduction. The sharp internal frontier now sits here. The selfadjoint quotient geometry already covers the gradient subcase of Instance C, while the genuinely unresolved case is non-selfadjoint positive-real hidden elimination. The right test

is whether reduced transfer functions inherit a canonical minimal-lift or hidden-projector structure in a Hardy, Herglotz, or storage-function setting.

Neural network compression and pruning. A trained neural network defines a map from inputs to outputs through a high-dimensional hidden representation. Pruning or compressing the network defines a quotient. A plausible test is whether Fisher information with respect to parameters descends to the pruned network, with the hidden contribution identifying which parameters are structurally invisible.

14.3 Far-transfer domains

Evolutionary population genetics with unobserved alleles. Observed allele frequencies in a population are a linear projection of the full allele frequency vector. A concrete test would be whether the drift and diffusion of the Wright-Fisher or Moran process, projected onto observed alleles, exhibit a quadratic shadow analogous to Instance B.

Cosmological parameter estimation with unobserved fields. The CMB power spectrum is determined by cosmological parameters, some of which are degenerate under observation. A natural question is whether the structural quotient of the parameter space carries a descended Fisher form characterising what is measurable.

Protein folding with hidden degrees of freedom. The observable (e.g. FRET-measured) degrees of freedom of a protein are a linear projection of its full conformational state. A natural test case is whether the observed fluctuation structure decomposes into a backbone contribution and a hidden-load correction.

References

- [1] J. R. Dunkley, *Finite Observation: A Spectral Theory of Nonequilibrium Visibility*, Preprint, 2026.
- [2] J. R. Dunkley, *Quotient Observation*, Preprint, 2026.
- [3] J. R. Dunkley, *Geometric Observation*, Preprint, 2026.
- [4] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. I. *Communications on Pure and Applied Mathematics*, 28(1):1–47, 1975.
- [5] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, second edition, 1998.
- [6] R. A. Fisher. Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society*, 22:700–725, 1925.
- [7] C. R. Rao. Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37:81–91, 1945.
- [8] T. J. Rothenberg. Identification in parametric models. *Econometrica*, 39(3):577–591, 1971.
- [9] S.-I. Amari and H. Nagaoka. *Methods of Information Geometry*. American Mathematical Society and Oxford University Press, 2000.
- [10] N. N. Čencov. *Statistical Decision Rules and Optimal Inference*. Translations of Mathematical Monographs, vol. 53, American Mathematical Society, 1982.
- [11] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

- [12] T. G. Kurtz. The relationship between stochastic and deterministic models for chemical reactions. *The Journal of Chemical Physics*, 57(7):2976–2978, 1972.
- [13] D. F. Anderson and T. G. Kurtz. Continuous time Markov chain models for chemical reaction networks. In H. Koepl, G. Setti, M. di Bernardo, and D. Densmore, editors, *Design and Analysis of Biomolecular Circuits*, pages 3–42. Springer, 2011.
- [14] D. T. Gillespie. The chemical Langevin equation. *The Journal of Chemical Physics*, 113(1):297–306, 2000.
- [15] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, 1986.
- [16] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland, third edition, 2007.
- [17] J. G. Kemeny and J. L. Snell. *Finite Markov Chains*. Van Nostrand, 1960.
- [18] P. Buchholz. Exact and ordinary lumpability in finite Markov chains. *Journal of Applied Probability*, 31(1):59–75, 1994.
- [19] J. C. Willems. Dissipative dynamical systems. Part I: General theory. *Archive for Rational Mechanics and Analysis*, 45:321–351, 1972.
- [20] V. A. Yakubovich. The solution of certain matrix inequalities in automatic control theory. *Doklady Akademii Nauk SSSR*, 143:1304–1307, 1962.
- [21] R. E. Kalman. Lyapunov functions for the problem of Lur’e in automatic control. *Proceedings of the National Academy of Sciences*, 49(2):201–205, 1963.
- [22] V. M. Popov. Absolute stability of nonlinear systems of automatic control. *Automation and Remote Control*, 22:857–875, 1962.
- [23] B. D. O. Anderson and S. Vongpanitlerd. *Network Analysis and Synthesis: A Modern Systems Theory Approach*. Prentice-Hall, 1973.
- [24] E. V. Haynsworth. Determination of the inertia of a partitioned Hermitian matrix. *Linear Algebra and its Applications*, 1(1):73–81, 1968.
- [25] D. E. Crabtree and E. V. Haynsworth. An identity for the Schur complement of a matrix. *Proceedings of the American Mathematical Society*, 22(2):364–366, 1969.
- [26] F. Zhang, editor. *The Schur Complement and Its Applications*. Springer, 2005.
- [27] W. F. Donoghue. *Monotone Matrix Functions and Analytic Continuation*. Springer, 1974.
- [28] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, third edition, 2003.
- [29] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3–4):321–377, 1936.
- [30] S. L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.
- [31] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky. Latent variable graphical model selection via convex optimization. *The Annals of Statistics*, 40(4):1935–1967, 2012.
- [32] T. Kailath. *Linear Systems*. Prentice-Hall, 1980.
- [33] B. C. Moore. Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Transactions on Automatic Control*, 26(1):17–32, 1981.
- [34] A. C. Antoulas. *Approximation of Large-Scale Dynamical Systems*. SIAM, 2005.
- [35] K. Glover. All optimal Hankel-norm approximations of linear multivariable systems and their L^∞ -error bounds. *International Journal of Control*, 39(6):1115–1193, 1984.

- [36] G. Kron. *Diakoptics: The Piecewise Solution of Large-Scale Systems*. Macdonald, 1963.
- [37] P. G. Doyle and J. L. Snell. *Random Walks and Electric Networks*. Mathematical Association of America, 1984.
- [38] P. Tetali. Random walks and the effective resistance of networks. *Journal of Theoretical Probability*, 4(1):101–109, 1991.
- [39] E. B. Curtis, D. Ingerman, and J. A. Morrow. Circular planar graphs and resistor networks. *Linear Algebra and its Applications*, 283(1–3):115–150, 1998.
- [40] A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. SIAM, 1994.
- [41] R. Lyons and Y. Peres. *Probability on Trees and Networks*. Cambridge University Press, 2017.
- [42] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.
- [43] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2:299–318, 1967.
- [44] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, second edition, 2006.
- [45] O. Kallenberg. *Foundations of Modern Probability*. Springer, third edition, 2021.
- [46] R. Penrose. A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(3):406–413, 1955.
- [47] A. Albert. Conditions for positive and nonnegative definiteness in terms of pseudoinverses. *SIAM Journal on Applied Mathematics*, 17(2):434–440, 1969.
- [48] A. Ben-Israel and T. N. E. Greville. *Generalized Inverses: Theory and Applications*. Springer, second edition, 2003.
- [49] G. Lindblad. Completely positive maps and entropy inequalities. *Communications in Mathematical Physics*, 40:147–151, 1975.
- [50] A. Uhlmann. Relative entropy and the Wigner–Yanase–Dyson–Lieb concavity in an interpolation theory. *Communications in Mathematical Physics*, 54:21–32, 1977.
- [51] D. Petz. Sufficient subalgebras and the relative entropy of states of a von Neumann algebra. *Communications in Mathematical Physics*, 105:123–131, 1986.
- [52] L. P. Kadanoff. Scaling laws for Ising models near T_c . *Physics Physique Fizika*, 2(6):263–272, 1966.
- [53] K. G. Wilson and J. Kogut. The renormalization group and the ϵ expansion. *Physics Reports*, 12(2):75–199, 1974.
- [54] J. Polchinski. Renormalization and effective Lagrangians. *Nuclear Physics B*, 231(2):269–295, 1984.
- [55] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, second edition, 2009.
- [56] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. MIT Press, second edition, 2000.
- [57] J. Tian and J. Pearl. A general identification condition for causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI-02)*, pages 567–573. AAAI Press, 2002.