Canonical Stein Geometry:
Burg Projections, Convolution, and Optimal Metrics

Tianle Liu*

4 September 2026

Abstract

Moment maps furnish a distinguished positive Stein kernel although the Stein identity is generally nonunique. Fixing the moment-map kernel τ_{mm} and $G = \tau_{\mathrm{mm}}^{-1}$, we prove that τ_{mm} uniquely maximizes relative expected log determinant over the closed positive Stein kernels satisfying the no-flux calibration. The deficit is an integrated Burg divergence and controls affine-relative trace distance. The kernel defines a reversible diffusion metric, and Burg uniquely selects the Ornstein–Uhlenbeck metric from a continuum of optimal Gaussian Poincaré metrics. Burg divergence also satisfies a two-gap conditional tower identity. For independent sums this becomes an exact three-gap convolution formula, yielding a Gaussian-rigid determinant monotone and quantitative stability under ellipticity. For covariance-normalized stochastic localization, the same two gaps control the signed error in an exact finite-time Burg–Parseval identity. The projection theorem extends directly to all one-dimensional log-concave laws, their products, and invertible affine images, while a separate varying-law compactness theorem records what survives without inverse-metric control.

1 Introduction

Let μ be a centered probability measure supported on an open set $\Omega \subset \mathbb{R}^n$. A symmetric matrix field τ is a local Stein kernel if

$$\int x \cdot \Phi(x)\ \mathrm{d}\mu(x) = \int \tau(x) : \nabla\Phi(x)\ \mathrm{d}\mu(x) \qquad \big(\Phi \in C_c^1(\Omega;\mathbb{R}^n)\big). \tag{1}$$

Its mean, when the usual cutoff is legitimate, is $\mathrm{Cov}(\mu)$. The identity is underdetermined in dimensions greater than one: divergence-free matrix fields can be added without changing it.

The moment-measure theorem of Cordero-Erausquin and Klartag [3], and Fathi's use of it for Stein kernels [11], single out a positive field τ_{mm}. If ψ is the target-side Legendre potential, then

$$\tau_{\mathrm{mm}} = (\nabla^2\psi)^{-1}.$$

Here "canonical" means the field singled out by the moment-measure construction. The source potential is unique up to translation, an ambiguity that does not change the target-side field, and Subsection 3.2 shows covariance under invertible affine maps. It does not mean that the bare Stein identity determines a unique kernel. Put $G = \tau_{\mathrm{mm}}^{-1}$. On the no-flux calibrated slice of the positive Stein cone, the following matrix Bregman identity holds [2, 8]:

$$-\int \log\det\Big(G^{1/2}\tau G^{1/2}\Big)\ \mathrm{d}\mu = \int d_{\mathrm{LD}}\Big(G^{1/2}\tau G^{1/2}, \mathrm{I}\Big)\ \mathrm{d}\mu, \tag{2}$$

*Department of Statistics and Operations Research, UNC Chapel Hill. Email: tianletl@unc.edu.

where

$$d_{\mathrm{LD}}(Z, \mathrm{I}) = \operatorname{Tr} Z - \log\det Z - n.$$

The right side is nonnegative and vanishes only at $\tau = \tau_{\mathrm{mm}}$. Unlike an absolute log-determinant formulation, (2) remains meaningful when τ is singular or $\log\det \tau_{\mathrm{mm}}$ is not integrable.

The theorem has two stability consequences. At a fixed law, the admissible positive slice is closed in the affine-relative metric

$$d_G(\tau, \sigma) = \int \left\| G^{1/2}(\tau - \sigma) G^{1/2} \right\|_{S_1} \mathrm{d}\mu, \tag{3}$$

and

$$d_G(\tau, \tau_{\mathrm{mm}}) \le 2\sqrt{n\, \Delta_{\mathrm{B}}(\tau \mid \tau_{\mathrm{mm}})}. \tag{4}$$

There is also a target-side nonsmooth result. In one dimension the canonical kernel of an arbitrary centered nondegenerate log-concave law is the intrinsic tail-moment quotient

$$\kappa(x) = \frac{\int_x^b u\rho(u)\,\mathrm{d}u}{\rho(x)}.$$

Tensorization and affine covariance therefore give an approximation-independent canonical kernel for every product and invertible affine image of products, including Laplace products and uniform measures on boxes and parallelotopes. The closed Burg theorem and its quantitative rigidity hold for this class; the one-dimensional formula and its weighted Poincaré role are discussed in [20, 9].

At a fixed law, every sufficiently regular positive Stein kernel is a reversible diffusion coefficient. Fathi's kernel gives a constant-one weighted Poincaré metric, but that metric optimization need not be unique. We exhibit a continuum of normalized optimal metrics for the standard Gaussian and show that the Burg objective selects the Ornstein–Uhlenbeck metric.

The conditional form of (2) is a two-gap tower identity. We apply it to convolution. If X_i are independent, $\sum_i a_i^2 = 1$, and $Z = \sum_i a_i X_i$, then, under the global Stein boundary condition stated in Section 6, the conditional sum of their canonical kernels is a Stein kernel of Z. When that kernel is also calibrated for the output moment map, inserting it between the canonical output kernel and the input kernels gives three nonnegative gaps: a canonical projection gap, a conditioning gap, and a pointwise matrix mixture gap. In particular, with

$$\mathscr{V}(\mu) = \int \log\det \tau_{\mathrm{mm}}^{\mu}\,\mathrm{d}\mu,$$

we obtain

$$\mathscr{V}\left(\mathcal{L}\left(\sum_i a_i X_i\right)\right) \ge \sum_i a_i^2 \mathscr{V}(\mathcal{L}(X_i)). \tag{5}$$

For common covariance, the determinant deficit contracts, with equality only for Gaussian inputs. The three terms also give the stability estimate below. This determinant monotone is distinct from the convolution bounds and central-limit monotonicity for the scalar Poincaré constant in [5].

The continuous counterpart uses covariance-normalized stochastic localization. At each finite time, the scaled posterior canonical kernel is averaged back to the initial law, producing the same canonical and fiber gaps as in the tower. A posterior-score calculation then joins their sum to a Parseval decomposition; under the stated envelope hypotheses, the two Burg gaps control the signed error in the initial canonical metric.

Relation to prior work

The classical inverse-Hessian weighted Poincaré inequality goes back to Brascamp and Lieb [1]. Cordero-Erausquin and Klartag later established existence and uniqueness for moment measures [3], while Santambrogio gave an optimal-transport variational treatment [19]. Klartag developed the associated moment-map geometry and Poincaré inequalities [14]; related Hessian metrics in optimal transport were studied by Kolesnikov [15]. Fathi's observation that the target-side inverse Hessian is a Stein kernel, together with his weighted Poincaré inequality, is the starting point here [11]. The present question is different: with that kernel fixed as reference, which variational principle selects it from the other positive Stein kernels?

Stein kernels also enter normal approximation, transport, entropy, and functional inequalities. Ledoux, Nourdin, and Peccati related Stein discrepancy to logarithmic Sobolev and transport inequalities [16]. Courtade, Fathi, and Pananjady proved existence under a spectral gap and selected a unique Jacobian-form kernel by a quadratic Sobolev minimization [4]. Mijoule, Raič, Reinert, and Swan developed the multivariate density framework and recorded explicit nonunique Gaussian kernels [18]. In one dimension, Saumard obtained the intrinsic kernel and its weighted inequalities [20], and Döbler characterized existence and uniqueness for general laws [9]. These results explain both sides of the selection problem: uniqueness is intrinsic in one dimension, whereas a separate criterion is needed in higher dimensions.

The closest stability and metric comparisons address different variables. Fathi and Mikulincer vary the diffusion coefficient and its invariant measure [12]; Delalande and Farinelli study a regularized moment measure problem and stability of its minimizers [7]. Lelièvre, Pavliotis, Robin, Santet, and Stoltz optimize reversible diffusion coefficients through the spectral gap and sampling performance [17]. Cui, Tong, and Zahm instead minimize the weighted Poincaré constant under a trace budget and identify normalized optimizers as Stein kernels [6]. Matrix LogDet divergence belongs to the established theory of Bregman matrix approximation [8]. Here it is used on a calibrated Stein slice, where the moment-map kernel becomes the unique Burg maximizer and the deficit is an exact projection gap.

Finally, stochastic localization was introduced by Eldan as a posterior martingale method for log-concave measures [10]. Its adaptive Gaussian-channel interpretation is related to the feedback-information identity of Kadota, Zakai, and Ziv [13]. In this paper localization serves as the continuous conditional family on which the static Burg projection can be iterated. The same chain rule governs independent sums and posterior localization; the additional ingredients are, respectively, the pointwise matrix-mixture gap and the score–residual Parseval calculation.

The calibration condition is essential. Both the feasible class and the relative objective use the given field $G = \tau_{\mathrm{mm}}^{-1}$. Thus the main theorem does not recover τ_{mm} from the compact-test Stein class alone; Remark 2.4 shows that such a statement is false. Membership of convolution, localization, or optimal-metric kernels in the calibrated class must be checked separately.

For arbitrary varying laws, a uniform L^2 bound still gives only subsequential positive Stein kernels by a joint weak-limit and conditional-barycenter construction; without control of inverse metrics it does not identify different subsequential limits.

The paper is organized as follows. Section 2 introduces the moment-map kernel, the calibrated class, and the nonsmooth affine-product construction. Section 3 establishes its variational, quantitative, affine, and differential characterizations. Section 4 gives the fixed-law Dirichlet, diffusion, and optimal-Poincaré interpretations. Section 5 proves the conditional Burg chain rule. Sections 6 and 7 develop its discrete convolution and continuous stochastic-localization consequences. Section 8 closes with the varying-law compactness theorem.

2 Moment maps and the calibrated Stein class

All matrices are real and symmetric unless stated otherwise. The positive definite and semidefinite cones are denoted by $\mathbb{S}^n_{++}$ and $\mathbb{S}^n_+$. The Schatten trace norm is $\|\cdot\|_{S_1}$.

Definition 2.1 (Regular moment representation)**.** Let $\Omega, Y \subset \mathbb{R}^n$ be open convex sets. A centered law $\mu = \rho(x)\,\mathrm{d}x$ on Ω has a *regular moment representation* if there is a strictly convex C^2 Legendre function $\varphi : Y \to \mathbb{R}$ such that

$$\nu(\,\mathrm{d}y) = Z^{-1}e^{-\varphi(y)}\,\mathrm{d}y, \qquad \mu = (\nabla\varphi)_{\#}\nu,$$

and $\nabla\varphi : Y \to \Omega$ is a C^1 diffeomorphism. With $\psi = \varphi^*$, define

$$\tau_\star(x) = \tau_{\mathrm{mm}}(x) = \nabla^2\varphi(\nabla\psi(x)) = (\nabla^2\psi(x))^{-1}, \qquad G = \tau_\star^{-1}. \tag{6}$$

Fathi's integration by parts gives, first for smooth compactly supported tests and then by approximation,

$$\int x \cdot \Phi(x)\,\mathrm{d}\mu(x) = \int \tau_\star(x) : \nabla\Phi(x)\,\mathrm{d}\mu(x). \tag{7}$$

Thus $\tau_\star$ is a continuous positive-definite Stein kernel.

The Burg calculation uses one global pairing that compactly supported Stein tests do not automatically provide. We state it as an explicit calibration condition.

Definition 2.2 (No-flux calibrated positive Stein class)**.** For the regular moment representation above, let $\mathcal{C}_G(\mu)$ be the set of measurable fields $\tau : \Omega \to \mathbb{S}^n_+$ such that

(i) τ satisfies (1);

(ii) $\int \mathrm{Tr}(G\tau)\,\mathrm{d}\mu = n$.

For $\tau, \sigma \in \mathcal{C}_G(\mu)$ define d_G by (3).

Proposition 2.3 (Canonical kernel for nonsmooth affine products)**.** *Let μ_i be a centered, nondegenerate one-dimensional log-concave law. Write $I_i = (a_i, b_i)$ for the interior of its support and ρ_i for the continuous positive representative of its density on I_i. Define*

$$F_i(x) = \int_x^{b_i} u\rho_i(u)\,\mathrm{d}u = -\int_{a_i}^x u\rho_i(u)\,\mathrm{d}u, \qquad \kappa_i(x) = \frac{F_i(x)}{\rho_i(x)}. \tag{8}$$

Then κ_i is continuous and positive on I_i, and it is both the one-dimensional Stein kernel and the target-side moment-map kernel of μ_i. Consequently, if A is invertible and $\mu = A_{\#}(\mu_1 \otimes \cdots \otimes \mu_n)$, then

$$\tau_\star(Ax) = A\,\mathrm{diag}\big(\kappa_1(x_1), \ldots, \kappa_n(x_n)\big)A^\mathsf{T} \tag{9}$$

is its continuous positive-definite moment-map kernel. The construction is intrinsic to the factors, and $\log\det\tau_\star \in L^1(\mu)$.

Proof. Centering gives the two expressions for F_i. They show that $F_i > 0$ on I_i, while $F_i' = -x\rho_i$ gives

$$\int_{I_i} \kappa_i f'\,\mathrm{d}\mu_i = \int_{I_i} F_i f'\,\mathrm{d}x = \int_{I_i} x f(x)\rho_i(x)\,\mathrm{d}x \qquad (f \in C_c^1(I_i)).$$

Thus κ_i is a Stein kernel.

Fix one factor and suppress the index. Let

$$y(x)=\int_{x_0}^{x}\frac{\mathrm{d}u}{\kappa(u)},\qquad Y=y(I),$$

and write $x=x(y)$ for its inverse. Define φ up to an additive constant by $\varphi'(y)=x(y)$. Then

$$\varphi''(y)=\frac{\mathrm{d}x}{\mathrm{d}y}=\kappa(x(y))>0.$$

Moreover,

$$\frac{\mathrm{d}}{\mathrm{d}x}\varphi(y(x))=\frac{x}{\kappa(x)}=-\frac{\mathrm{d}}{\mathrm{d}x}\log F(x),$$

so $\varphi(y(x))=-\log F(x)+C$. The function F tends to zero at both endpoints of I. Thus φ diverges at every finite endpoint of Y; if such an endpoint corresponds to an infinite endpoint of I, then $|\varphi'|=|x|\to\infty$, while a finite endpoint of I would make φ' bounded and force that endpoint of Y to be infinite. Hence φ is Legendre and essentially continuous, and its gradient is a C^1 diffeomorphism from Y to I.

Under $x=\varphi'(y)$ the pullback of μ has density

$$q(y)=\rho(x)\frac{\mathrm{d}x}{\mathrm{d}y}=\rho(x)\kappa(x)=F(x).$$

Since

$$\frac{\mathrm{d}}{\mathrm{d}y}\log q(y)=\frac{F'(x)}{F(x)}\frac{\mathrm{d}x}{\mathrm{d}y}=-x=-\varphi'(y),$$

one has $q=Z^{-1}e^{-\varphi}$. Hence μ is the moment measure of φ and $\kappa=\varphi''\circ(\varphi')^{-1}$. Moment-measure uniqueness applies; translating the source potential does not change this target-side field.

Summing the one-dimensional potentials tensorizes the construction, and an invertible linear change gives (9). Moment-measure uniqueness again shows that this is the canonical target-side field, independently of the chosen product presentation. The formula depends only on the factor densities, so it has no regularization or subsequence choice. Finally,

$$\log\kappa(X)=\log q(Y)-\log\rho(X).$$

Both q and ρ are nondegenerate one-dimensional log-concave densities; their absolute log densities are integrable against their own laws. Summing over the factors and adding $2\log|\det A|$ proves the final assertion. $\square$

The metric is finite because positivity and calibration give

$$\left\|G^{1/2}(\tau-\sigma)G^{1/2}\right\|_{S_1}\le\mathrm{Tr}(G\tau)+\mathrm{Tr}(G\sigma).$$

Condition (ii) is automatic in the standard regular class whenever the Stein identity for τ and $\tau_\star$ can be tested, by cutoff, against $\nabla\psi$. Indeed, with $D=\tau-\tau_\star$,

$$\int\mathrm{Tr}(GD)\,\mathrm{d}\mu=\int D:\nabla^2\psi\,\mathrm{d}\mu=0. \tag{10}$$

A sufficient concrete condition is an exhaustion $\chi_R \in C_c^\infty(\Omega)$, $0 \le \chi_R \le 1$, $\chi_R \to 1$, for which the bulk terms are uniformly integrable and

$$\int \rho D_{ij}(\partial_j \chi_R)(\partial_i \psi)\,\mathrm{d}x \longrightarrow 0.$$

Thus $\mathcal{C}_G(\mu)$ contains the d_G-closure of the regular cutoff class, but its definition also admits calibrated weak kernels not known to arise by that approximation.

Calibration is a boundary condition, not a consequence of the compact-test identity in complete generality. In weak differential form the difference $D = \tau - \tau_\star$ obeys $\operatorname{div}(\rho D) = 0$ row by row. When a normal trace is available, the cutoff condition above is precisely the vanishing of its pairing with $\nabla\psi$ at the boundary. This motivates the term "no-flux calibrated."

Remark 2.4 (Why the calibration cannot be dropped). The obstruction already appears in one dimension. Take

$$Y = (-\pi/2, \pi/2), \qquad \varphi(y) = -2\log\cos y, \qquad x = \varphi'(y) = 2\tan y.$$

The source density is proportional to $\cos^2 y$, the centered target has the smooth density

$$\rho(x) = \frac{1}{\pi}\left(1 + \frac{x^2}{4}\right)^{-2}$$

on $\mathbb{R}$, and

$$\tau_\star(x) = \varphi''(y) = 2 + \frac{x^2}{2}.$$

For every $c > 0$,

$$\tau_c(x) = \tau_\star(x) + \frac{c}{\rho(x)} \tag{11}$$

is positive. Here

$$\rho(x)\tau_\star(x) = \frac{2}{\pi}\left(1 + \frac{x^2}{4}\right)^{-1}, \qquad (\rho\tau_\star)'(x) = -x\rho(x).$$

Consequently, for every $f \in C_c^1(\mathbb{R})$,

$$\begin{aligned}\int_{\mathbb{R}} \tau_c f'\,\mathrm{d}\mu &= \int_{\mathbb{R}} (\rho\tau_\star + c) f'\,\mathrm{d}x \\ &= -\int_{\mathbb{R}} (\rho\tau_\star)' f\,\mathrm{d}x + c\int_{\mathbb{R}} f'\,\mathrm{d}x = \int_{\mathbb{R}} x f\,\mathrm{d}\mu.\end{aligned}$$

The boundary terms vanish because f is compactly supported. Thus τ_c satisfies the bare compact-test Stein identity. Nevertheless, with $G = \tau_\star^{-1}$,

$$\int G\tau_c\,\mathrm{d}\mu = 1 + c\int_{\mathbb{R}} G(x)\,\mathrm{d}x = 1 + c\,|Y| = 1 + c\pi.$$

Thus $\tau_c \notin \mathcal{C}_G(\mu)$ and, pointwise, $\log\tau_c > \log\tau_\star$. Both logarithms are integrable because $\rho(x) = O(|x|^{-4})$ and $\tau_c(x) = O(|x|^4)$. Maximum log determinant is therefore false over the bare compact-test class. A zero-flux, cutoff, or equivalent calibration hypothesis is indispensable.

Proposition 2.5 (Closedness). *The calibrated positive Stein class is closed under d_G. More precisely, if $\tau_k \in \mathcal{C}_G(\mu)$ and $d_G(\tau_k, \tau) \to 0$, then, after choosing a measurable representative, $\tau \in \mathcal{C}_G(\mu)$.*

Proof. Set $Z_k = G^{1/2}\tau_k G^{1/2}$ and $Z = G^{1/2}\tau G^{1/2}$. Trace-norm L^1 convergence has an almost everywhere convergent subsequence. Since $\mathbb{S}^n_+$ is closed, Z and τ are positive semidefinite. Also

$$\left|\int \operatorname{Tr}(Z_k - Z)\,\mathrm{d}\mu\right| \le \int \|Z_k - Z\|_{S_1}\,\mathrm{d}\mu \longrightarrow 0,$$

so calibration persists.

For $\Phi \in C^1_c(\Omega;\mathbb{R}^n)$ with support K, continuity and positive definiteness of $\tau_\star$ imply that the field

$$G^{-1/2}\nabla\Phi G^{-1/2}$$

is bounded on K. Trace duality gives

$$\left|\int (\tau_k - \tau) : \nabla\Phi\,\mathrm{d}\mu\right| \le \sup_K \left\|G^{-1/2}\nabla\Phi G^{-1/2}\right\|_{\mathrm{op}} d_G(\tau_k, \tau).$$

Passing to the limit in the Stein identity completes the proof. □

3 Canonical selection and rigidity

3.1 The closed Burg projection and quantitative rigidity

For $Z \in \mathbb{S}^n_+$, set

$$d_{\mathrm{LD}}(Z, \mathrm{I}) = \operatorname{Tr} Z - \log\det Z - n \in [0, \infty], \tag{12}$$

with $\log\det Z = -\infty$ when Z is singular. For $\tau \in \mathcal{C}_G(\mu)$ define the extended relative log-volume

$$\mathcal{L}_{\tau_\star}(\tau) = \int \log\det(G^{1/2}\tau G^{1/2})\,\mathrm{d}\mu \in [-\infty, \infty). \tag{13}$$

The integral is unambiguous: its positive part is bounded by $\operatorname{Tr}(G\tau)$ and hence is integrable.

Theorem 3.1 (Closed Burg projection)**.** *For every $\tau \in \mathcal{C}_G(\mu)$,*

$$\boxed{-\mathcal{L}_{\tau_\star}(\tau) = \Delta_{\mathrm{B}}(\tau \mid \tau_\star) := \int d_{\mathrm{LD}}(G^{1/2}\tau G^{1/2}, \mathrm{I})\,\mathrm{d}\mu.} \tag{14}$$

In particular, $\mathcal{L}_{\tau_\star} \le 0$, with equality only at $\tau = \tau_\star$ almost everywhere. The deficit is lower semicontinuous, and the relative log-volume is upper semicontinuous, in d_G.

Proof. Put $Z = G^{1/2}\tau G^{1/2}$, $g = \log\det Z$, and $a = \operatorname{Tr} Z - n$. Calibration gives $a \in L^1(\mu)$ and $\int a\,\mathrm{d}\mu = 0$, while $g^+ \le \operatorname{Tr} Z$ gives $g^+ \in L^1(\mu)$. Pointwise, with the extended convention at singular matrices,

$$d_{\mathrm{LD}}(Z, \mathrm{I}) = a - g = (a - g^+) + g^- \ge 0.$$

The first summand on the right is integrable. The addition rule for an integrable function and a nonnegative extended function therefore gives, in $[0, \infty]$,

$$\begin{aligned}\int d_{\mathrm{LD}}(Z, \mathrm{I})\,\mathrm{d}\mu &= \int (a - g^+)\,\mathrm{d}\mu + \int g^-\,\mathrm{d}\mu \\ &= -\int g^+\,\mathrm{d}\mu + \int g^-\,\mathrm{d}\mu = -\int g\,\mathrm{d}\mu.\end{aligned}$$

This proves (14) also when Z is singular on a set of positive measure or $\int g \,\mathrm{d}\mu = -\infty$. Since $x - \log x - 1$ is nonnegative and vanishes only at $x = 1$, equality forces $Z = \mathrm{I}$ almost everywhere and the relative log-volume lies in $[-\infty, 0]$.

If $Z_k \to Z$ in trace-norm L^1, first choose a subsequence realizing the liminf of the deficits and then a further almost everywhere convergent subsequence. The extended function $Z \mapsto d_{\mathrm{LD}}(Z, \mathrm{I})$ is lower semicontinuous on $\mathbb{S}^n_+$. Fatou's lemma proves lower semicontinuity of the deficit. The identity just proved then gives

$$\limsup_k \mathcal{L}_{\tau_\star}(\tau_k) = -\liminf_k \Delta_{\mathrm{B}}(\tau_k \mid \tau_\star) \le \mathcal{L}_{\tau_\star}(\tau),$$

which is the asserted upper semicontinuity. □

If $\log\det \tau_\star \in L^1(\mu)$, then the relative formula becomes the absolute identity

$$\int \log\det \tau_\star \,\mathrm{d}\mu - \int \log\det \tau \,\mathrm{d}\mu = \Delta_{\mathrm{B}}(\tau \mid \tau_\star), \tag{15}$$

where the second integral may be $-\infty$. Indeed, with $g = \log\det(G^{1/2}\tau G^{1/2})$, the determinant product formula gives

$$\log\det\tau = \log\det\tau_\star + g$$

pointwise in the extended sense, including at singular τ. Since $\log\det\tau_\star \in L^1$ and $g^+ \in L^1$, the positive part of $\log\det\tau$ is integrable and addition of the L^1 term commutes with the extended integral. Thus

$$\int \log\det\tau \,\mathrm{d}\mu = \int \log\det\tau_\star \,\mathrm{d}\mu + \int g \,\mathrm{d}\mu,$$

with either side allowed to equal $-\infty$. Combining this equality with (14) proves (15) in the non-finite branch as well.

Theorem 3.2 (Global quantitative rigidity)**.** *Let $\tau \in \mathcal{C}_G(\mu)$ and $Z = G^{1/2}\tau G^{1/2}$. Then*

$$\Delta_{\mathrm{B}}(\tau \mid \tau_\star) \ge \int \left\| Z^{1/2} - \mathrm{I} \right\|_{\mathrm{HS}}^2 \mathrm{d}\mu, \tag{16}$$

$$d_G(\tau, \tau_\star) \le 2\sqrt{n\,\Delta_{\mathrm{B}}(\tau \mid \tau_\star)}. \tag{17}$$

Consequently every maximizing sequence converges to $\tau_\star$ in d_G.

Proof. For $x > 0$,

$$x - \log x - 1 - (\sqrt{x} - 1)^2 = 2(\sqrt{x} - 1 - \log\sqrt{x}) \ge 0.$$

At $x = 0$ the same inequality holds by the extended convention and lower semicontinuity. Functional calculus and summation over eigenvalues give (16). Next, $Z - \mathrm{I} = (Z^{1/2} - \mathrm{I})(Z^{1/2} + \mathrm{I})$, so Schatten Hölder and Cauchy–Schwarz give

$$\begin{aligned}\int \|Z - \mathrm{I}\|_{S_1} \,\mathrm{d}\mu \le {} & \left(\int \left\| Z^{1/2} - \mathrm{I} \right\|_{\mathrm{HS}}^2 \mathrm{d}\mu \right)^{1/2} \\ & \times \left(\int \left\| Z^{1/2} + \mathrm{I} \right\|_{\mathrm{HS}}^2 \mathrm{d}\mu \right)^{1/2}.\end{aligned}$$

The second squared factor is at most $4n$: calibration gives $\int \operatorname{Tr} Z = n$, while $2\operatorname{Tr} Z^{1/2} \le \operatorname{Tr} Z + n$. This proves (17). □

Corollary 3.3 (Stability of the canonical Dirichlet form)**.** *If a smooth f satisfies $\left\|\tau_\star^{1/2}\nabla f\right\|^2 \le K$ almost everywhere, then*

$$\left|\int \nabla f^\mathsf{T}(\tau - \tau_\star)\nabla f \,\mathrm{d}\mu\right| \le 2K\sqrt{n\,\Delta_\mathrm{B}(\tau \mid \tau_\star)}. \tag{18}$$

Proof. With $u = \tau_\star^{1/2}\nabla f$, the integrand is $u^\mathsf{T}(Z-\mathrm{I})u$. Bound it by $K\,\|Z-\mathrm{I}\|_{S_1}$ and use (17). □

Corollary 3.4 (Nonsmooth affine-product Burg theorem)**.** *For every law in Proposition 2.3, the class $\mathcal{C}_G(\mu)$ is closed in d_G, the kernel (9) is the unique maximizer of relative expected log determinant on that class, and (16)– (18) hold. Singular positive-semidefinite competitors are included. Since* $\log\det\tau_\star \in L^1(\mu)$*, the absolute identity (15) holds as well.*

Proof. Proposition 2.3 supplies a C^2 moment representation with a continuous positive-definite reference field. These are exactly the reference-field properties used in Proposition 2.5, Theorem 3.1, and Theorem 3.2. □

There is no reverse matrix estimate from trace calibration alone. In dimension one, let $Z_k = e^{-k}$ on an event of probability $1/k$ and let $Z_k = c_k$ otherwise, with $\mathbb{E}Z_k = 1$. Then $\mathbb{E}\,|Z_k - 1| \to 0$ but $\mathbb{E}d_{\mathrm{LD}}(Z_k, 1) = -\mathbb{E}\log Z_k \to 1$. Thus d_G convergence need not force convergence of the corresponding matrix Burg values in the ambient trace-calibrated matrix space. This example is not asserted to arise from Stein kernels in $\mathcal{C}_G(\mu)$ at one fixed law. A reverse statement inside that class would require additional structure; lower spectral control or logarithmic uniform integrability gives a sufficient condition.

3.2 Affine covariance

For an invertible affine change $y = A(x - \mathbb{E}X)$, kernels transform as $\tau^A(y) = A\tau(x)A^\mathsf{T}$. The canonical kernel obeys the same rule. (A further translation merely recenters the Stein identity.) If $G^A = (A\tau_\star A^\mathsf{T})^{-1}$, then

$$U = (G^A)^{1/2}AG^{-1/2}$$

is orthogonal and

$$(G^A)^{1/2}\tau^A(G^A)^{1/2} = U(G^{1/2}\tau G^{1/2})U^\mathsf{T}.$$

Hence $\mathcal{C}_G$ transforms covariantly, while d_G, the relative volume, and the Burg deficit are invariant.

3.3 Inverse-Codazzi rigidity

The Burg theorem is variational. A complementary differential characterization explains why the inverse field is the natural geometry. For a C^1 matrix field H, write

$$(\operatorname{Curl} H)_{i;jk} = \partial_i H_{jk} - \partial_j H_{ik}.$$

Theorem 3.5 (Inverse-Codazzi characterization)**.** *Let $\Omega \subset \mathbb{R}^n$ be open, convex, and connected, and let $\mu = \rho\,\mathrm{d}x$ be centered, with $\rho \in C^1(\Omega)$ strictly positive. Suppose that $\tau : \Omega \to \mathbb{S}^n_{++}$ is a C^1 Stein kernel and $H = \tau^{-1}$ satisfies*

$$\partial_i H_{jk} = \partial_j H_{ik}. \tag{19}$$

Assume that the potential ψ constructed in the proof is Legendre and that $\nabla\psi : \Omega \to Y$ is a global C^2 diffeomorphism onto an open convex set Y. Assume also that $\varphi = \psi^*$ is essentially *continuous,* $0 < \int_Y e^{-\varphi}\,\mathrm{d}y < \infty$, *and that the hypotheses of moment-measure uniqueness in [3] hold. Then* $\tau = \tau_{\mathrm{mm}}$. *Conversely, every regular moment-map kernel whose dual potential* ψ *is* C^3 *has* $\mathrm{Curl}(\tau_{\mathrm{mm}}^{-1}) = 0$.

Proof. For each k, (19) says that the one-form $H_{ik}\,\mathrm{d}x_i$ is closed. The Poincaré lemma gives a vector field u with $H_{ik} = \partial_i u_k$. Symmetry makes $u_k\,\mathrm{d}x_k$ closed, hence $u = \nabla\psi$ and

$$H = \nabla^2\psi. \tag{20}$$

Positive definiteness makes ψ strictly convex. Under the global hypotheses, $\varphi = \psi^*$ is defined on an open convex set Y, and

$$x = \nabla\varphi(y), \qquad \nabla^2\varphi(y) = \tau(x).$$

Pull μ back under $\nabla\varphi$ and write

$$\widetilde{\rho}(y) = \rho(\nabla\varphi(y))\det\nabla^2\varphi(y).$$

Hessian symmetry gives

$$\partial_{y_i}\log\det\nabla^2\varphi(y) = \partial_{x_j}\tau_{ij}(x). \tag{21}$$

The weak Stein equation is $\partial_j(\rho\tau_{ij}) = -x_i\rho$. Combining it with (21) yields

$$\partial_{y_i}\log\widetilde{\rho}(y) = -x_i = -\partial_i\varphi(y).$$

Thus $\widetilde{\rho} = Z^{-1}e^{-\varphi}$, so μ is the moment measure of φ. Moment-measure uniqueness [3] identifies the representation, and $\tau = \tau_{\mathrm{mm}}$. Conversely, $\tau_{\mathrm{mm}}^{-1} = \nabla^2\psi$, whose third derivatives have the required symmetry. □

The conclusion uses the global Legendre assumptions. A locally inverse-Hessian field need not produce a global moment representation, and conditional averaging need not preserve the inverse-Codazzi condition. The calibrated class is convex and closed under deterministic convex combinations. Thus inverse-Codazzi rigidity gives a differential characterization, while the Burg characterization remains stable under averaging once calibration has been verified.

4 Stein metrics, reversible diffusions, and Poincaré optimality

Before introducing conditional operations, we record two fixed-law meanings of the Burg principle.

4.1 Reversible Stein diffusions

A sufficiently regular positive-definite Stein kernel defines a symmetric Dirichlet form. Set

$$L_\tau f = \tau : \nabla^2 f - x\cdot\nabla f, \qquad \mathcal{E}_\tau(f,g) = \int \nabla f^{\mathsf{T}}\tau\nabla g\,\mathrm{d}\mu. \tag{22}$$

Proposition 4.1 (Stein kernels as reversible metrics)**.** *Let* τ *be a symmetric positive-definite local Stein kernel whose entries are locally integrable. For compactly supported smooth* f, g,

$$\int f L_\tau g\,\mathrm{d}\mu = -\mathcal{E}_\tau(f,g). \tag{23}$$

Thus L_τ *is symmetric and* μ *is invariant at the Dirichlet-form level. If the Stein identity extends by cutoff to linear tests and* $\tau \in L^1(\mu)$, *then* $\int\tau\,\mathrm{d}\mu = \mathrm{Cov}(\mu)$. *If, separately, the form is closable* *and conservative—or classical local regularity gives a well-posed nonexplosive diffusion—then its generator is realized by*

$$\mathrm{d}X_t = -X_t\,\mathrm{d}t + \sqrt{2\tau(X_t)}\,\mathrm{d}B_t.$$

Provided its global mean-kernel identity is justified, the moment-map metric additionally satisfies

$$\operatorname{Var}_\mu f \le \mathcal{E}_{\tau_\star}(f,f), \tag{24}$$

and the best constant is one because affine functions give equality.

Proof. Apply the ith Stein identity to $f\,\partial_i g$ and sum over i. The Hessian terms cancel from $\int f L_\tau g$, leaving $-\int \nabla f^\mathsf{T} \tau \nabla g$. Under the stated cutoff, applying the identity to coordinate-linear vector tests gives the mean formula. Inequality (24) is Fathi's weighted Poincaré theorem [11]; equality on affine functions follows from the mean identity. □

For an Euler step with conditional covariance $2h\tau(x)$,

$$h(X_{k+1} \mid X_k = x) = \frac{n}{2}\log(4\pi e h) + \frac{1}{2}\log\det\tau(x). \tag{25}$$

Thus, when the absolute log determinants are integrable, the Burg maximizer also maximizes mean one-step noise entropy among calibrated kernels for which the corresponding diffusion is well posed. In the extended setting, the statement concerns the relative volume (13). Corollary 3.3 controls the Dirichlet form for functions satisfying its bounded-gradient hypothesis; it gives no semigroup or pathwise estimate.

4.2 Optimal Poincaré metrics

For a positive matrix field W, let

$$C(\mu, W) = \sup_{f\not\equiv\mathrm{const}} \frac{\operatorname{Var}_\mu f}{\int \nabla f^\mathsf{T} W \nabla f\,\mathrm{d}\mu}, \tag{26}$$

with the usual form-domain convention. The normalized optimal-metric problem of Cui–Tong–Zahm [6] minimizes this constant under

$$\int \operatorname{Tr} W\,\mathrm{d}\mu = \operatorname{Tr}\operatorname{Cov}(\mu). \tag{27}$$

The trace budget and affine test functions show that the infimum is at least one. Fathi's inequality (24) shows that the moment-map metric attains one. Cui–Tong–Zahm prove, under their stated regularity, that every metric attaining the normalized value one is a positive-semidefinite Stein kernel. Their optimization alone does not assert uniqueness.

Corollary 4.2 (Secondary optimization among Poincaré minimizers)**.** *Among normalized Cui–Tong–Zahm optimizers that additionally belong to $\mathcal{C}_G(\mu)$, the moment-map metric is the unique maximizer of relative expected log determinant. In other words, it uniquely solves the second stage of the restricted lexicographic problem*

first minimize $C(\mu, W)$, *then, among calibrated minimizers, maximize $\mathcal{L}_{\tau_\star}(W)$.*

The intersection in the statement is essential: the trace normalization (27) is not the same as the moment-Hessian calibration $\int \mathrm{Tr}(GW) = n$ unless G is constant. Theorem 3.1 applies once that calibration is verified, for example by the cutoff criterion in Section 2.

For the standard Gaussian, the normalized Poincaré problem is nonunique. Let γ be standard Gaussian measure on $\mathbb{R}^n$, $n \geq 2$, and for $\alpha \geq 0$ define

$$W_\alpha(x) = \frac{1}{1+\alpha}\mathrm{I} + \frac{\alpha}{(1+\alpha)(n-1)}\left(|x|^2\,\mathrm{I} - xx^{\mathsf{T}}\right). \tag{28}$$

This is a subfamily of the explicit Gaussian Stein kernels in [18, Example 4.18].

Theorem 4.3 (A continuum of optimal Gaussian metrics)**.** *For every $0 \leq \alpha \leq 1$, W_α is positive definite, $\int W_\alpha \, \mathrm{d}\gamma = \mathrm{I}$, and*

$$\mathrm{Var}_\gamma f \leq \int \nabla f^{\mathsf{T}} W_\alpha \nabla f \, \mathrm{d}\gamma. \tag{29}$$

The best constant is exactly one, so the family consists of distinct normalized optimal metrics. For $\alpha > 1$, its weighted Poincaré constant is strictly larger than one.

The canonical Gaussian kernel is $W_0 = \mathrm{I}$. For every $\alpha > 0$,

$$\int \log\det W_\alpha \, \mathrm{d}\gamma < 0 = \int \log\det W_0 \, \mathrm{d}\gamma. \tag{30}$$

Thus the Burg criterion uniquely selects the Ornstein–Uhlenbeck metric among these normalized Poincaré minimizers.

Proof. The radial eigenvalue of W_α is $1/(1+\alpha)$, and its tangential eigenvalue is

$$\frac{n-1+\alpha\,|x|^2}{(1+\alpha)(n-1)},$$

so the field is positive definite. Direct differentiation gives

$$\mathrm{div}\, W_\alpha - W_\alpha x = -x, \tag{31}$$

and Gaussian integration gives $\int W_\alpha \, \mathrm{d}\gamma = \mathrm{I}$. Thus it is a normalized Gaussian Stein kernel.

For $i < j$, write $L_{ij} = x_i \partial_j - x_j \partial_i$. Pointwise,

$$\sum_{i<j} (L_{ij} f)^2 = |x|^2 \, |\nabla f|^2 - (x \cdot \nabla f)^2. \tag{32}$$

We use the strengthened Gaussian inequality

$$\mathrm{Var}_\gamma f \leq \frac{1}{2}\int |\nabla f|^2 \, \mathrm{d}\gamma + \frac{1}{2(n-1)} \sum_{i<j} \int (L_{ij} f)^2 \, \mathrm{d}\gamma. \tag{33}$$

To prove it, decompose a centered f into Hermite chaoses. On the first chaos, the two terms on the right add exactly to the variance. On every chaos of order at least two, the gradient term with coefficient $1/2$ already dominates the variance. The angular operators preserve chaos, so orthogonality and density complete the proof.

Equations (28) and (32) give

$$\int \nabla f^{\mathsf{T}} W_\alpha \nabla f \,\mathrm{d}\gamma = \frac{1}{1+\alpha}\int |\nabla f|^2 \,\mathrm{d}\gamma + \frac{\alpha}{(1+\alpha)(n-1)} \sum_{i<j} \int (L_{ij} f)^2 \,\mathrm{d}\gamma.$$

For $0 \le \alpha \le 1$, this is the convex combination, with coefficient $2\alpha/(1+\alpha)$, of the standard Gaussian Dirichlet form and the right side of (33). Hence (29) holds. Linear functions give equality. If $\alpha > 1$, the radial quadratic $f(x) = |x|^2 - n$ has variance $2n$, zero angular energy, and weighted energy $4n/(1+\alpha) < 2n$.

W_α is a calibrated Stein kernel relative to the canonical Gaussian metric I. Strictness in Theorem 3.1 gives (30) for $\alpha > 0$. □

These results distinguish three notions of uniqueness. Moment-measure uniqueness concerns the source potential up to translation. The Courtade–Fathi–Pananjady construction selects a Jacobian-form kernel for a quadratic functional. Normalized Poincaré minimization can be nonunique, even for the standard Gaussian. On the calibrated class, the Burg objective has the unique maximizer τ_{mm}.

We now pass from fixed-law geometry to conditional averaging, where Burg divergence admits a chain rule.

5 The conditional Burg tower

The following conditional identity separates the canonical projection deficit from variation within the conditional fibers.

Theorem 5.1 (Two-gap Burg tower)**.** *Let (X, T) be defined on a probability space $(\mathcal{Z}, Q)$, with X distributed according to μ and $T : \mathcal{Z} \to \mathbb{S}^n_{++}$. Put*

$$\bar\tau(x) = \mathbb{E}_Q[T \mid \mathsf{X} = x].$$

Assume that $\bar\tau \in \mathcal{C}_G(\mu)$, $\log\det\tau_\star \in L^1(\mu)$, $\mathbb{E}_Q \operatorname{Tr} T < \infty$, and $\mathbb{E}_Q |\log\det T| < \infty$. Then

$$\boxed{\int \log\det\tau_\star \,\mathrm{d}\mu - \mathbb{E}_Q \log\det T = J_{\mathrm{can}} + J_{\mathrm{fib}},} \tag{34}$$

where

$$J_{\mathrm{can}} = \int d_{\mathrm{LD}}\left(\tau_\star^{-1/2} \bar\tau \tau_\star^{-1/2}, \mathrm{I}\right) \mathrm{d}\mu, \tag{35}$$

$$J_{\mathrm{fib}} = \mathbb{E}_Q d_{\mathrm{LD}}\left(\bar\tau(\mathsf{X})^{-1/2} T \bar\tau(\mathsf{X})^{-1/2}, \mathrm{I}\right). \tag{36}$$

Both terms are nonnegative. Moreover, $J_{\mathrm{can}} = 0$ exactly when $\bar\tau = \tau_\star$ almost everywhere, and $J_{\mathrm{fib}} = 0$ exactly when $T = \bar\tau(\mathsf{X})$ almost surely.

Proof. An integrable almost surely positive-definite random matrix has an almost surely positive-definite conditional expectation. Indeed, if $\bar\tau$ were singular on a set of positive measure, measurable spectral calculus would give a measurable unit vector $v(\mathsf{X})$ in its null space there. Then

$\mathbb{E}_Q[v(\mathsf{X})^\mathsf{T} T v(\mathsf{X}) \mid \mathsf{X}] = 0$ on that set, contradicting $T \succ 0$ almost surely. Thus $\bar{\tau}^{-1/2}$ is defined almost everywhere. Conditional concavity gives $\log\det\bar{\tau}(\mathsf{X}) \geq \mathbb{E}_Q[\log\det T \mid \mathsf{X}]$, while its positive part is bounded by $\operatorname{Tr}\bar{\tau}$. The assumptions therefore make $\log\det\bar{\tau} \in L^1(\mu)$. Insert $\int \log\det\bar{\tau}\,\mathrm{d}\mu$ between the two terms on the left of (34). The first difference is Theorem 3.1. For the second, the relative matrix

$$R = \bar{\tau}(\mathsf{X})^{-1/2} T \bar{\tau}(\mathsf{X})^{-1/2}$$

has conditional mean I given X. Hence $\mathbb{E}_Q \operatorname{Tr} R = n$, and expanding $d_{\mathrm{LD}}(R, \mathrm{I})$ gives the second difference. Strictness of the scalar log-determinant divergence gives the equality cases. □

The matrix argument used in Theorem 3.2 applies also at the fiber level; T need not itself be a Stein competitor. In particular,

$$d_G(\bar{\tau}, \tau_\star)^2 \leq 4nJ_{\mathrm{can}}, \tag{37}$$

$$\left(\mathbb{E}_Q \left\| \bar{\tau}(\mathsf{X})^{-1/2}(T - \bar{\tau}(\mathsf{X}))\bar{\tau}(\mathsf{X})^{-1/2} \right\|_{S_1}\right)^2 \leq 4nJ_{\mathrm{fib}}. \tag{38}$$

The references in these two distances are different. No triangle inequality between T and $\tau_\star$ follows without relative ellipticity.

We first apply this chain rule to independent sums.

6 An exact convolution law

Let $X_1, \ldots, X_m$ be independent centered random vectors in $\mathbb{R}^n$. Choose nonzero a_i with

$$w_i = a_i^2, \qquad \sum_{i=1}^m w_i = 1,$$

and put $Z = \sum_i a_i X_i$. Assume that the input laws μ_i and the output law λ have finite second moments and regular moment representations with moment-map kernels τ_i and τ_Z. Assume the input kernels satisfy the global Stein identity against $C_c^1(\mathbb{R}^n; \mathbb{R}^n)$—equivalently for this argument, that their local identities extend by boundary cutoff to the composite tests obtained after conditioning on the other summands. Assume also the global mean-kernel identities

$$\mathbb{E}\tau_i(X_i) = \operatorname{Cov}(X_i), \qquad \mathbb{E}\tau_Z(Z) = \operatorname{Cov}(Z), \tag{39}$$

as justified by source integration by parts or a global cutoff, and assume

$$\mathbb{E}\left|\log\det\tau_i(X_i)\right| < \infty, \qquad \mathbb{E}\left|\log\det\tau_Z(Z)\right| < \infty. \tag{40}$$

On the product space define

$$S = \sum_{i=1}^m w_i \tau_i(X_i), \qquad \bar{\tau}(z) = \mathbb{E}[S \mid Z = z]. \tag{41}$$

Independent-sum integration by parts gives

$$\mathbb{E}[Z \cdot \Phi(Z)] = \mathbb{E}[\bar{\tau}(Z) : \nabla\Phi(Z)], \tag{42}$$

so $\bar{\tau}$ is a positive-definite Stein kernel of λ. We assume that it is calibrated, $\bar{\tau} \in \mathcal{C}_{\tau_Z^{-1}}(\lambda)$. This last condition is automatic under the cutoff criterion following Definition 2.2; it is recorded explicitly because a compact-test Stein identity alone does not justify an unbounded Hessian test.

Conditional-sum Stein kernels, including their optimal-metric behavior, are already present in the Stein-CLT literature and in [6, Corollary 1]. The following theorem refines that construction by decomposing its log-volume difference into three nonnegative terms.

Lemma 6.1 (Constant regular moment-map kernels)**.** *If the moment-map kernel of a regular moment representation is a constant matrix $C \succ 0$ almost everywhere, then the target law is Gaussian $N(0, C)$.*

Proof. The target density is positive and the kernel is continuous, so it is identically C on the connected support. Pulling back by the global moment-map diffeomorphism gives $\nabla^2\varphi = C$ on the connected source domain Y. Hence, for some $b \in \mathbb{R}^n$ and $a \in \mathbb{R}$,

$$\nabla\varphi(y) = Cy + b, \qquad \varphi(y) = \frac{1}{2} y^\mathsf{T} C y + b \cdot y + a \quad (y \in Y).$$

If Y were a proper nonempty open convex subset of $\mathbb{R}^n$, it would have a finite boundary point: start in Y, join that point to any point outside Y, and take the first exit along the segment. Along a sequence in Y approaching this boundary point, $Cy + b$ stays bounded. This contradicts the essential smoothness in the Legendre hypothesis, which requires the gradient norm to diverge at a finite boundary. Therefore $Y = \mathbb{R}^n$. The source law is $N(-C^{-1}b, C^{-1})$, and its image under $y \mapsto Cy + b$ is $N(0, C)$. □

For a regular law η, write

$$\mathscr{V}(\eta) = \int \log\det \tau_{\mathrm{mm}}^{\eta} \, \mathrm{d}\eta. \tag{43}$$

Theorem 6.2 (Three-gap Burg convolution identity)**.** *Under the preceding assumptions,*

$$\boxed{\mathscr{V}(\lambda) - \sum_{i=1}^{m} w_i \mathscr{V}(\mu_i) = J_{\mathrm{can}} + J_{\mathrm{cond}} + J_{\mathrm{mix}},} \tag{44}$$

where

$$J_{\mathrm{can}} = \mathbb{E} d_{\mathrm{LD}}\Big(\tau_Z(Z)^{-1/2}\bar{\tau}(Z)\tau_Z(Z)^{-1/2}, \mathrm{I}\Big), \tag{45}$$

$$J_{\mathrm{cond}} = \mathbb{E} d_{\mathrm{LD}}\Big(\bar{\tau}(Z)^{-1/2} S \bar{\tau}(Z)^{-1/2}, \mathrm{I}\Big), \tag{46}$$

$$J_{\mathrm{mix}} = \mathbb{E} \sum_{i=1}^{m} w_i d_{\mathrm{LD}}\Big(S^{-1/2}\tau_i(X_i)S^{-1/2}, \mathrm{I}\Big). \tag{47}$$

All three terms are nonnegative. In particular,

$$\mathscr{V}(\lambda) \geq \sum_i w_i \mathscr{V}(\mu_i). \tag{48}$$

If at least two weights are positive, equality holds exactly when all participating inputs are centered Gaussians with one common covariance matrix.

Proof. The trace integrability following from (39) and the concavity bounds

$$\log\det S \geq \sum_i w_i \log\det \tau_i(X_i), \qquad \log\det \bar{\tau}(Z) \geq \mathbb{E}[\log\det S \mid Z]$$

make all intermediate logarithmic terms below integrable under (40). Insert two intermediate terms:

$$\begin{aligned}\mathscr{V}(\lambda) - \sum_i w_i \mathscr{V}(\mu_i) = {} & [\mathscr{V}(\lambda) - \mathbb{E}\log\det\bar{\tau}(Z)] \\ & + [\mathbb{E}\log\det\bar{\tau}(Z) - \mathbb{E}\log\det S] \\ & + [\mathbb{E}\log\det S - \sum_i w_i \mathbb{E}\log\det \tau_i(X_i)].\end{aligned}$$

The first two brackets are Theorem 5.1 applied on the product space with $\mathsf{X} = Z$ and $T = S$: its fiber gap is J_{cond} because $\mathbb{E}[S \mid Z] = \bar{\tau}(Z)$. In the third bracket,

$$\sum_i w_i S^{-1/2} \tau_i(X_i) S^{-1/2} = \mathrm{I}$$

pointwise, so the weighted relative trace terms cancel. This proves (44).

If equality holds, $J_{\mathrm{mix}} = 0$, so $\tau_i(X_i) = S$ almost surely for every participating i. In particular, the independent random matrices $\tau_i(X_i)$ are almost surely equal and therefore deterministic, with a common value C. Their means are the input covariance matrices by (39), and Lemma 6.1 makes every participating input $N(0, C)$. The converse is immediate. □

If the moment sources ν_i, ν_Z and all differential entropies are finite, the moment-map change of variables gives $\mathscr{V}(\eta) = h(\eta) - h(\nu_\eta)$. Theorem 6.2 therefore also yields

$$h(\lambda) - h(\nu_Z) \geq \sum_i w_i \big[h(\mu_i) - h(\nu_i)\big]. \tag{49}$$

6.1 The canonical determinant deficit

For a regular centered law with covariance $\Sigma \succ 0$, assume explicitly that

$$\int \tau_{\mathrm{mm}} \,\mathrm{d}\mu = \Sigma, \qquad \int |\log\det \tau_{\mathrm{mm}}| \,\mathrm{d}\mu < \infty. \tag{50}$$

The first identity requires source integration by parts or a legitimate global cutoff; it is not inferred from the local compact-test identity on a domain with boundary. Define

$$\mathfrak{D}_{\det}(\mu) = \log\det\Sigma - \mathscr{V}(\mu). \tag{51}$$

This quantity is affine invariant. With $Z = \Sigma^{-1/2} \tau_{\mathrm{mm}} \Sigma^{-1/2}$, (50) gives $\int \operatorname{Tr} Z \,\mathrm{d}\mu = n$, and hence

$$\mathfrak{D}_{\det}(\mu) = -\int \log\det Z \,\mathrm{d}\mu = \int d_{\mathrm{LD}}(Z, \mathrm{I}) \,\mathrm{d}\mu \geq 0. \tag{52}$$

Equality holds only when $Z = \mathrm{I}$ almost surely, or $\tau_{\mathrm{mm}} = \Sigma$ almost surely. Lemma 6.1 then gives $\mu = N(0, \Sigma)$.

Corollary 6.3 (Convolution monotonicity and equality cases). *If the inputs in Theorem 6.2 have a common covariance* Σ*, then*

$$\boxed{\mathfrak{D}_{\det}(\lambda) \le \sum_i w_i \mathfrak{D}_{\det}(\mu_i).} \tag{53}$$

For iid inputs and at least two positive weights, the inequality is strict unless the common law is Gaussian. In particular, the deficit strictly decreases along dyadic normalized convolution of a non-Gaussian law.

Without the common-covariance assumption, writing $\Sigma_i = \mathrm{Cov}(\mu_i)$ and $\Sigma_Z = \sum_i w_i \Sigma_i$ gives the corresponding unequal-covariance bound

$$\mathfrak{D}_{\det}(\lambda) \le \sum_i w_i \mathfrak{D}_{\det}(\mu_i) + \left[\log\det\left(\sum_i w_i \Sigma_i\right) - \sum_i w_i \log\det \Sigma_i\right]. \tag{54}$$

The bracket is the nonnegative covariance Jensen gap.

This argument gives dyadic monotonicity and the weighted one-step inequality (53), but does not compare consecutive normalized sums in general.

6.2 Central-limit convergence and a rate

Let X_i be iid with covariance Σ and whiten to $\Sigma = \mathrm{I}$. Put

$$Z_N = N^{-1/2} \sum_{i=1}^{N} X_i, \qquad M_N = \frac{1}{N} \sum_{i=1}^{N} \tau_{\mathrm{mm}}(X_i).$$

Lemma 6.4 (Empirical log determinant). *Let* T_i *be iid random matrices in* $\mathbb{S}^n_{++}$ *with* $\mathbb{E}T_1 = \mathrm{I}$ *and* $\mathbb{E}\left|\log\det T_1\right| < \infty$*, and set* $A_N = N^{-1}\sum_{i=1}^{N} T_i$*. Then*

$$\mathbb{E}\log\det A_N \longrightarrow 0. \tag{55}$$

If, in addition, $T_1 \succeq m\mathrm{I}$ *almost surely for some* $m > 0$ *and* $\mathbb{E}\left\|T_1 - \mathrm{I}\right\|_{\mathrm{HS}}^2 < \infty$*, then*

$$0 \le -\mathbb{E}\log\det A_N \le \frac{1}{2m^2 N}\mathbb{E}\left\|T_1 - \mathrm{I}\right\|_{\mathrm{HS}}^2. \tag{56}$$

Proof. For a positive-semidefinite matrix, $2\left|(T_1)_{jk}\right| \le (T_1)_{jj} + (T_1)_{kk}$. Thus every entry is integrable, and the scalar strong law applied entrywise gives $A_N \to \mathrm{I}$ almost surely.

We use the following elementary uniform-integrability fact. If Y_i are iid copies of an integrable scalar variable and $\bar{Y}_N = N^{-1}\sum_i Y_i$, write $Y_i = B_i^{(K)} + R_i^{(K)}$, where $B_i^{(K)} = Y_i \mathbf{1}_{\{|Y_i| \le K\}}$. Since $\left|N^{-1}\sum_i B_i^{(K)}\right| \le K$, on $\{\left|\bar{Y}_N\right| > 2K\}$ the average of the remainders has magnitude larger than K. The triangle inequality then gives

$$\sup_N \mathbb{E}\left[\left|\bar{Y}_N\right|; \left|\bar{Y}_N\right| > 2K\right] \le 2\mathbb{E}\left[|Y_1|; |Y_1| > K\right] \longrightarrow 0.$$

Hence the sample averages are uniformly integrable.

Apply this fact first to $Y_i = \operatorname{Tr} T_i$. Since $(\log\det A_N)^+ \le \operatorname{Tr} A_N$, the positive parts of $\log\det A_N$ are uniformly integrable. Next put $L_i = \log\det T_i$. Concavity of $\log\det$ gives

$$\log\det A_N \ge \frac{1}{N}\sum_{i=1}^{N} L_i,$$

so $(\log\det A_N)^- \le (N^{-1}\sum_i L_i)^-$, and the negative parts are uniformly integrable by the same fact. Since $\log\det A_N \to 0$ almost surely, uniform integrability and Vitali's theorem prove (55).

For the rate, put $H_N = A_N - \mathrm{I}$ and $A_{N,t} = \mathrm{I} + tH_N$. The mean identity implies $m \le 1$, so $A_{N,t} \succeq m\mathrm{I}$ for $0 \le t \le 1$. Moreover,

$$\mathrm{Tr}(A_{N,t}^{-1}H_N A_{N,t}^{-1}H_N) = \left\| A_{N,t}^{-1/2}H_N A_{N,t}^{-1/2}\right\|_{\mathrm{HS}}^2 \le m^{-2}\left\|H_N\right\|_{\mathrm{HS}}^2.$$

Taylor's integral formula therefore gives

$$\begin{aligned}-\log\det A_N &= -\operatorname{Tr} H_N + \int_0^1 (1-t)\operatorname{Tr}(A_{N,t}^{-1}H_N A_{N,t}^{-1}H_N)\,\mathrm{d}t \\ &\le -\operatorname{Tr} H_N + \frac{1}{2m^2}\left\|H_N\right\|_{\mathrm{HS}}^2.\end{aligned}$$

Taking expectations, independence and $\mathbb{E}(T_i - \mathrm{I}) = 0$ give

$$\mathbb{E}\left\|H_N\right\|_{\mathrm{HS}}^2 = \frac{1}{N}\mathbb{E}\left\|T_1 - \mathrm{I}\right\|_{\mathrm{HS}}^2.$$

Finally, Jensen's inequality gives $\mathbb{E}\log\det A_N \le 0$, proving (56). □

Proposition 6.5 (Determinant deficit along the CLT)**.** *Suppose that the input law and every output law $\mathcal{L}(Z_N)$ satisfy the regularity, global mean-kernel, and absolute log-integrability hypotheses of Theorem 6.2, and that the conditional-sum kernel is calibrated for each N. Then*

$$\mathfrak{D}_{\det}(\mathcal{L}(Z_N)) \longrightarrow 0. \tag{57}$$

If additionally $\tau_{\mathrm{mm}}(X) \succeq m\mathrm{I}$ almost surely for some $m > 0$ and $\mathbb{E}\left\|\tau_{\mathrm{mm}}(X) - \mathrm{I}\right\|_{\mathrm{HS}}^2 < \infty$, then

$$\mathfrak{D}_{\det}(\mathcal{L}(Z_N)) \le \frac{1}{2m^2 N}\mathbb{E}\left\|\tau_{\mathrm{mm}}(X) - \mathrm{I}\right\|_{\mathrm{HS}}^2. \tag{58}$$

Proof. Apply Theorem 6.2 with $a_i = N^{-1/2}$ and $w_i = 1/N$. Then the product-space matrix S in (41) is M_N, and the first two gaps give the exact identity

$$\mathscr{V}(\mathcal{L}(Z_N)) - \mathbb{E}\log\det M_N = J_{\mathrm{can}} + J_{\mathrm{cond}} \ge 0. \tag{59}$$

The input mean-kernel identity gives $\mathbb{E}\tau_{\mathrm{mm}}(X) = \mathrm{I}$, so Lemma 6.4 applies to $T_i = \tau_{\mathrm{mm}}(X_i)$. The exact representation (52), applied to the output, gives

$$0 \le \mathfrak{D}_{\det}(\mathcal{L}(Z_N)) = -\mathscr{V}(\mathcal{L}(Z_N)) \le -\mathbb{E}\log\det M_N.$$

The limit and, under the additional lower and second-moment hypotheses, the rate now follow from (55) and (56). □

6.3 Stability under ellipticity

Under uniform ellipticity, the mixture term controls the Stein discrepancies of the two inputs.

Theorem 6.6 (Stability under ellipticity)**.** *Under the hypotheses of Theorem 6.2, suppose there are two whitened inputs, let their weights be $w \in (0,1)$ and $1-w$, and assume*

$$m\mathrm{I} \preceq \tau_i(X_i) \preceq M\mathrm{I} \qquad (i=1,2)$$

almost surely. Then

$$W_2(\mu_1,\gamma)^2 + W_2(\mu_2,\gamma)^2 \le \frac{2M^3}{m\,w(1-w)}\left[\mathscr{V}(\lambda) - w\mathscr{V}(\mu_1) - (1-w)\mathscr{V}(\mu_2)\right]. \tag{60}$$

Proof. For $x \in [m/M, M/m]$,

$$x - \log x - 1 \ge \frac{m}{2M}(x-1)^2.$$

Since $m\mathrm{I} \preceq S \preceq M\mathrm{I}$, functional calculus and (47) imply

$$J_{\mathrm{mix}} \ge \frac{m}{2M^3} w(1-w)\mathbb{E}\left\|\tau_1(X_1) - \tau_2(X_2)\right\|_{\mathrm{HS}}^2. \tag{61}$$

Independence and $\mathbb{E}\tau_i = \mathrm{I}$ turn the last expectation into

$$\mathbb{E}\left\|\tau_1 - \mathrm{I}\right\|_{\mathrm{HS}}^2 + \mathbb{E}\left\|\tau_2 - \mathrm{I}\right\|_{\mathrm{HS}}^2.$$

The Stein-discrepancy transport bound $W_2(\mu_i,\gamma)^2 \le \mathbb{E}\left\|\tau_i - \mathrm{I}\right\|_{\mathrm{HS}}^2$ [4], together with $J_{\mathrm{mix}} \le J_{\mathrm{can}} + J_{\mathrm{cond}} + J_{\mathrm{mix}}$, proves the result. □

Convolution is the discrete application of the tower: conditioning produces the first two gaps, and matrix mixing supplies the third. We next turn to a continuous conditional family, where the same two tower gaps enter a time-integrated variance identity.

7 Burg–Parseval localization

The tower identity and the score–residual square function meet naturally in covariance-normalized stochastic localization. Let ρ_t solve

$$\mathrm{d}\rho_t(x) = \rho_t(x)\left\langle A_t^{-1/2}(x-a_t), \mathrm{d}B_t\right\rangle, \qquad a_t = \int x\rho_t(x)\,\mathrm{d}x, \quad A_t = \mathrm{Cov}_{\rho_t}(X), \tag{62}$$

from $\rho_0 = \rho$. Fix a deterministic time t for which the process can be obtained by stopped exhaustion, and assume the following finite-time regularity conditions.

Assumption 7.1 (Admissible finite horizon)**.** At the chosen time t:

(i) the stopped solutions in (62) converge to jointly measurable densities $(\omega,x) \mapsto \rho_s^\omega(x)$ for $0 \le s \le t$, with $A_s \succ 0$ and unit mass, and the stochastic Fubini interchanges below are valid;

(ii) $\rho_s(x)$ and $e^s\rho_s(x)(x-a_s)$ pass from their stopped local martingale identities to true expectation identities, so in particular $\mathbb{E}\rho_t(x) = \rho(x)$ and (64) holds for almost every x;

(iii) every translated posterior has a regular moment representation, and its kernel $(\omega,x) \mapsto \tau_t^\omega(x)$ is jointly measurable;

(iv) with Q_t defined below, $\mathbb{E}_{Q_t}\operatorname{Tr}(e^t\tau_t(X)) < \infty$, $\mathbb{E}_{Q_t}\left|\log\det(e^t\tau_t(X))\right| < \infty$, and all posterior Stein and log-determinant integrals admit the displayed Fubini interchanges.

Only the process up to time t is used.

Itô's formula gives $\mathrm{d}a_t = A_t^{1/2}\,\mathrm{d}B_t$ and shows that

$$e^t\rho_t(x)(x-a_t) \tag{63}$$

is a local martingale. Consequently admissibility of the horizon gives

$$e^t\mathbb{E}[\rho_t(x)(x-a_t)] = \rho(x)x. \tag{64}$$

For each translated posterior at time t, write τ_t for its moment-map kernel, evaluated in the original x coordinate, and ν_t for its moment source. Define

$$\bar{\tau}_t(x) = e^t\frac{\mathbb{E}[\rho_t(x)\tau_t(x)]}{\rho(x)}. \tag{65}$$

Posterior Stein integration by parts and (64) show that $\bar{\tau}_t$ is a Stein kernel of μ. Put the joint law

$$Q_t(\,\mathrm{d}\omega,\,\mathrm{d}x) = \rho_t^\omega(x)\,\mathrm{d}x\,\mathbb{P}(\,\mathrm{d}\omega).$$

Its x-marginal is μ, and

$$\bar{\tau}_t(x) = \mathbb{E}_{Q_t}[e^t\tau_t(x)\mid X=x]. \tag{66}$$

Corollary 7.2 (Finite-time Burg–Bellman identity)**.** *Assume Assumption 7.1, $\bar{\tau}_t\in\mathcal{C}_G(\mu)$, and $\log\det\tau_\star\in L^1(\mu)$. Then*

$$\boxed{\mathscr{V}(\mu) - nt - \mathbb{E}\mathscr{V}(\mu_t) = J_{\mathrm{can}}(t) + J_{\mathrm{fib}}(t),} \tag{67}$$

where

$$J_{\mathrm{can}}(t) = \int d_{\mathrm{LD}}\left(\tau_\star^{-1/2}\bar{\tau}_t\tau_\star^{-1/2},\mathrm{I}\right)\,\mathrm{d}\mu,$$
$$J_{\mathrm{fib}}(t) = \mathbb{E}_{Q_t}d_{\mathrm{LD}}\left(\bar{\tau}_t(X)^{-1/2}e^t\tau_t(X)\bar{\tau}_t(X)^{-1/2},\mathrm{I}\right).$$

If

$$\mathfrak{B}(t) = J_{\mathrm{can}}(t) + J_{\mathrm{fib}}(t), \qquad P_t = \tau_\star(X)^{-1/2}e^t\tau_t(X)\tau_\star(X)^{-1/2}, \tag{68}$$

then the same total gap has the one-reference representation

$$\boxed{\mathfrak{B}(t) = \mathbb{E}_{Q_t}d_{\mathrm{LD}}(P_t,\mathrm{I}).} \tag{69}$$

In particular,

$$\mathscr{V}(\mu) \geq \int \log\det\bar{\tau}_t\,\mathrm{d}\mu \geq nt + \mathbb{E}\mathscr{V}(\mu_t). \tag{70}$$

The two native trace distances in (37) and (38) have squared sum at most

$$4n\,[J_{\mathrm{can}}(t) + J_{\mathrm{fib}}(t)]. \tag{71}$$

In the single initial metric one also has

$$\mathbb{E}_{Q_t}\left\|P_t-\mathrm{I}\right\|_{S_1} \leq 2\sqrt{n\mathfrak{B}(t)}. \tag{72}$$

Proof. Apply Theorem 5.1 with $T = e^t\tau_t(X)$ under Q_t, and use $\log\det(e^t\tau_t) = nt + \log\det\tau_t$. The quantitative statement is (37)–(38). Calibration and (66) give

$$\mathbb{E}_{Q_t}\operatorname{Tr}P_t = \int \operatorname{Tr}(\tau_\star^{-1}\bar{\tau}_t)\,\mathrm{d}\mu = n,$$

while log-determinant additivity gives

$$-\mathbb{E}_{Q_t}\log\det P_t = \mathscr{V}(\mu) - nt - \mathbb{E}\mathscr{V}(\mu_t) = \mathfrak{B}(t).$$

Expanding d_{LD} proves (69), and the proof of (17), applied on the product space, gives (72). $\square$

Suppose in addition that all differential entropies are finite and that the adaptive Gaussian-channel identity

$$\mathbb{E}h(\mu_t) = h(\mu) - \frac{nt}{2} \tag{73}$$

holds. If it is justified for a particular localization by a stopped feedback-information argument in the spirit of [13], then change of variables in the moment map gives $\mathscr{V}(\eta) = h(\eta) - h(\nu_\eta)$. Therefore

$$\mathbb{E}h(\nu_t) - h(\nu_0) - \frac{nt}{2} = J_{\mathrm{can}}(t) + J_{\mathrm{fib}}(t). \tag{74}$$

This identity is finite-time; it uses no terminal limit and does not imply that either gap vanishes.

7.1 The finite-time Burg–Parseval bridge

For a posterior quantity, write $\mathbb{E}_s$ for expectation under μ_s. The following assumption records the stochastic-integrability requirements used below.

Assumption 7.3 (Admissible Burg–Parseval horizon)**.** Fix $T > 0$ and $f \in C_c^2(\Omega)$. Assume the following on $[0, T]$.

(i) Assumption 7.1 holds in a jointly measurable form for almost every $s \le T$, with posterior kernels $(s, \omega, x) \mapsto \tau_s^\omega(x)$ and back-projections $\bar{\tau}_s \in \mathcal{C}_G(\mu)$.

(ii) One has $\log\det\tau_\star \in L^1(\mu)$ and, for almost every $s \le T$,

$$\mathbb{E}_{Q_s}\operatorname{Tr}(e^s\tau_s(X)) < \infty, \qquad \mathbb{E}_{Q_s}|\log\det(e^s\tau_s(X))| < \infty.$$

These quantities and the required Fubini interchanges are integrable in time, and $\mathfrak{B}(s) = J_{\mathrm{can}}(s) + J_{\mathrm{fib}}(s) < \infty$ is measurable.

(iii) The stopped score and posterior-variance martingales below pass to expectation, and the posterior Stein identities can be tested against f and affine cutoffs. In particular,

$$\mathbb{E}_s\tau_s = A_s, \qquad \mathbb{E}_s[\tau_s\nabla f] = \mathbb{E}_s[(f - \mathbb{E}_s f)(X - a_s)]. \tag{75}$$

Set

$$m_s = \mathbb{E}_s f, \qquad Y_s = X - a_s, \qquad b_s = \mathbb{E}_s[(f - m_s)Y_s], \quad c_s = A_s^{-1}b_s, \tag{76}$$

$$\Xi_s = \mathbb{E}_s[(f - m_s - c_s\cdot Y_s)Y_s\otimes Y_s], \tag{77}$$

$$\rho_s^f = \left\|A_s^{-1/2}\Xi_s A_s^{-1/2}\right\|_{\mathrm{HS}}^2, \qquad S_s = b_s^{\mathsf{T}}A_s^{-1}b_s. \tag{78}$$

Write $q(x) = \nabla f(x)$, $r_s(x) = q(x) - c_s$, and

$$\mathcal{E}_\star(f) = \int q^\mathsf{T} \tau_\star q \,\mathrm{d}\mu, \tag{79}$$

$$\mathcal{H}_\star^T(f) = \int_0^T e^{-s} \mathbb{E}_{Q_s}[r_s(X)^\mathsf{T} \tau_\star(X) r_s(X)] \,\mathrm{d}s. \tag{80}$$

The following elementary spectral estimate will convert the Burg information into a directional error bound. Define

$$\omega_n(a) = \min\left\{2\sqrt{na},\, a + \sqrt{a^2 + 2a}\right\}, \qquad a \geq 0. \tag{81}$$

Lemma 7.4 (Burg control of a random direction)**.** *Let P be a random matrix in $\mathbb{S}^n_{++}$ such that $\mathbb{E}\operatorname{Tr} P = n$ and $\mathbb{E} d_{\mathrm{LD}}(P, \mathrm{I}) = B < \infty$. Then*

$$\mathbb{E}\left\|P - \mathrm{I}\right\|_{\mathrm{op}} \leq \omega_n(B). \tag{82}$$

Consequently, if a random vector u satisfies $|u|^2 \leq K$, then

$$\left|\mathbb{E}[u^\mathsf{T}(P - \mathrm{I})u]\right| \leq K\omega_n(B). \tag{83}$$

Proof. The proof of (17), which uses only $\mathbb{E}\operatorname{Tr} P = n$, gives $\mathbb{E}\left\|P - \mathrm{I}\right\|_{\mathrm{op}} \leq 2\sqrt{nB}$. For the other bound, put $\phi(x) = x - \log x - 1$. If $0 < x \leq 1$, then $1 - x \leq \sqrt{2\phi(x)}$. If $x = 1 + y \geq 1$, then

$$\phi(1 + y) = \int_0^y \frac{u}{1 + u}\,\mathrm{d}u \geq \frac{y^2}{2(1 + y)},$$

and hence $y \leq \phi(1 + y) + \sqrt{\phi(1 + y)^2 + 2\phi(1 + y)}$. Applying these inequalities to the eigenvalue furthest from one gives

$$\left\|P - \mathrm{I}\right\|_{\mathrm{op}} \leq d_{\mathrm{LD}}(P, \mathrm{I}) + \sqrt{d_{\mathrm{LD}}(P, \mathrm{I})^2 + 2d_{\mathrm{LD}}(P, \mathrm{I})}.$$

The scalar function on the right is increasing and concave. Jensen's inequality gives the second term in (81); taking the better of the two estimates proves (82). Equation (83) follows from $\left|u^\mathsf{T}(P - \mathrm{I})u\right| \leq K\left\|P - \mathrm{I}\right\|_{\mathrm{op}}$. □

Theorem 7.5 (Finite-time Burg–Parseval bridge)**.** *Under Assumption 7.3, the score residual has the exact finite-time decomposition*

$$\operatorname{Var}_\mu f = S_0 + \mathbb{E}\int_0^T \rho_s^f \,\mathrm{d}s + \mathbb{E}[\operatorname{Var}_T(f) - S_T], \tag{84}$$

where the last term is nonnegative. It is joined to the canonical Dirichlet form by

$$\boxed{\mathbb{E}\int_0^T \rho_s^f \,\mathrm{d}s = (1 - e^{-T})\mathcal{E}_\star(f) + \mathbb{E}S_T - S_0 - \mathcal{H}_\star^T(f) + \mathcal{R}_T(f),} \tag{85}$$

with the signed error

$$\mathcal{R}_T(f) = \int_0^T e^{-s}\mathbb{E}_{Q_s}\Big[q(X)^\mathsf{T}(e^s\tau_s(X) - \tau_\star(X))q(X) - r_s(X)^\mathsf{T}(e^s\tau_s(X) - \tau_\star(X))r_s(X)\Big]\,\mathrm{d}s. \tag{86}$$

Suppose deterministic envelopes K_q, K_r satisfy, for almost every s,

$$q(X)^{\mathsf{T}}\tau_\star(X)q(X) \le K_q(s), \qquad r_s(X)^{\mathsf{T}}\tau_\star(X)r_s(X) \le K_r(s) \quad Q_s\text{-almost surely.} \tag{87}$$

Then the two Burg gaps control the bridge error through their sum:

$$\boxed{|\mathcal{R}_T(f)| \le \int_0^T e^{-s}[K_q(s)+K_r(s)]\omega_n(J_{\mathrm{can}}(s)+J_{\mathrm{fib}}(s))\,\mathrm{d}s.} \tag{88}$$

Proof. Put

$$\mathcal{R}_s^f = \mathbb{E}_s[(f-m_s)Y_s\otimes Y_s], \qquad \mathcal{T}_s[v] = \mathbb{E}_s[(v\cdot Y_s)Y_s\otimes Y_s],$$

and, for the standard basis (e_k), set

$$R_{s,k} = \mathcal{R}_s^f A_s^{-1/2}e_k, \qquad T_{s,k} = \mathcal{T}_s[A_s^{-1/2}e_k].$$

The localization equations give

$$\mathrm{d}b_s = \sum_k R_{s,k}\,\mathrm{d}B_{s,k} - b_s\,\mathrm{d}s, \qquad \mathrm{d}A_s = \sum_k T_{s,k}\,\mathrm{d}B_{s,k} - A_s\,\mathrm{d}s.$$

Applying the matrix inverse rule to $S_s = b_s^{\mathsf{T}}A_s^{-1}b_s$ gives drift

$$-S_s + \sum_k (R_{s,k} - T_{s,k}c_s)^{\mathsf{T}}A_s^{-1}(R_{s,k}-T_{s,k}c_s).$$

Since $R_{s,k} - T_{s,k}c_s = \Xi_s A_s^{-1/2}e_k$, the sum is ρ_s^f. After taking expectations,

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}S_s = -\mathbb{E}S_s + \mathbb{E}\rho_s^f. \tag{89}$$

On the other hand, the conditional mean m_s has martingale coefficient $b_s^{\mathsf{T}}A_s^{-1/2}$, so

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}\operatorname{Var}_s(f) = -\mathbb{E}S_s. \tag{90}$$

Subtracting (89) from (90) and integrating proves (84). The remainder is nonnegative because $c_s\cdot Y_s$ is the orthogonal projection of $f-m_s$ onto the posterior linear functions.

By (75), $b_s = \mathbb{E}_s[\tau_s q]$ and $A_s = \mathbb{E}_s\tau_s$. Expanding the square therefore gives the pathwise identity

$$S_s = \mathbb{E}_s[q^{\mathsf{T}}\tau_s q - r_s^{\mathsf{T}}\tau_s r_s]. \tag{91}$$

Since Q_s has x-marginal μ, insertion of $\tau_\star$ into (91) gives

$$\begin{aligned}\mathbb{E}S_s &= e^{-s}\mathcal{E}_\star(f) - e^{-s}\mathbb{E}_{Q_s}[r_s^{\mathsf{T}}\tau_\star r_s]\\ &\quad + e^{-s}\mathbb{E}_{Q_s}\left[q^{\mathsf{T}}(e^s\tau_s-\tau_\star)q - r_s^{\mathsf{T}}(e^s\tau_s-\tau_\star)r_s\right].\end{aligned}$$

Now integrate $\mathbb{E}\rho_s^f = (\mathbb{E}S_s)' + \mathbb{E}S_s$ from (89). This proves (85).

For the quantitative statement, normalize $P_s = \tau_\star^{-1/2}e^s\tau_s\tau_\star^{-1/2}$. Equation (69) and Lemma 7.4, applied at each time, bound the two terms in (86) by $K_q(s)\omega_n(\mathfrak{B}(s))$ and $K_r(s)\omega_n(\mathfrak{B}(s))$. Integration gives (88). □

Corollary 7.6 (Burg-stable Parseval and Poincaré deficit)**.** *Assume the hypotheses above on every compact time interval, that the posterior is the conditional law of X, and that $\mathbb{E}\operatorname{Var}_s(f) \to 0$. If the right side of (88) is integrable on $[0, \infty)$, then*

$$\boxed{\operatorname{Var}_\mu f = S_0 + \mathbb{E}\int_0^\infty \rho_s^f \, \mathrm{d}s = \mathcal{E}_\star(f) - \mathcal{H}_\star^\infty(f) + \mathcal{R}_\infty(f).} \tag{92}$$

In particular,

$$|[\mathcal{E}_\star(f) - \operatorname{Var}_\mu f] - \mathcal{H}_\star^\infty(f)| \le \int_0^\infty e^{-s}[K_q(s) + K_r(s)]\omega_n(J_{\mathrm{can}}(s) + J_{\mathrm{fib}}(s)) \, \mathrm{d}s. \tag{93}$$

If $A_0 = \mathrm{I}$, then $S_0 = |\mathbb{E}[Xf(X)]|^2$.

Proof. The projection property gives $0 \le S_s \le \operatorname{Var}_s(f)$, so $\mathbb{E}S_s \to 0$. Letting $T \to \infty$ in (84) gives the score–residual Parseval identity. The assumed bound makes (86) absolutely convergent, while monotone convergence applies to (80). Letting $T \to \infty$ in (85) gives the second equality in (92); rearrangement and (88) give (93). $\square$

Corollary 7.7 (Future Parseval tail)**.** *Fix a deterministic time T at which Assumption 7.1 and the hypotheses of Corollary 7.2 hold, so in particular $\bar{\tau}_T \in \mathcal{C}_G(\mu)$ and (69) is valid. Put*

$$z_T = \tau_\star(X)^{1/2} r_T(X), \qquad H_T = \mathbb{E}_{Q_T} |z_T|^2, \qquad K_T = \operatorname*{ess\,sup}_{Q_T} |z_T|^2,$$

and assume $K_T < \infty$. Then the moment-map weighted Poincaré inequality for the posterior gives

$$\boxed{\mathbb{E}[\operatorname{Var}_T(f) - S_T] \le e^{-T}\left[H_T + K_T\big(2\sqrt{\mathfrak{B}(T)} + \mathfrak{B}(T)\big)\right].} \tag{94}$$

Under the global revelation hypotheses of Corollary 7.6, the left side is exactly

$$\mathbb{E}\int_T^\infty \rho_s^f \, \mathrm{d}s. \tag{95}$$

Proof. Let $h_T = f - m_T - c_T \cdot Y_T$. Posterior orthogonal regression gives $\operatorname{Var}_T(f) - S_T = \operatorname{Var}_T(h_T)$, while $\nabla h_T = r_T$. Fathi's weighted Poincaré inequality and the definition of P_T yield

$$\mathbb{E}\operatorname{Var}_T(h_T) \le e^{-T}\mathbb{E}_{Q_T}[z_T^\mathsf{T} P_T z_T].$$

Write $P_T^{1/2} = \mathrm{I} + D_T$. The root estimate in (16), now used with the one-reference identity (69), gives

$$\mathbb{E}_{Q_T} \|D_T\|_{\mathrm{HS}}^2 \le \mathfrak{B}(T).$$

Since

$$z_T^\mathsf{T} P_T z_T = |z_T|^2 + 2z_T^\mathsf{T} D_T z_T + |D_T z_T|^2,$$

Cauchy–Schwarz and the definition of K_T prove (94). Integrating the derivative of $\mathbb{E}[\operatorname{Var}_s(f) - S_s]$ from T to infinity proves (95). $\square$

Remark 7.8 (The baseline cannot be removed). The following computation is valid in the natural Gaussian L^2 domain of the identities; equivalently, apply them first to even compactly supported cutoffs and pass to the limit. For standard Gaussian localization, $\tau_\star = \mathrm{I}$ and $e^s \tau_s = \mathrm{I}$, so both Burg gaps vanish at every time. Nevertheless, for $f(x) = (x_1^2 - 1)/\sqrt{2}$ the affine score at time zero is zero and the Parseval residual integral equals $\mathrm{Var}_\gamma f = 1$. Thus no estimate of the nonlinear score-residual lifetime by the Burg gaps alone can hold, even in the perfectly elliptic Gaussian case. The positive term $\mathcal{H}_\star$ in (92) is indispensable.

The preceding localization results use regular posterior moment maps. We close the theorem development by isolating what remains stable when the law and its Stein kernel vary and the inverse reference metric is not controlled.

8 Approximation and nonsmooth limit kernels

Proposition 2.3 treats structured nonsmooth targets directly. For a general varying target law, however, the inverse reference metric also moves. The following proposition gives the compactness statement available without inverse-metric control.

Proposition 8.1 (Weak stability of positive Stein kernels). *Let $\mu_k \to \mu$ in W_2, with all laws centered. Suppose that $X_k \sim \mu_k$ has a symmetric positive-semidefinite* global *Stein kernel τ_k, meaning that the identity holds against $C_c^\infty(\mathbb{R}^n; \mathbb{R}^n)$, and*

$$\sup_k \mathbb{E} \, \|\tau_k(X_k)\|_{\mathrm{HS}}^2 < \infty. \tag{96}$$

Then the joint laws of $(X_k, \tau_k(X_k))$ are tight. After passing to a subsequence that realizes the lower limit in (99) and then to a further joint weak limit (X, M), the conditional barycenter

$$\tau_\infty(X) = \mathbb{E}[M \mid X] \tag{97}$$

is a symmetric positive-semidefinite Stein kernel of μ and

$$\mathbb{E}\tau_\infty(X) = \mathrm{Cov}(\mu), \tag{98}$$

$$\mathbb{E} \, \|\tau_\infty(X)\|_{\mathrm{HS}}^2 \le \liminf_k \mathbb{E} \, \|\tau_k(X_k)\|_{\mathrm{HS}}^2 \, . \tag{99}$$

If, with one constant C, all k satisfy

$$\mathrm{Var}_{\mu_k} f \le C \int \nabla f^{\mathsf{T}} \tau_k \nabla f \, \mathrm{d}\mu_k,$$

then the same inequality holds for μ and τ_∞ on $C_c^\infty(\mathbb{R}^n)$. Every original subsequence has a further joint limit with the analogous energy bound relative to that subsequence.

Proof. The W_2 convergence makes the first coordinates tight, and (96) makes the matrix coordinates tight. Prokhorov's theorem gives a joint limit. The positive cone is closed, so $M \succeq 0$. For $\Phi \in C_c^\infty(\mathbb{R}^n; \mathbb{R}^n)$, truncate the matrix coordinate in the Stein identity and pass to the joint limit; the L^2 bound supplies uniform integrability. This gives

$$\mathbb{E}[X \cdot \Phi(X)] = \mathbb{E}[M : \nabla\Phi(X)].$$

Conditioning on X proves the Stein identity for (97) and preserves positivity. To obtain the mean identity, test the ith scalar Stein equation at level k with $x_j\chi(x/R)$, where χ is a smooth radial cutoff. The error is bounded by a constant times

$$\mathbb{E}\big[\|\tau_k(X_k)\|_{\mathrm{HS}}\,|X_k|\,\mathbf{1}_{\{|X_k|\geq R\}}\big]$$

and tends to zero as $R\to\infty$ by Cauchy–Schwarz. Hence $\mathbb{E}\tau_k(X_k)=\mathrm{Cov}(\mu_k)$. Uniform L^2 control of the matrix coordinate gives $\mathbb{E}M=\lim_k\mathbb{E}\tau_k(X_k)$, while W_2 convergence gives $\mathrm{Cov}(\mu_k)\to\mathrm{Cov}(\mu)$. Conditioning then proves (98). Conditional Jensen and lower semicontinuity give (99). For a fixed compactly supported smooth f, the energy integrand is continuous and grows at most linearly in the matrix coordinate; pass to the joint limit and then condition on X. Variances converge by weak convergence, proving the last assertion. □

In particular, if μ_k are regular log-concave approximations and τ_k are their moment-map kernels, a uniform L^2 estimate produces a positive Stein kernel of the limiting log-concave law and preserves Fathi's constant-one weighted Poincaré inequality. We call any such τ_∞ a *weak moment-map limit kernel*. This is a compactness statement for positive Stein kernels, complementary to coefficient and moment-measure stability estimates such as [12, 7]. It does not imply a general nonsmooth Burg theorem: a weak limit of graphs can retain extra randomness, the conditional barycenter can depend on the chosen subsequence, and inversion and log determinant are unstable under spectral collapse. Identifying these limits with a nonsmooth moment-map Hessian would require additional control of the inverse metrics.

9 Limitations and open questions

The calibration in the Burg theorem remains a genuine boundary hypothesis and must be verified for conditional convolution and localization kernels. Proposition 2.3 gives a canonical nonsmooth kernel for all one-dimensional log-concave laws and their affine products, but the general multidimensional approximation problem still lacks the inverse-metric control needed to identify different subsequential Hessian limits. The localization bridge requires a jointly measurable path of posterior kernels, calibration for almost every time, and directional envelopes or another uniform-integrability condition; mean Burg information alone cannot control an arbitrarily correlated adapted direction. Its canonical-gradient baseline is unavoidable even for Gaussian measure, and the theorem does not compare that baseline with the Euclidean Dirichlet form or bound the full score-residual lifetime. Open questions include nonsmooth moment-map stability, intrinsic propagation of calibration, and a history-sensitive conversion of the Burg–Parseval identity.

Disclosure on the use of generative AI

OpenAI Codex was used in preparing this research draft for literature searches, theorem exploration, algebraic checks, proof review, and copyediting. The author takes full responsibility for all results in the manuscript.

References

[1] H. J. Brascamp and E. H. Lieb, On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation, *J. Funct. Anal.* **22**(4) (1976), 366–389. https://doi.org/10.1016/0022-1236(76)90004-5.

[2] L. M. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, *USSR Comput. Math. Math. Phys.* **7**(3) (1967), 200–217. https://doi.org/10.1016/0041-5553(67)90040-7.

[3] D. Cordero-Erausquin and B. Klartag, Moment measures, *J. Funct. Anal.* **268**(12) (2015), 3834–3866. https://doi.org/10.1016/j.jfa.2015.04.001.

[4] T. A. Courtade, M. Fathi, and A. Pananjady, Existence of Stein kernels under a spectral gap, and discrepancy bounds, *Ann. Inst. H. Poincaré Probab. Statist.* **55**(2) (2019), 777–790. https://doi.org/10.1214/18-AIHP898.

[5] T. A. Courtade, Bounds on the Poincaré constant for convolution measures, *Ann. Inst. H. Poincaré Probab. Statist.* **56**(1) (2020), 566–579. https://doi.org/10.1214/19-AIHP973.

[6] T. Cui, X. Tong, and O. Zahm, Optimal Riemannian metric for Poincaré inequalities and how to ideally precondition Langevin dynamics, preprint (2024), revised 13 March 2025, arXiv:2404.02554v2. https://doi.org/10.48550/arXiv.2404.02554.

[7] A. Delalande and S. Farinelli, Regularized moment measures, *Potential Anal.* **64** (2026), article 40. https://doi.org/10.1007/s11118-025-10274-5.

[8] I. S. Dhillon and J. A. Tropp, Matrix nearness problems with Bregman divergences, *SIAM J. Matrix Anal. Appl.* **29**(4) (2007), 1120–1146. https://doi.org/10.1137/060649021.

[9] C. Döbler, On existence and uniqueness of univariate Stein kernels, *Electron. Commun. Probab.* **30** (2025). https://doi.org/10.1214/25-ECP707.

[10] R. Eldan, Thin shell implies spectral gap up to polylog via a stochastic localization scheme, *Geom. Funct. Anal.* **23**(2) (2013), 532–569. https://doi.org/10.1007/s00039-013-0214-y.

[11] M. Fathi, Stein kernels and moment maps, *Ann. Probab.* **47**(4) (2019), 2172–2185. https://doi.org/10.1214/18-AOP1305.

[12] M. Fathi and D. Mikulincer, Stability estimates for invariant measures of diffusion processes, with applications to stability of moment measures and Stein kernels, *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* **23**(3) (2022), 1417–1445. https://doi.org/10.2422/2036-2145.202011_016.

[13] T. T. Kadota, M. Zakai, and J. Ziv, Mutual information of the white Gaussian channel with and without feedback, *IEEE Trans. Inform. Theory* **17**(4) (1971), 368–371. https://doi.org/10.1109/TIT.1971.1054670.

[14] B. Klartag, Poincaré inequalities and moment maps, *Ann. Fac. Sci. Toulouse Math. (6)* **22**(1) (2013), 1–41. https://doi.org/10.5802/afst.1366.

[15] A. V. Kolesnikov, Hessian metrics, $CD(K,N)$-spaces, and optimal transportation of log-concave measures, *Discrete Contin. Dyn. Syst.* **34**(4) (2014), 1511–1532. https://doi.org/10.3934/dcds.2014.34.1511.

[16] M. Ledoux, I. Nourdin, and G. Peccati, Stein's method, logarithmic Sobolev and transport inequalities, *Geom. Funct. Anal.* **25**(1) (2015), 256–306. https://doi.org/10.1007/s00039-015-0312-0.

[17] T. Lelièvre, G. A. Pavliotis, G. Robin, R. Santet, and G. Stoltz, Optimizing the diffusion coefficient of overdamped Langevin dynamics, *Math. Comp.* **95**(360) (2026), 1829–1886. https://doi.org/10.1090/mcom/4098.

[18] G. Mijoule, M. Raič, G. Reinert, and Y. Swan, Stein's density method for multivariate continuous distributions, *Electron. J. Probab.* **28** (2023), paper 59, 1–40. https://doi.org/10.1214/22-EJP883.

[19] F. Santambrogio, Dealing with moment measures via entropy and optimal transport, *J. Funct. Anal.* **271**(2) (2016), 418–436. https://doi.org/10.1016/j.jfa.2016.04.009.

[20] A. Saumard, Weighted Poincaré inequalities, concentration inequalities and tail bounds related to Stein kernels in dimension one, *Bernoulli* **25**(4B) (2019), 3978–4006. https://doi.org/10.3150/19-BEJ1117.