Rigidity and Quantitative Stability of the Sliced Wasserstein Deficit

Bang-Xian Han*

May 23, 2026

Abstract

The sliced Wasserstein distance $SW_2(\mu, \nu)$ compares high-dimensional probability measures by averaging one-dimensional optimal transport distances over linear projections. Although sliced Wasserstein distances are now standard computational tools in statistics, imaging, and machine learning, the rigidity behind the elementary comparison

$$SW_2^2(\mu, \nu) \leq \frac{1}{d} W_2^2(\mu, \nu)$$

has not been systematically studied.

Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, $d \geq 2$, with $\mu \ll \mathcal{L}^d$, and define the sliced Wasserstein deficit by

$$\mathcal{D}(\mu, \nu) := \frac{1}{d} W_2^2(\mu, \nu) - SW_2^2(\mu, \nu) \geq 0.$$

We prove that $\mathcal{D}(\mu, \nu) = 0$ if and only if the Brenier map $T = \nabla\varphi$ from μ to ν is homothetic affine,

$$T(x) = \lambda x + b \qquad \mu\text{-a.e.},$$

for some $\lambda \geq 0$ and $b \in \mathbb{R}^d$.

For quantitative stability, we introduce the sliced Poincaré–Korn constant $\kappa_{\mathrm{SPK}}(\mu)$, defined as the spectral gap of an averaged ridge-projection quadratic form on gradient fields modulo the family $\{\lambda x + b\}$. Whenever this constant is positive, we prove a stability estimate for the sliced Wasserstein deficit, up to a one-dimensional Lipschitz scale for the projected monotone transports. We establish positive SPK bounds for the standard Gaussian, isotropic Gaussian measures, bounded perturbations of the Gaussian, and compact classes of gradient fields for fixed source measures.

Finally, we show that anisotropic Gaussians give a sharp obstruction: neither a Bakry–Émery lower curvature bound nor a usual Poincaré inequality alone can imply a global sliced Poincaré–Korn inequality.

Keywords: sliced Wasserstein distance; optimal transport; rigidity; stability; spectral gap; sliced Poincaré–Korn inequality

MSC 2020: 49Q22, 52A40, 60E15

Contents

*School of Mathematics, Shandong University, Jinan, 250100, China. Email: hanbx@sdu.edu.cn.

1 Introduction

1.1 Background

For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, the quadratic Wasserstein distance (see, e.g., [30, 31]) is defined by

$$W_2^2(\mu, \nu) := \inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \, \mathrm{d}\gamma(x, y),$$

where $\Pi(\mu, \nu)$ denotes the set of transport plans. For $\theta \in \mathbb{S}^{d-1}$, write

$$\pi_\theta(x) = \theta \cdot x, \qquad \mu_\theta = (\pi_\theta)_\# \mu.$$

The sliced Wasserstein distance (see [1]) is defined by

$$SW_2^2(\mu, \nu) := \int_{\mathbb{S}^{d-1}} W_2^2(\mu_\theta, \nu_\theta) \, \mathrm{d}\sigma(\theta),$$

where σ is the normalized surface measure on $\mathbb{S}^{d-1}$.

Note that

$$\int_{\mathbb{S}^{d-1}} \theta \otimes \theta \, \mathrm{d}\sigma(\theta) = \frac{1}{d} I_d, \qquad \int_{\mathbb{S}^{d-1}} |\theta \cdot z|^2 \, \mathrm{d}\sigma(\theta) = \frac{1}{d}|z|^2.$$

For every coupling $\gamma \in \Pi(\mu, \nu)$,

$$W_2^2(\mu_\theta, \nu_\theta) \le \int |\theta \cdot (x - y)|^2 \, \mathrm{d}\gamma(x, y).$$

Averaging in θ and optimizing over γ gives the elementary inequality

$$\mathcal{D}(\mu, \nu) := \frac{1}{d} W_2^2(\mu, \nu) - SW_2^2(\mu, \nu) \ge 0, \tag{SWD}$$

and we call $\mathcal{D}(\mu, \nu)$ the *sliced Wasserstein deficit.*

If $\mu \ll \mathcal{L}^d$ and $T = \nabla\varphi$ is the Brenier map from μ to ν, then

$$\mathcal{D}(\mu,\nu) = \int_{\mathbb{S}^{d-1}} \left[\int_{\mathbb{R}^d} |\theta \cdot (T(x) - x)|^2 \, \mathrm{d}\mu(x) - W_2^2(\mu_\theta, \nu_\theta) \right] \mathrm{d}\sigma(\theta). \tag{1.2}$$

Thus $\mathcal{D}$ is an average of non-negative one-dimensional optimality gaps.

Sliced Wasserstein distances were introduced and developed as computationally tractable variants of optimal transport by reducing the comparison of high-dimensional measures to one-dimensional projections. Early mathematical and computational developments include the work of Rabin–Peyré–Delon–Bernot on barycenters and texture mixing [29], Bonnotte's thesis on unidimensional and evolution methods [2], and the Radon/sliced barycenter framework of Bonneel–Rabin–Peyré–Pfister [1]. The distance and its variants have since become standard tools in imaging, statistics, and machine learning; for example, in large-scale generative modelling and domain adaptation [20, 25, 26]. Generalized, max, distributional, and tree-sliced variants were introduced to improve expressivity while retaining one-dimensional computability; see [21, 27, 22] and references therein.

The present paper is concerned not with computation but with the rigidity and stability theory behind the elementary inequality (SWD)–asking whether approximate equality in every slice forces the high-dimensional Brenier map to be globally affine—was previously unknown except in the trivial one-dimensional case. This places the paper in the broader tradition of deficit and quantitative-stability problems in optimal transport and geometric analysis. Related examples include quantitative stability for transport maps, stability in functional inequalities such as Talagrand or logarithmic Sobolev inequalities, mass-transport approaches to quantitative isoperimetry, and rigidity under curvature lower bounds; see [13, 11, 7, 24, 9, 4].

1.2 Main results

The first theorem identifies equality cases in (SWD).

Theorem 1.1 (Rigidity)**.** *Let $d \geq 2$, let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, assume $\mu \ll \mathcal{L}^d$, and let $T = \nabla\varphi$ be the Brenier map from μ to ν. Then*

$$\mathcal{D}(\mu,\nu) = 0 \quad \Longleftrightarrow \quad T(x) = \lambda x + b \quad \mu\text{-a.e.}$$

for some $\lambda \geq 0$ and $b \in \mathbb{R}^d$.

The proof is given in Section 2. The argument does not assume smoothness of T and avoids differentiating the Brenier potential. Zero deficit says that the projected plan

$$(\theta \cdot X, \theta \cdot T(X))$$

is one-dimensionally optimal for almost every direction. One-dimensional cyclical monotonicity then implies directional monotonicity for almost every θ. A simple geometric lemma forces every chord $T(x) - T(y)$ to be a non-negative multiple of $x - y$, and a three-point argument makes this multiple constant.

For stability we introduce the ridge defect

$$\mathcal{R}_\mu(u) := \int_{\mathbb{S}^{d-1}} \mathbb{E}_\mu\big[\mathrm{Var}(\theta \cdot u(X) \mid \theta \cdot X)\big] \, \mathrm{d}\sigma(\theta) \tag{1.3}$$

for vector fields $u \in L^2(\mu; \mathbb{R}^d)$. Equivalently,

$$\mathcal{R}_\mu(u) = \int_{\mathbb{S}^{d-1}} \inf_{h \in L^2(\mu_\theta)} \int_{\mathbb{R}^d} |\theta \cdot u(x) - h(\theta \cdot x)|^2 \, \mathrm{d}\mu(x) \, \mathrm{d}\sigma(\theta). \tag{1.4}$$

Let

$$\mathcal{A}_d := \{x \mapsto \lambda x + b : \lambda \in \mathbb{R},\ b \in \mathbb{R}^d\}.$$

Here and below, following Otto's calculus, define the the tangent space at $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ by

$$\mathrm{Tan}_\mu := \overline{\{\nabla\varphi : \varphi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\mu;\mathbb{R}^d)}.$$

The sliced Poincaré–Korn constant is

$$\kappa_{\mathrm{SPK}}(\mu) := \inf_{u \in \mathrm{Tan}_\mu} \frac{\mathcal{R}_\mu(u)}{\mathrm{dist}^2_{L^2(\mu)}(u, \mathcal{A}_d)}, \tag{1.5}$$

where terms with zero denominator are omitted. The normalized constant is

$$\overline{\kappa}_{\mathrm{SPK}}(\mu) := d\, \kappa_{\mathrm{SPK}}(\mu).$$

Our main quantitative theorem under the SPK hypothesis is the following.

Theorem 1.2 (Stability under SPK)**.** *Let $\mu \ll \mathcal{L}^d$, let $T = \nabla\varphi$ be the Brenier map from μ to ν, assume $T \in \mathrm{Tan}_\mu$, and for σ-a.e. θ let τ_θ be the monotone transport map from μ_θ to ν_θ. Suppose*

$$\Lambda := \operatorname*{ess\,sup}_{\theta \in \mathbb{S}^{d-1}} \mathrm{Lip}(\tau_\theta) < \infty.$$

If $\kappa_{\mathrm{SPK}}(\mu) > 0$, then

$$\mathrm{dist}^2_{L^2(\mu)}(T, \mathcal{A}_d) \le \frac{\Lambda}{\kappa_{\mathrm{SPK}}(\mu)}\, \mathcal{D}(\mu, \nu). \tag{1.6}$$

Equivalently,

$$\mathrm{dist}^2_{L^2(\mu)}(T, \mathcal{A}_d) \le \frac{d\Lambda}{\overline{\kappa}_{\mathrm{SPK}}(\mu)}\, \mathcal{D}(\mu, \nu).$$

The proof uses a one-dimensional Fenchel-gap inequality in Section 3. The factor Λ is a scale parameter and cannot be removed in this form: replacing a non-affine Brenier map by a large multiple multiplies the distance to $\mathcal{A}_d$ quadratically, while the one-dimensional Fenchel gaps scale only linearly.

Remark 1.3 (On the assumption $T \in \mathrm{Tan}_\mu$ in Theorem 1.2). For a general Brenier map $T = \nabla\phi$ with $\mu \ll \mathcal{L}^d$, it is not automatic that $T \in \mathrm{Tan}_\mu$; this requires that the convex potential ϕ can be approximated by smooth functions in the weighted Sobolev norm $\|\nabla(\phi - \psi)\|_{L^2(\mu)} \to 0$.

The assumption is satisfied in most natural settings: if $\mu = \rho\mathcal{L}^d$ with ρ locally bounded away from zero on a connected open set, or if μ is log-concave with full-dimensional support, then C_c^∞-gradients are dense in the space of $L^2(\mu)$ gradient fields, and consequently every Brenier map belongs to Tan_μ. On the other hand, if μ is supported on a disconnected set or if its density degenerates on large sets, Tan_μ may be strictly smaller than the space of all $L^2(\mu)$ gradient fields, and the theorem should then be applied only to those Brenier maps that happen to lie in the closure. Thus the hypothesis $T \in \mathrm{Tan}_\mu$ is a mild regularity condition on the source measure rather than on the map itself.

The Gaussian as the model case. In probability theory the standard Gaussian measure serves as the canonical setting in which to test conjectures and develop tools: it is the stage for the central limit theorem, for the Poincaré and logarithmic Sobolev inequalities, and for the Wiener–Itô chaos decomposition. In optimal transport, Caffarelli's contraction theorem and the Gaussian isoperimetric inequality provide archetypal rigidity and stability results. It is therefore natural that the first complete quantitative stability theorem for the sliced Wasserstein deficit should be established for the Gaussian source γ_d.

Theorem 1.4 (Gaussian sliced Poincaré–Korn inequality)**.** *For the standard Gaussian measure $\gamma_d = N(0, I_d)$, $d \geq 2$,*

$$\mathcal{R}_{\gamma_d}(u) \geq \frac{d-1}{d(d+2)} \operatorname{dist}^2_{L^2(\gamma_d)}(u, \mathcal{A}_d) \tag{1.7}$$

for every gradient field $u \in \mathrm{Tan}_{\gamma_d}$. Hence

$$\overline{\kappa}_{\mathrm{SPK}}(\gamma_d) \geq \frac{d-1}{d+2}.$$

Combining Theorem 1.4 with Theorem 1.2 yields Gaussian stability:

$$\operatorname{dist}^2_{L^2(\gamma_d)}(T, \mathcal{A}_d) \leq \frac{d(d+2)}{d-1} \Lambda \, \mathcal{D}(\gamma_d, \nu). \tag{1.8}$$

The proof of Theorem 1.4 is based on the Hermite decomposition and a Funk–Hecke tensor estimate. For affine gradient maps we compute the exact constant; see Section 4.

Finally, we show that isotropic normalization is essential.

Proposition 1.5 (Anisotropic Gaussian obstruction)**.** *There exists a family of centered Gaussian measures μ_ε on $\mathbb{R}^2$ satisfying a uniform Bakry–Émery condition* $\mathrm{BE}(1, \infty)$ *and a uniform Poincaré inequality, but*

$$\kappa_{\mathrm{SPK}}(\mu_\varepsilon) \to 0 \qquad \text{as} \ \ \varepsilon \downarrow 0.$$

Consequently no lower bound for κ_{SPK} can depend only on a Bakry–Émery lower curvature bound or on the usual Poincaré constant.

Summary This paper consists of the following four main contributions.

(1) We prove a zero-deficit rigidity theorem for general absolutely continuous source measures. The proof uses only one-dimensional cyclical monotonicity and a geometric three-point argument, hence it does not require smoothness of the Brenier map. We also prove a higher-dimensional version of this rigidity by a different method.

(2) We introduce the sliced Poincaré–Korn inequality (constant) and show that any positive lower bound for this constant immediately implies quantitative stability for the sliced Wasserstein deficit, up to a one-dimensional regularity factor.

(3) We prove the SPK inequality for Gaussian measures. This yields a complete stability theorem for Gaussian measures (and their bounded perturbations).

(4) We show by an explicit anisotropic Gaussian measure that neither a Bakry–Émery lower curvature bound nor the usual Poincaré constant can imply a universal sliced Poincaré–Korn bound without an isotropic or comparable non-degeneracy normalization.

Further research The anisotropic Gaussian counterexample in Section 5.3 shows that isotropicity, or at least a comparable non-degeneracy normalization, is essential. It is natural to ask whether isotropic log-concave measures satisfy a dimension-free sliced Poincaré–Korn inequality. This question is philosophically close to the KLS and slicing problems; see the localization lemma of Kannan–Lovász–Simonovits [18], Eldan's stochastic localization [10], Lee–Vempala's work on KLS via stochastic localization [23], and the recent resolution of Bourgain's slicing problem by Klartag–Lehec [19]. At present, plausible first targets include unconditional measures, product measures, uniformly log-concave measures with covariance normalization, and heat-flow regularizations. Ciosmak's leaf decomposition [6] may also be relevant.

Organization Section 2 proves the zero-deficit rigidity theorems, including a higher-dimensional generalization. Section 3 introduces the sliced Poincaré–Korn constant and proves stability under the SPK hypothesis, including a discussion of the Lipschitz parameter Λ. Section 4 proves the Gaussian SPK inequality and Gaussian stability. Section 5 gives further examples: bounded Gaussian perturbations, compact classes for fixed measures, and the anisotropic Gaussian counterexample. Appendix contains the linear algebra computations.

2 Rigidity of the Sliced Wasserstein Deficit

In this section we prove Theorem 1.1. The proof is deliberately written without differentiability assumptions on the Brenier map. In Theorem 2.5 we will use a different strategy to deal with the higher-dimensional problem.

2.1 Deficit decomposition

Let $\mu \ll \mathcal{L}^d$ and let $T = \nabla\varphi$ be the Brenier map from μ to ν. For every direction $\theta \in \mathbb{S}^{d-1}$ define

$$g_\theta(T) := \int |\theta \cdot (T(x) - x)|^2 \, \mathrm{d}\mu(x) - W_2^2(\mu_\theta, \nu_\theta).$$

Then $g_\theta(T) \geq 0$ and

$$\mathcal{D}(\mu, \nu) = \int_{\mathbb{S}^{d-1}} g_\theta(T) \, \mathrm{d}\sigma(\theta). \tag{2.1}$$

Indeed, the projected coupling $(\theta \cdot X, \theta \cdot T(X))$ is an admissible coupling between μ_θ and ν_θ, so $g_\theta \geq 0$. Averaging in θ gives

$$\int_{\mathbb{S}^{d-1}} \int |\theta \cdot (T(x) - x)|^2 \, \mathrm{d}\mu(x) \, \mathrm{d}\sigma(\theta) = \frac{1}{d} \int |T(x) - x|^2 \, \mathrm{d}\mu(x) = \frac{1}{d} W_2^2(\mu, \nu),$$

which proves (2.1).

Lemma 2.1 (One-dimensional monotonicity)**.** *Let $\alpha, \beta \in \mathcal{P}_2(\mathbb{R})$ and let $\eta \in \Pi(\alpha, \beta)$ be optimal for the cost $|s - t|^2$. Then*

$$(s - s')(t - t') \geq 0$$

for $\eta \otimes \eta$-a.e. $((s, t), (s', t'))$.

Proof. By cyclical monotonicity,

$$|s - t|^2 + |s' - t'|^2 \leq |s - t'|^2 + |s' - t|^2$$

for $\eta \otimes \eta$-a.e. pairs. Expanding gives

$$2(s - s')(t - t') \geq 0.$$

□

Lemma 2.2 (Directional signs force parallelism)**.** *Let $a, b \in \mathbb{R}^d$ with $a \neq 0$. If*

$$(\theta \cdot a)(\theta \cdot b) \geq 0$$

for σ-a.e. $\theta \in \mathbb{S}^{d-1}$, then $b = \lambda a$ for some $\lambda \geq 0$.

Proof. If a and b are not collinear, the two hemispheres $\{\theta : \theta \cdot a > 0\}$ and $\{\theta : \theta \cdot b < 0\}$ have an intersection of positive surface measure. On this intersection the product is negative. Thus $b = \lambda a$ for some $\lambda \in \mathbb{R}$. The sign condition forces $\lambda \geq 0$. □

Lemma 2.3 (Constancy from three points)**.** *Let $E \subset \mathbb{R}^d$ with $d \geq 2$, let m be a Borel measure on E with $m \ll \mathcal{L}^d$, and let $T : E \to \mathbb{R}^d$ be measurable. Suppose that for $m \otimes m$-a.e. pair (x, y),*

$$T(x) - T(y) = \lambda(x, y)(x - y)$$

for some $\lambda(x, y) \geq 0$. Then $\lambda(x, y) = \lambda_0$ for a constant $\lambda_0 \geq 0$ and almost every pair (x, y).

Proof. For $m^{\otimes 3}$-a.e. triple (x, y, z), the vectors $x - y$ and $y - z$ are linearly independent. For such a triple we have

$$T(x) - T(y) = \lambda_{xy}(x - y), \quad T(y) - T(z) = \lambda_{yz}(y - z),$$

and

$$T(x) - T(z) = \lambda_{xz}(x - z) = \lambda_{xz}(x - y) + \lambda_{xz}(y - z).$$

Since also $T(x) - T(z) = (T(x) - T(y)) + (T(y) - T(z))$, linear independence gives

$$\lambda_{xy} = \lambda_{yz} = \lambda_{xz}$$

for almost every triple. Fubini then implies that $\lambda(x, y)$ is a.e. constant: for a.e. y, the equality $\lambda(x, y) = \lambda(y, z)$ holds for almost every (x, z), so both sides must equal a number depending only on y; symmetry removes the dependence on y. □

Proof of Theorem 1.1. Assume first $\mathcal{D}(\mu, \nu) = 0$. Since $g_\theta \geq 0$ and $\int g_\theta \, d\sigma = 0$, we have $g_\theta = 0$ for σ-a.e. θ. Hence, for almost every θ, the projected plan

$$(\theta \cdot X, \theta \cdot T(X))$$

is an optimal coupling between μ_θ and ν_θ. Lemma 2.1 yields

$$(\theta \cdot (x - y))(\theta \cdot (T(x) - T(y))) \geq 0 \tag{2.2}$$

for $\mu \otimes \mu$-a.e. (x, y) and σ-a.e. θ. By Fubini, for $\mu \otimes \mu$-a.e. pair (x, y) the sign condition (2.2) holds for σ-a.e. θ. Since $\mu \ll \mathcal{L}^d$, $x \neq y$ for almost every pair. Lemma 2.2 gives

$$T(x) - T(y) = \lambda(x, y)(x - y), \qquad \lambda(x, y) \geq 0.$$

Lemma 2.3 yields $\lambda(x, y) = \lambda_0$ for almost every pair. Fixing one point by Fubini gives $T(x) = \lambda_0 x + b$ for μ-a.e. x.

Conversely, suppose $T(x) = \lambda x + b$ with $\lambda \geq 0$. Then for every θ,

$$\theta \cdot T(x) = \lambda(\theta \cdot x) + \theta \cdot b$$

is the one-dimensional monotone transport from μ_θ to ν_θ. Therefore $g_\theta = 0$ for every θ and $\mathcal{D}(\mu, \nu) = 0$. □

Remark 2.4 (Relation with the ridge formulation)*.* Zero deficit also implies a ridge factorization

$$\theta \cdot T(x) = \tau_\theta(\theta \cdot x)$$

for almost every direction if the one-dimensional optimal map is unique. If $T = \nabla\varphi$ were smooth, differentiating along directions $v \perp \theta$ would give

$$v^T D^2\varphi(x)\theta = 0.$$

For almost every θ this forces $D^2\varphi(x)$ to be scalar, hence $T(x) = \lambda x + b$. The proof above is a non-smooth version of the same compatibility principle, expressed through monotone couplings rather than derivatives.

2.2 Higher-dimensional rigidity

The one-dimensional rigidity theorem has a higher-dimensional analogue. We record it separately because its proof reveals another useful mechanism: the ridge factorization of projected Brenier maps forces the Hessian measure of the convex potential to be scalar. This argument is closer to the differential-geometric intuition behind the equality case and will also be useful when considering higher-dimensional sliced distances.

Let $G_{d,k}$ be the Grassmannian of k-dimensional linear subspaces of $\mathbb{R}^d$, let $\pi_{d,k}$ be its Haar probability measure, and let P_E denote the orthogonal projection onto E. Define

$$SW_{2,k}^2(\mu,\nu) := \int_{G_{d,k}} W_2^2((P_E)_\#\mu,(P_E)_\#\nu)\ \mathrm{d}\pi_{d,k}(E).$$

Since

$$\int_{G_{d,k}} |P_E z|^2\ \mathrm{d}\pi_{d,k}(E) = \frac{k}{d}|z|^2,$$

we have

$$SW_{2,k}^2(\mu,\nu) \le \frac{k}{d} W_2^2(\mu,\nu),$$

and we define the Grassmannian deficit

$$\mathcal{D}_k(\mu,\nu) := \frac{k}{d} W_2^2(\mu,\nu) - SW_{2,k}^2(\mu,\nu). \tag{2.3}$$

For $k = 1$ this is the original deficit $\mathcal{D}$.

Theorem 2.5 (Grassmannian rigidity)**.** *Let $1 \le k \le d-1$. Let $\mu = \rho\mathcal{L}^d$ be concentrated on a connected open set $\Omega \subset \mathbb{R}^d$, with $\rho > 0$ a.e. on Ω, and let $T = \nabla\varphi$ be the Brenier map from μ to ν. If*

$$\mathcal{D}_k(\mu,\nu) = 0,$$

then

$$T(x) = \lambda x + b \qquad \mu\text{-a.e.}$$

for some $\lambda \ge 0$ and $b \in \mathbb{R}^d$. Conversely, every map of this form has zero k-dimensional deficit.

Lemma 2.6 (Linear algebra on the Grassmannian)**.** *Let M be a symmetric $d \times d$ matrix and let $1 \le k \le d-1$. If*

$$P_E M (I - P_E) = 0$$

for $\pi_{d,k}$-a.e. $E \in G_{d,k}$, then $M = \alpha I$ for some $\alpha \in \mathbb{R}$. Moreover,

$$\int_{G_{d,k}} \|P_E M (I - P_E)\|_{\mathrm{HS}}^2\ \mathrm{d}\pi_{d,k}(E) = \frac{k(d-k)}{(d-1)(d+2)} \left\| M - \frac{\operatorname{Tr} M}{d} I \right\|_{\mathrm{HS}}^2.$$

Proof. This is proved in Appendix B. ☐

Proof of Theorem 2.5. Let $X \sim \mu$. For $E \in G_{d,k}$ put

$$g_E(T) := \mathbb{E}|P_E(T(X) - X)|^2 - W_2^2((P_E)_\#\mu,(P_E)_\#\nu).$$

Then $g_E(T) \ge 0$ and

$$\mathcal{D}_k(\mu,\nu) = \int_{G_{d,k}} g_E(T)\ \mathrm{d}\pi_{d,k}(E).$$

Hence $\mathcal{D}_k(\mu,\nu) = 0$ implies $g_E(T) = 0$ for a.e. E. For such E, the coupling

$$(P_E X, P_E T(X))$$

is an optimal quadratic coupling between $(P_E)_{\#}\mu$ and $(P_E)_{\#}\nu$. Since $(P_E)_{\#}\mu$ is absolutely continuous on E, Brenier's theorem gives a convex function $f_E : E \to \mathbb{R}$ such that

$$P_E T(x) = \nabla_E f_E(P_E x) \qquad \mu\text{-a.e. } x. \tag{2.4}$$

Because $\rho > 0$ a.e. on Ω, this identity also holds for Lebesgue-a.e. $x \in \Omega$.

Now use the convex potential. Since φ is convex, by Alexandrov's theorem, its distributional Hessian $D^2\varphi$ is a symmetric positive semidefinite matrix-valued Radon measure on Ω. Fix such a space E and take $v \in E$, $w \in E^{\perp}$. From (2.4), the scalar function $v \cdot T(x) = v \cdot \nabla\varphi(x)$ depends only on $P_E x$. Hence its derivative in the direction w vanishes in distributions:

$$\partial_w(v \cdot \nabla\varphi) = 0.$$

Equivalently,

$$v^T D^2\varphi\, w = 0 \qquad \text{as a signed Radon measure.} \tag{2.5}$$

Thus

$$P_E D^2\varphi(I - P_E) = 0$$

as a matrix-valued measure for a.e. E.

Let $\mathfrak{m} := \operatorname{Tr}(D^2\varphi)$ be a Radon measure and write the polar decomposition

$$D^2\varphi = M\mathfrak{m},$$

where $M(x)$ is symmetric positive semidefinite and $\operatorname{Tr} M(x) = 1$ for $\mathfrak{m}$-a.e. x. By Fubini, (2.5) implies that for $\mathfrak{m}$-a.e. x,

$$P_E M(x)(I - P_E) = 0$$

for $\pi_{d,k}$-a.e. E. Lemma 2.6 gives $M(x) = I/d$ for $\mathfrak{m}$-a.e. x. Consequently

$$D^2\varphi = \omega I$$

for the scalar Radon measure $\omega := \mathfrak{m}/d$.

It remains to show that ω is a constant multiple of Lebesgue measure. In distributions,

$$\partial_{ij}\varphi = 0 \quad (i \neq j), \qquad \partial_{11}\varphi = \cdots = \partial_{dd}\varphi = \omega.$$

Fix j and choose $i \neq j$. Then in distributions,

$$\partial_j\omega = \partial_j\partial_{ii}\varphi = \partial_i\partial_{ji}\varphi = 0.$$

Thus $\nabla\omega = 0$ in $\mathcal{D}'(\Omega)$. Since Ω is connected, $\omega = \lambda\mathcal{L}^d$ for some constant $\lambda \geq 0$. Hence

$$D^2\varphi = \lambda I$$

in distributions on Ω, and therefore

$$\varphi(x) = \frac{\lambda}{2}|x|^2 + b \cdot x + c$$

on Ω. Thus $T = \nabla\varphi = \lambda x + b$ μ-a.e.

Conversely, if $T(x) = \lambda x + b$ with $\lambda \geq 0$, then for every E

$$P_E T(x) = \lambda P_E x + P_E b$$

is the Brenier map from $(P_E)_{\#}\mu$ to $(P_E)_{\#}\nu$. Hence $g_E(T) = 0$ for every E and $\mathcal{D}_k(\mu, \nu) = 0$. $\square$

Remark 2.7. The proof for $k = 1$ recovers the ridge-differentiation heuristic in Section 2: the identity $\theta \cdot \nabla\varphi = h_\theta(\theta \cdot x)$ says that the Hessian measure has no mixed block between $\mathbb{R}\theta$ and $\theta^\perp$. The Grassmannian formulation shows that this is not a peculiarity of one-dimensional slicing; any family of projected optimalities on $G_{d,k}$, $1 \le k \le d-1$, forces the same scalar Hessian measure.

Remark 2.8 (Assumptions in the Grassmannian rigidity theorem). Theorem 2.5 requires μ to be concentrated on a connected open set Ω with $\rho > 0$ a.e., which is stronger than the sole absolute continuity $\mu \ll \mathcal{L}^d$ used in Theorem 1.1. The extra regularity is needed because the proof of Theorem 2.5 operates on the Hessian measure $D^2\phi$ of the Brenier potential. This relies on Alexandrov's theorem and on the ability to differentiate the convex potential in the sense of distributions on Ω, which necessitates both connectedness and the non-degeneracy of ρ. By contrast, Theorem 1.1 uses only one-dimensional cyclical monotonicity and a purely geometric three-point argument, and therefore does not require any differential structure.

3 Sliced Poincaré–Korn constant

3.1 Definition and spectral interpretation

Let $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and let

$$\mathrm{Tan}_\mu := \overline{\{\nabla\varphi : \varphi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\mu;\mathbb{R}^d)}.$$

For $u \in \mathrm{Tan}_\mu$ define the ridge defect by (1.3). The sliced Poincaré–Korn constant (or sliced Poincaré–Korn spectral gap) is defined as

$$\kappa_{\mathrm{SPK}}(\mu) := \inf_{\substack{u \in \mathrm{Tan}_\mu \\ \mathrm{dist}_{L^2(\mu)}(u,\mathcal{A}_d) > 0}} \frac{\mathcal{R}_\mu(u)}{\mathrm{dist}^2_{L^2(\mu)}(u,\mathcal{A}_d)}.$$

Because the average over directions carries a natural factor $1/d$, the normalized quantity

$$\overline{\kappa}_{\mathrm{SPK}}(\mu) = d\,\kappa_{\mathrm{SPK}}(\mu)$$

is the quantity expected to be dimension-free in isotropic log-concave classes. For example, the Gaussian stability Theorem 1.4 gives

$$\overline{\kappa}_{\mathrm{SPK}}(\gamma_d) \ge \frac{d-1}{d+2}.$$

Equivalently, we say that μ satisfies sliced Poincaré–Korn inequality $\mathrm{SPK}(\kappa)$ if

$$\mathrm{dist}^2_{L^2(\mu)}(u,\mathcal{A}_d) \le \frac{1}{\kappa}\mathcal{R}_\mu(u) \qquad \forall u \in \mathrm{Tan}_\mu. \tag{SPK}$$

The terminology "sliced Poincaré–Korn" is motivated by the analogy with Korn-type inequalities in elasticity and classical Poincaré inequality. Classical Korn inequalities control vector fields modulo rigid motions by the symmetric part of the gradient, and play a fundamental role in linearized elasticity; see, for instance, Ciarlet's text [5] and Dacorogna's direct-methods treatment [8]. Nonlinear geometric rigidity estimates, such as the theorem of Friesecke–James–Müller [14], provide a modern quantitative version modulo rotations. In our setting the null space is not the Euclidean rigid-motion space but the homothetic affine family

$$\mathcal{A}_d = \{\lambda x + b\},$$

and the "strain" is replaced by an averaged conditional-variance defect over one-dimensional projections. Thus the sliced Poincaré–Korn constant should be viewed as a new spectral gap, rather than as a direct consequence of a combination of Korn inequality and Poincaré inequality.

The following kernel lemma identifies the null space of the ridge defect on weak gradient fields. It is placed here because it explains why the quotient by $\mathcal{A}_d$ is the natural one in the SPK inequality; it will also be used in the compactness criterion in Section 5.

Lemma 3.1 (Kernel of the ridge defect on weak gradient fields)**.** *Let $\Omega \subset \mathbb{R}^d$ be connected and open, and let $\mu = \rho\mathcal{L}^d$ be a probability measure with $\mu(\Omega) = 1$. Assume that ρ is locally bounded below on Ω. Let $v \in L^2(\mu; \mathbb{R}^d)$ be a distributional gradient on Ω. If the ridge defect $\mathcal{R}_\mu(v) = 0$, then $v(x) = \lambda x + b$ for μ-a.e. $x \in \Omega$.*

Proof. Since ρ is locally bounded below, $v \in L^1_{\mathrm{loc}}(\Omega)$. The identity $\mathcal{R}_\mu(v) = 0$ means that for σ-a.e. θ there exists a function h_θ such that

$$\theta \cdot v(x) = h_\theta(\theta \cdot x) \qquad \text{for } \mu\text{-a.e. } x.$$

Thus, for every $w \perp \theta$,

$$\partial_w(\theta \cdot v) = 0 \qquad \text{in } \mathcal{D}'(\Omega).$$

Because v is a distributional gradient, its distributional derivative Dv is symmetric. Hence

$$(I - \theta \otimes \theta) Dv\, \theta = 0 \qquad \text{in } \mathcal{D}'(\Omega)$$

for almost every θ. Testing against smooth compactly supported functions and using the elementary linear-algebra fact that a symmetric matrix M satisfying $(I - \theta \otimes \theta)M\theta = 0$ for a.e. θ must be a scalar multiple of the identity, we obtain

$$Dv = \omega I$$

for some scalar distribution ω. The off-diagonal equations give $\partial_i v_j = 0$ for $i \neq j$, while the diagonal equations give $\partial_1 v_1 = \cdots = \partial_d v_d = \omega$. Taking mixed derivatives yields $\partial_i \omega = 0$ for every i. Since Ω is connected, $\omega = \lambda$ is constant. Therefore $D(v - \lambda x) = 0$, and $v = \lambda x + b$ on Ω in the sense of distributions, hence μ-a.e. □

3.2 Fenchel-gap stability in one dimension

The following lemma is fundamental in the proof of our stability theorem.

Lemma 3.2 (One-dimensional Fenchel gap)**.** *Let A, B be real-valued random variables in L^2, and let τ be the monotone transport map from the law of A to the law of B. If* $\mathrm{Lip}(\tau) \leq \Lambda$, *then*

$$\mathbb{E}|B - \tau(A)|^2 \leq \Lambda \left(\mathbb{E}|A - B|^2 - W_2^2(\mathrm{Law}(A), \mathrm{Law}(B)) \right). \tag{3.2}$$

Proof. Let f be a convex potential such that $f' = \tau$ a.e. The one-dimensional Kantorovich duality for the quadratic cost gives

$$\mathbb{E}|A - B|^2 - W_2^2(\mathrm{Law}(A), \mathrm{Law}(B)) = 2\mathbb{E}\big[f(A) + f^*(B) - AB\big].$$

The expression in brackets is the Fenchel gap. Since f' is Λ-Lipschitz and monotone, f^* is $1/\Lambda$-strongly convex on the relevant interval (with the usual generalized interpretation if f' is not strictly increasing). The Fenchel gap therefore controls the squared distance to the subgradient relation:

$$f(A) + f^*(B) - AB \geq \frac{1}{2\Lambda}|B - f'(A)|^2 = \frac{1}{2\Lambda}|B - \tau(A)|^2.$$

Taking expectations proves (3.2). □

3.3 Stability under SPK

We now prove Theorem 1.2.

Proof of Theorem 1.2. For each direction θ, applying the lemma with

$$A = \theta \cdot X, \qquad B = \theta \cdot T(X)$$

gives

$$\int |\theta \cdot T(x) - \tau_\theta(\theta \cdot x)|^2 \, \mathrm{d}\mu(x) \le \Lambda \, g_\theta(T) \tag{3.3}$$

whenever $\mathrm{Lip}(\tau_\theta) \le \Lambda$. Since $h = \tau_\theta$ is an admissible ridge function,

$$\inf_h \int |\theta \cdot T(x) - h(\theta \cdot x)|^2 \, \mathrm{d}\mu(x) \le \Lambda g_\theta(T).$$

Integrating in θ yields

$$\mathcal{R}_\mu(T) \le \Lambda \mathcal{D}(\mu, \nu).$$

Applying (SPK) to $u = T$ gives

$$\mathrm{dist}^2_{L^2(\mu)}(T, \mathcal{A}_d) \le \frac{1}{\kappa_{\mathrm{SPK}}(\mu)} \mathcal{R}_\mu(T) \le \frac{\Lambda}{\kappa_{\mathrm{SPK}}(\mu)} \mathcal{D}(\mu, \nu).$$

□

Remark 3.3 (Why SPK is the right key). The first step, from the Wasserstein deficit to the ridge defect, is one-dimensional and robust. All high-dimensional geometry is contained in the lower bound for $\mathcal{R}_\mu$ modulo $\mathcal{A}_d$. Thus the stability problem splits cleanly into an optimal-transport input and a geometric spectral input.

3.4 On the one-dimensional Lipschitz parameter

The parameter

$$\Lambda = \operatorname*{ess\,sup}_{\theta \in \mathbb{S}^{d-1}} \mathrm{Lip}(\tau_\theta)$$

should be regarded as a one-dimensional regularity scale for the target marginals. It is not part of the high-dimensional SPK geometry. Lemma 3.2 uses it only to turn the Fenchel gap into an L^2 distance from the monotone graph.

In concrete situations Λ can often be bounded independently of the dimension. For instance, suppose the source is the standard Gaussian and the target has density e^{-W} with

$$\nabla^2 W \ge \alpha I_d$$

for some $\alpha > 0$. By the Prékopa theorem, each one-dimensional marginal ν_θ is again α-strongly log-concave on $\mathbb{R}$. Caffarelli's contraction theorem [3, 12] in dimension one then implies that the monotone map from $N(0,1)$ to ν_θ satisfies

$$\mathrm{Lip}(\tau_\theta) \le \alpha^{-1/2} \qquad \text{for every } \theta. \tag{3.4}$$

Thus in the Gaussian-source case, uniformly log-concave targets have a universal projected Lipschitz scale. In particular, if $\alpha \ge 1$, all projected rearrangements are contractions.

There is also a direct one-dimensional density criterion. If p_θ and q_θ are the densities of μ_θ and ν_θ, and the monotone map τ_θ is differentiable, then

$$\tau_\theta'(s) = \frac{p_\theta(s)}{q_\theta(\tau_\theta(s))}.$$

Hence a uniform lower bound on the target projected densities along the monotone image, relative to the source projected densities, gives a bound for Λ. The uniform Lipschitz assumption in Theorem 1.2 may therefore be replaced by any condition that gives the strong convexity of the one-dimensional dual potentials required in Lemma 3.2. We keep the Lipschitz form because it is simple and transparent.

4 Stability for Gaussian measures

Let $\gamma_d = N(0, I_d)$ be the standard Gaussian and $X \sim \gamma_d$. Let $A = A^T$ and $b \in \mathbb{R}^d$, and set

$$T(x) = Ax + b.$$

For affine gradient maps the ridge defect can be computed exactly.

Proposition 4.1 (Exact Gaussian linear calculation)**.** *Let $A = A^T$. Then*

$$\mathcal{R}_{\gamma_d}(Ax + b) = \frac{d\,\mathrm{Tr}(A^2) - (\mathrm{Tr}\,A)^2}{d(d+2)}. \tag{4.1}$$

Moreover,

$$\inf_{\lambda, b_0} \int |Ax + b - (\lambda x + b_0)|^2 \,\mathrm{d}\gamma_d(x) = \mathrm{Tr}(A^2) - \frac{(\mathrm{Tr}\,A)^2}{d}. \tag{4.2}$$

Consequently,

$$\mathcal{R}_{\gamma_d}(Ax + b) = \frac{1}{d+2} \inf_{\lambda, b_0} \|Ax + b - (\lambda x + b_0)\|^2_{L^2(\gamma_d)}. \tag{4.3}$$

Proof. Fix $\theta \in \mathbb{S}^{d-1}$ and decompose

$$X = Z\theta + Y, \qquad Z = \theta \cdot X \sim N(0, 1), \qquad Y \sim N(0, I_d - \theta \otimes \theta),$$

with Z and Y independent. Then

$$\theta \cdot T(X) = Z\theta^T A\theta + \theta^T AY + \theta \cdot b.$$

Thus

$$\mathrm{Var}(\theta \cdot T(X) \mid \theta \cdot X) = \mathbb{E}[(\theta^T AY)^2] = |A\theta|^2 - (\theta^T A\theta)^2.$$

Using

$$\int_{\mathbb{S}^{d-1}} |A\theta|^2 \,\mathrm{d}\sigma(\theta) = \frac{\mathrm{Tr}(A^2)}{d}$$

and the fourth-moment formula

$$\int_{\mathbb{S}^{d-1}} (\theta^T A\theta)^2 \,\mathrm{d}\sigma(\theta) = \frac{(\mathrm{Tr}\,A)^2 + 2\,\mathrm{Tr}(A^2)}{d(d+2)},$$

which is the degree-two instance of the standard spherical moment/Funk–Hecke computation recalled in Appendix A. We obtain (4.1). The optimal translation in (4.2) is $b_0 = b$, and the optimal scalar is $\lambda = \mathrm{Tr}(A)/d$. This gives (4.2) and (4.3). □

Remark 4.2. The spherical averages in Proposition 4.1 are the $n = 2$ case of the tensor calculation in Appendix A. Equivalently, this is the decomposition of a symmetric matrix into its scalar trace and trace-free irreducible parts under the action of $O(d)$.

4.1 SPK for standard Gaussian

Let Tan_{γ_d} be the tangent space at γ_d. We give the proof in some detail because this is the main positive model case of the paper.

We use probabilists' Hermite polynomials, Wick tensors, and the Wiener–Itô chaos decomposition; these standard conventions may be found, for example, in Janson's account of Gaussian Hilbert spaces [17] or Nualart's text on Malliavin calculus [28]. If $A_n \in \mathrm{Sym}^n(\mathbb{R}^d)$, write

$$\psi_n(x) = \langle A_n, {:}x^{\otimes n}{:}\rangle$$

for the corresponding element of the n-th homogeneous Wiener chaos. Then

$$\|\nabla\psi_n\|^2_{L^2(\gamma_d)} = n\,n!\,\|A_n\|^2. \tag{4.4}$$

For each $\theta \in \mathbb{S}^{d-1}$ let P_θ denote conditional expectation with respect to $\theta \cdot X$.

Lemma 4.3 (Projection of one Gaussian chaos)**.** *Let $\psi_n = \langle A_n, {:}X^{\otimes n}{:}\rangle$ with $n \geq 1$. Then*

$$P_\theta(\theta \cdot \nabla\psi_n) = n\,(A_n : \theta^{\otimes n})\,H_{n-1}(\theta \cdot X), \tag{4.5}$$

where H_{n-1} is the one-dimensional probabilists' Hermite polynomial. Consequently,

$$\|P_\theta(\theta \cdot \nabla\psi_n)\|^2_{L^2(\gamma_d)} = n\,n!\,(A_n : \theta^{\otimes n})^2. \tag{4.6}$$

Proof. The directional component of the gradient is

$$\theta \cdot \nabla\psi_n = n\,\big\langle A_n(\theta, \cdot, \ldots, \cdot), {:}X^{\otimes(n-1)}{:}\big\rangle,$$

where $A_n(\theta, \cdot, \ldots, \cdot)$ denotes contraction of one tensor index against θ. The conditional expectation of a Wick monomial onto the one-dimensional Gaussian variable $Z = \theta \cdot X$ is its orthogonal projection onto the chaos generated by Z:

$$\mathbb{E}\big[{:}X^{\otimes(n-1)}{:} \mid Z\big] = H_{n-1}(Z)\,\theta^{\otimes(n-1)}.$$

Contracting with $A_n(\theta, \cdot, \ldots, \cdot)$ gives (4.5). Since

$$\|H_{n-1}(Z)\|^2_{L^2} = (n-1)!,$$

we obtain

$$\|P_\theta(\theta \cdot \nabla\psi_n)\|_2^2 = n^2(A_n : \theta^{\otimes n})^2(n-1)! = n\,n!\,(A_n : \theta^{\otimes n})^2.$$

□

Lemma 4.4 (Chaos contribution to the ridge defect)**.** *For $\psi_n = \langle A_n, {:}X^{\otimes n}{:}\rangle$,*

$$\int_{\mathbb{S}^{d-1}} \|(I - P_\theta)(\theta \cdot \nabla\psi_n)\|^2_{L^2(\gamma_d)}\,\mathrm{d}\sigma(\theta) = n\,n!\left[\frac{1}{d}\|A_n\|^2 - \int_{\mathbb{S}^{d-1}} (A_n : \theta^{\otimes n})^2\,\mathrm{d}\sigma(\theta)\right]. \tag{4.7}$$

Proof. First,

$$\int_{\mathbb{S}^{d-1}} \|\theta \cdot \nabla\psi_n\|^2_{L^2(\gamma_d)}\,\mathrm{d}\sigma(\theta) = \frac{1}{d}\|\nabla\psi_n\|^2_{L^2(\gamma_d)} = \frac{n\,n!}{d}\|A_n\|^2$$

by (4.4) and the normalization of σ. Second, conditional expectation is an orthogonal projection, so

$$\|(I - P_\theta)(\theta \cdot \nabla\psi_n)\|_2^2 = \|\theta \cdot \nabla\psi_n\|_2^2 - \|P_\theta(\theta \cdot \nabla\psi_n)\|_2^2.$$

Integrating in θ and using Lemma 4.3 gives (4.7). □

The next theorem is Theorem 1.4.

Theorem 4.5 (Gaussian sliced Poincaré–Korn inequality)**.** *For every $u \in \mathrm{Tan}_{\gamma_d}$,*

$$\mathrm{dist}^2_{L^2(\gamma_d)}(u, \mathcal{A}_d) \le \frac{d(d+2)}{d-1}\, \mathcal{R}_{\gamma_d}(u). \tag{4.8}$$

Proof. We first assume $u = \nabla\psi$ with ψ a polynomial. Write its homogeneous Wiener-chaos expansion

$$\psi = \sum_{n\ge 0} \psi_n, \qquad \psi_n \in \text{the } n\text{-th chaos}.$$

Then

$$u = \sum_{n\ge 1} \nabla\psi_n,$$

and these vector fields are mutually orthogonal in $L^2(\gamma_d;\mathbb{R}^d)$. For fixed θ, the conditional expectation P_θ preserves homogeneous chaoses, because it is the orthogonal projection onto the closed subspace generated by the one-dimensional Gaussian variable $\theta \cdot X$. Hence the quantities

$$(I - P_\theta)(\theta \cdot \nabla\psi_n)$$

are orthogonal in $L^2(\gamma_d)$ for different n. Therefore

$$\mathcal{R}_{\gamma_d}(u) = \sum_{n\ge 1} \int_{\mathbb{S}^{d-1}} \|(I - P_\theta)(\theta \cdot \nabla\psi_n)\|_2^2 \, \mathrm{d}\sigma(\theta). \tag{4.9}$$

Let $\psi_n = \langle A_n, {:}X^{\otimes n}{:}\rangle$. Lemma 4.4 gives the contribution of the n-th chaos. The rigid modes are exactly the following: $n = 1$ gives constant vector fields b, and the scalar trace part of $n = 2$ gives fields λx. These modes span $\mathcal{A}_d$ and contribute zero to the ridge defect. All other chaos components are orthogonal to $\mathcal{A}_d$.

For the non-rigid components, Lemma 4.6 yields

$$\int_{\mathbb{S}^{d-1}} (A_n : \theta^{\otimes n})^2 \, \mathrm{d}\sigma(\theta) \le \frac{3}{d(d+2)} \|A_n\|^2,$$

with the sharper value $2/[d(d+2)]$ on the trace-free part of $n = 2$. Inserting this into (4.7) gives, for every non-rigid component,

$$\int_{\mathbb{S}^{d-1}} \|(I - P_\theta)(\theta \cdot \nabla\psi_n)\|_2^2 \, \mathrm{d}\sigma(\theta) \ge \left(\frac{1}{d} - \frac{3}{d(d+2)}\right) n\, n! \|A_n\|^2.$$

By (4.4), this is

$$\ge \frac{d-1}{d(d+2)} \|\nabla\psi_n\|^2_{L^2(\gamma_d)}.$$

Summing over the orthogonal complement of the rigid modes and using (4.9), we obtain

$$\mathcal{R}_{\gamma_d}(u) \ge \frac{d-1}{d(d+2)}\, \mathrm{dist}^2_{L^2(\gamma_d)}(u, \mathcal{A}_d).$$

For a general $u \in \mathrm{Tan}_{\gamma_d}$, choose polynomial gradients u_j converging to u in $L^2(\gamma_d)$. The distance to the finite-dimensional closed space $\mathcal{A}_d$ is continuous in L^2. The ridge defect is lower semicontinuous, since for each θ it is the squared distance of $\theta \cdot u$ to a closed subspace of $L^2(\gamma_d)$, and then one integrates in θ. Passing to the limit gives (4.8). □

Lemma 4.6 (Spherical tensor estimate)**.** *Let A be a symmetric n-tensor on $\mathbb{R}^d$. If either $n \ge 3$, or $n = 2$ and A is trace-free, then*

$$\int_{\mathbb{S}^{d-1}} (A : \theta^{\otimes n})^2 \, \mathrm{d}\sigma(\theta) \le \frac{3}{d(d+2)} \|A\|^2.$$

For $n = 2$ and $\mathrm{Tr}\, A = 0$ one has the sharper identity

$$\int_{\mathbb{S}^{d-1}} (\theta^T A\theta)^2 \, \mathrm{d}\sigma(\theta) = \frac{2}{d(d+2)} \|A\|^2_{\mathrm{HS}}.$$

Proof. The trace-free $n = 2$ identity is the usual fourth-moment formula on the sphere. The general estimate is a standard finite-dimensional Funk–Hecke, or equivalently representation-theoretic, calculation for symmetric tensors; see Appendix A and the references therein for the normalization. The appendix shows that the eigenvalues of the operator

$$A \mapsto \int_{\mathbb{S}^{d-1}} (A : \theta^{\otimes n})\theta^{\otimes n} \, \mathrm{d}\sigma(\theta)$$

on the irreducible trace components of $\mathrm{Sym}^n(\mathbb{R}^d)$ are bounded by $3/[d(d+2)]$ except for the scalar trace component in degree $n = 2$, which corresponds exactly to the rigid field $x \mapsto \lambda x$. This proves the lemma. □

Remark 4.7 (Sharpness of the Gaussian SPK constant). Proposition 4.1 shows that on trace-free linear gradient fields the exact ratio is $1/(d + 2)$. The lower bound in Theorem 4.5 is weaker for $d > 2$ because higher Hermite modes may have a smaller spectral gap. We have not optimized the full Gaussian spectral gap in this paper.

The relative gap between the two coefficients is the factor $(d - 1)/d$, which tends to 1 as $d \to \infty$ but is strictly less than 1 for every finite $d \geq 2$. This discrepancy originates from the higher Wiener chaos modes ($n \geq 3$): in Lemma 4.6 we used the uniform estimate $3/[d(d+2)]$ for all $n \geq 3$, which is not sharp for every individual degree. Some higher-order trace components may have a smaller spherical projection energy, and the infimum in the SPK constant is dictated by the worst spectral gap across all non-rigid modes. Consequently, the optimal global constant $\kappa_{\mathrm{SPK}}(\gamma_d)$ lies strictly between $(d - 1)/[d(d + 2)]$ and $1/(d + 2)$, and its exact value remains an open problem.

4.2 Stability for standard Gaussian

Corollary 4.8 (Gaussian stability). *Let $T = \nabla\varphi \in \mathrm{Tan}_{\gamma_d}$ be the optimal transport map from γ_d to ν. Let τ_θ be the monotone transport map from $(\gamma_d)_\theta$ to ν_θ, and suppose*

$$\Lambda := \operatorname*{ess\,sup}_\theta \mathrm{Lip}(\tau_\theta) < \infty.$$

Then

$$\mathrm{dist}^2_{L^2(\gamma_d)}(T, \mathcal{A}_d) \leq \frac{d(d+2)}{d-1} \, \Lambda \, \mathcal{D}(\gamma_d, \nu).$$

Proof. For each θ, (3.3) gives

$$\int |\theta \cdot T(x) - \tau_\theta(\theta \cdot x)|^2 \, \mathrm{d}\gamma_d(x) \leq \Lambda g_\theta(T).$$

Since $h = \tau_\theta$ is admissible in the definition of $\mathcal{R}_{\gamma_d}(T)$,

$$\mathcal{R}_{\gamma_d}(T) \leq \Lambda \int g_\theta(T) \, \mathrm{d}\sigma(\theta) = \Lambda \mathcal{D}(\gamma_d, \nu).$$

Theorem 4.5 completes the proof. □

4.3 Isotropic Gaussian measures

Proposition 4.9 (Isotropic Gaussian measures). *Let*

$$\gamma_{a,\sigma} := N(a, \sigma^2 I_d), \qquad a \in \mathbb{R}^d, \quad \sigma > 0.$$

Then

$$\kappa_{\mathrm{SPK}}(\gamma_{a,\sigma}) \geq \frac{d-1}{d(d+2)}.$$

Equivalently, the normalized sliced Poincaré–Korn constant satisfies

$$\overline{\kappa}_{\mathrm{SPK}}(\gamma_{a,\sigma}) = d\,\kappa_{\mathrm{SPK}}(\gamma_{a,\sigma}) \geq \frac{d-1}{d+2}.$$

Consequently, the same quantitative stability estimate holds as in the standard Gaussian case.

Proof. The case $a = 0$ and $\sigma = 1$ is exactly the Gaussian sliced Poincaré–Korn inequality proved in Theorem 4.5. We reduce the general spherical case to the standard one.

Let $X \sim \gamma_{a,\sigma}$. Write

$$X = a + \sigma Z, \qquad Z \sim \gamma_d.$$

For a gradient field $u = \nabla_x \psi \in L^2(\gamma_{a,\sigma}; \mathbb{R}^d)$, define

$$v(z) := u(a + \sigma z).$$

Then v is also a gradient field with respect to the z-variable, since

$$v(z) = \nabla_z \left(\frac{1}{\sigma} \psi(a + \sigma z) \right).$$

For every $\theta \in \mathbb{S}^{d-1}$,

$$\theta \cdot X = \theta \cdot a + \sigma\, \theta \cdot Z.$$

Thus conditioning on $\theta \cdot X$ is equivalent to conditioning on $\theta \cdot Z$. Hence

$$\mathcal{R}_{\gamma_{a,\sigma}}(u) = \mathcal{R}_{\gamma_d}(v).$$

Moreover,

$$\mathrm{dist}^2_{L^2(\gamma_{a,\sigma})}(u, \mathcal{A}_d) = \mathrm{dist}^2_{L^2(\gamma_d)}(v, \mathcal{A}_d).$$

Indeed,

$$\lambda(a + \sigma z) + b = (\lambda\sigma) z + (\lambda a + b),$$

and as $\lambda \in \mathbb{R}$ and $b \in \mathbb{R}^d$ vary, the pair

$$\alpha = \lambda\sigma, \qquad \beta = \lambda a + b$$

runs over all $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^d$.

Applying Theorem 4.5 to v under γ_d, we obtain

$$\mathcal{R}_{\gamma_d}(v) \geq \frac{d-1}{d(d+2)} \,\mathrm{dist}^2_{L^2(\gamma_d)}(v, \mathcal{A}_d).$$

Using the two identities above gives

$$\mathcal{R}_{\gamma_{a,\sigma}}(u) \geq \frac{d-1}{d(d+2)} \,\mathrm{dist}^2_{L^2(\gamma_{a,\sigma})}(u, \mathcal{A}_d).$$

Therefore translations and scalar dilations do not change the sliced Poincaré–Korn constant, and

$$\kappa_{\mathrm{SPK}}(\gamma_{a,\sigma}) \geq \frac{d-1}{d(d+2)}.$$

For the Wasserstein stability estimate, if

$$S(z) = \frac{T(a + \sigma z) - a}{\sigma},$$

then

$$\mathrm{dist}^2_{L^2(\gamma_{a,\sigma})}(T, \mathcal{A}_d) = \sigma^2 \,\mathrm{dist}^2_{L^2(\gamma_d)}(S, \mathcal{A}_d),$$

and

$$\mathcal{D}\big(\gamma_{a,\sigma}, T_{\#}\gamma_{a,\sigma}\big) = \sigma^2 \mathcal{D}\big(\gamma_d, S_{\#}\gamma_d\big).$$

Thus the stability constant is unchanged. $\square$

5 More examples

5.1 Perturbations of the Gaussian

Proposition 5.1 (Bounded Gaussian perturbations)**.** *Assume*

$$\mathrm{d}\mu = \rho\,\mathrm{d}\gamma_d, \qquad 0 < m \le \rho \le M < \infty.$$

Then

$$\kappa_{\mathrm{SPK}}(\mu) \ge \frac{m}{M}\frac{d-1}{d(d+2)}.$$

Consequently, for Brenier maps $T : \mu \to \nu$ with $T \in \mathrm{Tan}_\mu$ and satisfying $\operatorname{ess\,sup}_\theta \mathrm{Lip}(\tau_\theta) \le \Lambda$,

$$\mathrm{dist}^2_{L^2(\mu)}(T, \mathcal{A}_d) \le \frac{M}{m}\frac{d(d+2)}{d-1}\,\Lambda\,\mathcal{D}(\mu,\nu).$$

Proof. For every direction and every ridge function $h(\theta \cdot x)$,

$$\int |\theta\cdot u - h(\theta\cdot x)|^2 \,\mathrm{d}\mu \ge m \int |\theta\cdot u - h(\theta\cdot x)|^2 \,\mathrm{d}\gamma_d.$$

Taking the infimum over h and integrating in θ gives

$$\mathcal{R}_\mu(u) \ge m\mathcal{R}_{\gamma_d}(u).$$

On the other hand,

$$\mathrm{dist}^2_{L^2(\mu)}(u, \mathcal{A}_d) \le M\,\mathrm{dist}^2_{L^2(\gamma_d)}(u, \mathcal{A}_d).$$

The Gaussian sliced Poincaré–Korn inequality gives the claim. □

5.2 Fixed measures and compact classes

The next criterion is useful when one wants a rigorous stability theorem for a fixed source measure, without claiming a dimension-free constant.

Proposition 5.2 (Compactness criterion)**.** *Let $\mu = \rho\mathcal{L}^d$ on a connected open set Ω with $\mu(\Omega) = 1$, and assume that ρ is locally bounded below on Ω. Let $\mathcal{C} \subset \mathrm{Tan}_\mu$ be a class of gradient fields such that*

$$\{u - a_u : u \in \mathcal{C}\}$$

is relatively compact in $L^2(\mu;\mathbb{R}^d)$, where $a_u \in \mathcal{A}_d$ is a nearest point to u in $\mathcal{A}_d$. Then there exists $\kappa(\mu,\mathcal{C}) > 0$ such that

$$\mathrm{dist}^2_{L^2(\mu)}(u, \mathcal{A}_d) \le \frac{1}{\kappa(\mu,\mathcal{C})}\mathcal{R}_\mu(u) \qquad \forall u \in \mathcal{C}.$$

Proof. Suppose not. Then there is a sequence $u_n \in \mathcal{C}$ with

$$\mathrm{dist}_{L^2(\mu)}(u_n, \mathcal{A}_d) = 1, \qquad \mathcal{R}_\mu(u_n) \to 0.$$

Subtract nearest affine fields and write $v_n = u_n - a_n$. By compactness, a subsequence converges in $L^2(\mu)$ to some v. Since the class of distributional gradients is closed under $L^2(\mu)$ convergence under the local lower bound on ρ, the limit v is again a distributional gradient on Ω. The ridge defect is continuous on $L^2(\mu;\mathbb{R}^d)$: for each θ it is the squared distance of $\theta\cdot u$ to a closed subspace of $L^2(\mu)$, and these distances are dominated by $|u|^2$. Hence $\mathcal{R}_\mu(v) = 0$ and $\mathrm{dist}(v, \mathcal{A}_d) = 1$. Lemma 3.1 gives $v \in \mathcal{A}_d$, a contradiction. □

Remark 5.3 (Natural compact classes). This is not a uniform geometric estimate. It is a compactness device for fixed measures and a priori compact families of fields. Several natural classes fit this framework.

First, if Ω is bounded and the density of μ is bounded above and below on Ω, then any class bounded in $W^{1,p}(\Omega;\mathbb{R}^d)$ with $p > d$, after subtracting nearest elements of $\mathcal{A}_d$, is relatively compact in $L^2(\mu)$ by Morrey–Rellich compactness. A similar statement holds for $W^{s,2}$ classes with $s > 0$ on bounded domains, with the usual compact embedding.

Second, on unbounded domains with confining tails, one may combine local compactness with a uniform tail estimate. For example, a class of gradient fields with uniform local Lipschitz bounds and a uniform quadratic growth bound is precompact in $L^2(\mu)$ whenever the measure has sufficiently strong tails, after fixing the affine normalization. This is the form relevant for regularized Brenier maps on strongly log-concave backgrounds.

Third, heat or Ornstein–Uhlenbeck regularization gives compact classes at positive time in many confining settings. If P_t is a Markov semigroup whose generator has compact resolvent on $L^2(\mu)$, then P_t maps bounded subsets of $L^2(\mu)$ into relatively compact subsets for every $t > 0$. For the Gaussian measure this is immediate from the Hermite expansion of the Ornstein–Uhlenbeck semigroup: the multiplier e^{-nt} tends to zero on the n-th chaos. Thus families such as $\{\nabla P_t f : \|f\|_{L^2} \le 1,\ t \ge t_0 > 0\}$ satisfy the compactness assumption.

Finally, finite-dimensional ansatz classes, for instance fields supported on finitely many Hermite modes or finitely many basis functions, also satisfy the hypothesis. In all these examples the constant $\kappa(\mu, \mathcal{C})$ is non-explicit and depends on the chosen class.

5.3 An anisotropic Gaussian counterexample

We now prove Proposition 1.5. Let

$$\mu_\varepsilon = N\left(0, \begin{pmatrix} \varepsilon^2 & 0 \\ 0 & 1 \end{pmatrix}\right), \qquad 0 < \varepsilon < 1.$$

Then $\mu_\varepsilon = e^{-V_\varepsilon}\,\mathrm{d}x/Z_\varepsilon$ with

$$V_\varepsilon(x_1, x_2) = \frac{x_1^2}{2\varepsilon^2} + \frac{x_2^2}{2}, \qquad \nabla^2 V_\varepsilon = \begin{pmatrix} \varepsilon^{-2} & 0 \\ 0 & 1 \end{pmatrix} \ge I.$$

Thus the Bakry–Émery curvature lower bound is uniform. The Poincaré constant is also uniformly bounded by 1.

Consider the gradient field

$$u(x_1, x_2) = (x_2, x_1) = \nabla(x_1 x_2).$$

Since $\mathbb{E}u = 0$ and

$$\mathbb{E}[X \cdot u(X)] = 2\mathbb{E}[X_1 X_2] = 0,$$

this field is orthogonal to $\mathcal{A}_2$ in $L^2(\mu_\varepsilon)$. Therefore

$$\mathrm{dist}^2_{L^2(\mu_\varepsilon)}(u, \mathcal{A}_2) = \int |u|^2 \,\mathrm{d}\mu_\varepsilon = 1 + \varepsilon^2. \tag{5.1}$$

Let $\theta = (\cos t, \sin t)$. Then

$$\theta \cdot X = \cos t\, X_1 + \sin t\, X_2, \qquad \theta \cdot u(X) = \cos t\, X_2 + \sin t\, X_1.$$

For jointly Gaussian variables,

$$\mathrm{Var}(Y \mid Z) = \mathrm{Var}(Y) - \frac{\mathrm{Cov}(Y, Z)^2}{\mathrm{Var}(Z)}.$$

A direct calculation gives

$$\operatorname{Var}\big(\theta \cdot u(X) \mid \theta \cdot X\big) = \frac{\varepsilon^2 \cos^2(2t)}{\varepsilon^2 \cos^2 t + \sin^2 t}. \tag{5.2}$$

Hence

$$\mathcal{R}_{\mu_\varepsilon}(u) = \frac{1}{2\pi} \int_0^{2\pi} \frac{\varepsilon^2 \cos^2(2t)}{\varepsilon^2 \cos^2 t + \sin^2 t} \, \mathrm{d}t \tag{5.3}$$

$$\leq \frac{1}{2\pi} \int_0^{2\pi} \frac{\varepsilon^2}{\varepsilon^2 \cos^2 t + \sin^2 t} \, \mathrm{d}t = \varepsilon. \tag{5.4}$$

Together with (5.1), this gives

$$\frac{\mathcal{R}_{\mu_\varepsilon}(u)}{\operatorname{dist}^2_{L^2(\mu_\varepsilon)}(u, \mathcal{A}_2)} \leq \frac{\varepsilon}{1+\varepsilon^2} \to 0.$$

Thus $\kappa_{\mathrm{SPK}}(\mu_\varepsilon) \to 0$.

The preceding SPK degeneration produces an actual stability counterexample. The computation is completely explicit because all projected marginals are one-dimensional Gaussians. Let

$$S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad T_\delta(x) = (I + \delta S)x.$$

For $|\delta| < 1$, the matrix $I + \delta S$ is positive definite, so

$$T_\delta = \nabla \left(\frac{1}{2}|x|^2 + \delta x_1 x_2 \right)$$

is a Brenier map. Let $\nu_{\varepsilon,\delta} = (T_\delta)_\# \mu_\varepsilon$. Then

$$T_\delta(x) - x = \delta u(x), \qquad \operatorname{dist}^2_{L^2(\mu_\varepsilon)}(T_\delta, \mathcal{A}_2) = \delta^2(1+\varepsilon^2).$$

Fix $\theta = (\cos t, \sin t)$ and put

$$Z = \theta \cdot X, \qquad Y = \theta \cdot u(X), \qquad X \sim \mu_\varepsilon.$$

Set

$$a = \operatorname{Var} Z = \varepsilon^2 \cos^2 t + \sin^2 t, \quad c = \operatorname{Cov}(Z, Y) = (1+\varepsilon^2) \sin t \cos t, \quad q = \operatorname{Var}(Y \mid Z).$$

Then the variance of $\theta \cdot T_\delta(X) = Z + \delta Y$ is

$$b_\delta = a + 2\delta c + \delta^2 \operatorname{Var} Y.$$

Since the Wasserstein distance between centered one-dimensional Gaussians of variances a and b_δ is $(\sqrt{b_\delta} - \sqrt{a})^2$, the directional deficit is

$$g_\theta(\delta) = \delta^2 \operatorname{Var} Y - (\sqrt{b_\delta} - \sqrt{a})^2. \tag{5.5}$$

Because $a \geq \varepsilon^2 > 0$, Taylor expansion in δ is uniform in t for each fixed ε. From (5.5),

$$\lim_{\delta \to 0} \frac{g_\theta(\delta)}{\delta^2} = \operatorname{Var} Y - \frac{c^2}{a} = \operatorname{Var}(Y \mid Z).$$

The last quantity is exactly the integrand in (5.2). Dominated convergence therefore gives the exact second variation

$$\lim_{\delta \to 0} \frac{\mathcal{D}(\mu_\varepsilon, \nu_{\varepsilon,\delta})}{\delta^2} = \mathcal{R}_{\mu_\varepsilon}(u). \tag{5.6}$$

Consequently, for every fixed ε and all sufficiently small δ,

$$\mathcal{D}(\mu_\varepsilon, \nu_{\varepsilon,\delta}) \le 2\delta^2 \mathcal{R}_{\mu_\varepsilon}(u) \le 2\delta^2\varepsilon.$$

It follows that

$$\frac{\operatorname{dist}^2_{L^2(\mu_\varepsilon)}(T_\delta, \mathcal{A}_2)}{\mathcal{D}(\mu_\varepsilon, \nu_{\varepsilon,\delta})} \ge \frac{1+\varepsilon^2}{2\varepsilon} \longrightarrow \infty.$$

Moreover, if $\delta = o(\varepsilon)$, the projected one-dimensional Gaussian maps have Lipschitz constants uniformly bounded. Indeed, their Lipschitz factors are $\sqrt{b_\delta/a}$, and the lower bound $a \ge \varepsilon^2$ gives $b_\delta/a = 1 + O(\delta/\varepsilon) + O(\delta^2/\varepsilon^2)$ uniformly in t. Thus the obstruction persists even when the one-dimensional scale parameter Λ is kept bounded.

Therefore no stability theorem of the form

$$\operatorname{dist}^2(T, \mathcal{A}_d) \le C\mathcal{D}(\mu, \nu)$$

can hold with C depending only on a Bakry–Émery lower curvature bound or on a usual Poincaré constant.

Remark 5.4 (Further directions). The counterexample above leaves two natural problems. First, one may seek positive SPK estimates under isotropic log-concavity assumptions, possibly using stochastic localization or leaf-decomposition methods. Second, the sharp Gaussian sliced Poincaré–Korn constant remains open beyond the exact affine calculation in Proposition 4.1.

A The spherical tensor estimate

We give the finite-dimensional computation used in Lemma 4.6. This is a standard representation-theoretic calculation: symmetric tensors are decomposed into irreducible trace components, and every $O(d)$-invariant quadratic form is diagonal on these components by Schur's lemma. In the language of tensor spherical harmonics, Higuchi's treatment of symmetric trace-free tensor spherical harmonics on spheres [15] gives a closely related systematic account; the classical diagonalization framework for spherical functions on Grassmann manifolds is represented by James–Constantine [16]. Let $\operatorname{Sym}^n(\mathbb{R}^d)$ be the space of symmetric n-tensors with the Hilbert–Schmidt inner product. For $A \in \operatorname{Sym}^n(\mathbb{R}^d)$ define

$$Q_n(A) := \int_{\mathbb{S}^{d-1}} (A : \theta^{\otimes n})^2 \, \mathrm{d}\sigma(\theta).$$

The quadratic form Q_n is invariant under the orthogonal group $O(d)$. The standard irreducible decomposition of symmetric tensors is

$$\operatorname{Sym}^n(\mathbb{R}^d) = \bigoplus_{j=0}^{\lfloor n/2 \rfloor} I^{\odot j} \odot \mathcal{H}_{n-2j}, \tag{A.1}$$

where $\mathcal{H}_m$ denotes the trace-free symmetric m-tensors, equivalently harmonic homogeneous polynomials of degree m, and $\odot$ is normalized so that

$$(I^{\odot j} \odot H) : \theta^{\otimes n} = H : \theta^{\otimes m} \qquad (m = n - 2j,\ |\theta| = 1).$$

By Schur's lemma, Q_n is diagonal on (A.1). If

$$A = I^{\odot j} \odot H, \qquad H \in \mathcal{H}_m, \qquad n = m + 2j,$$

then a direct contraction calculation gives the eigenvalue

$$\frac{Q_n(A)}{\|A\|^2} = \lambda^{(d)}_{m,j} := \frac{(m+2j)!\,\Gamma(d/2)}{2^{m+2j} j!\,\Gamma(m+j+d/2)}. \tag{A.2}$$

We recall the two ingredients in this calculation. First, for $H \in \mathcal{H}_m$,

$$\int_{\mathbb{S}^{d-1}} (H : \theta^{\otimes m})^2 \, \mathrm{d}\sigma(\theta) = \frac{m!\Gamma(d/2)}{2^m\Gamma(m+d/2)}\|H\|^2. \tag{A.3}$$

Second, with the above normalization of $I^{\odot j} \odot H$,

$$\|I^{\odot j} \odot H\|^2 = \frac{m!\, 2^{2j} j!\, \Gamma(m+j+d/2)}{(m+2j)!\, \Gamma(m+d/2)}\|H\|^2. \tag{A.4}$$

Dividing (A.3) by (A.4) gives (A.2).

The rigid scalar mode in degree two is $(m,j) = (0,1)$; its eigenvalue is $1/d$. The trace-free quadratic mode is $(m,j) = (2,0)$, and

$$\lambda_{2,0}^{(d)} = \frac{2}{d(d+2)}.$$

For $n \geq 3$, the maximum of (A.2) over all trace components is

$$\frac{3}{d(d+2)}.$$

Indeed, the value $3/[d(d+2)]$ is attained for $(m,j) = (1,1)$ in degree 3 and for $(m,j) = (0,2)$ in degree 4, and the elementary ratio comparison

$$\frac{\lambda_{m,j+1}^{(d)}}{\lambda_{m,j}^{(d)}} = \frac{(m+2j+2)(m+2j+1)}{4(j+1)(m+j+d/2)}$$

combined with the explicit $j = 0$ values shows that all remaining components are no larger. Consequently,

$$Q_n(A) \leq \frac{3}{d(d+2)}\|A\|^2$$

for every $n \geq 3$, and for $n = 2$ on the trace-free subspace. This proves Lemma 4.6.

B Grassmannian projection moments

In this appendix we give the detailed computation used in Lemma 2.6. The argument is a standard invariant-theoretic computation on the real Grassmannian: the quadratic form is diagonalized by the irreducible decomposition of symmetric two-tensors under the orthogonal group, and the remaining scalar is fixed by evaluating one trace-free test matrix; see, for instance, James–Constantine [16] for the classical spherical-function framework on Grassmannians.

Let $G_{d,k}$ be the Grassmannian of k-dimensional linear subspaces of $\mathbb{R}^d$, equipped with its $O(d)$-invariant probability measure $\pi_{d,k}$. For $E \in G_{d,k}$, let P_E denote the orthogonal projection onto E. For a symmetric $d \times d$ matrix M, set

$$Q(M) := \int_{G_{d,k}} \|P_E M(I - P_E)\|_{\mathrm{HS}}^2 \, \mathrm{d}\pi_{d,k}(E).$$

Step 1: invariant reduction. For every $R \in O(d)$, one has $P_{RE} = RP_E R^\mathsf{T}$. Since $\pi_{d,k}$ is $O(d)$-invariant, it follows that

$$Q(RMR^\mathsf{T}) = Q(M).$$

Thus Q is an $O(d)$-invariant quadratic form on the space S_d of real symmetric $d \times d$ matrices.

Moreover,

$$Q(\alpha I) = 0$$

for every $\alpha \in \mathbb{R}$, because $P_E(I - P_E) = 0$. Writing

$$S_d^0 := \{M \in S_d : \operatorname{Tr} M = 0\},$$

we have the $O(d)$-orthogonal decomposition

$$S_d = S_d^0 \oplus \mathbb{R} I.$$

Since Q vanishes on the scalar summand, it remains to understand its restriction to S_d^0.

The $O(d)$-representation S_d^0 is irreducible. By Schur's lemma, or equivalently by the uniqueness of the $O(d)$-invariant inner product on this irreducible representation, the polarization of $Q|_{S_d^0}$ must be a scalar multiple of the Hilbert–Schmidt inner product. Therefore there exists a constant $C(d,k) \geq 0$ such that

$$Q(M) = C(d,k) \left\| M - \frac{\operatorname{Tr} M}{d} I \right\|_{\mathrm{HS}}^2 \qquad \text{for all } M \in S_d. \tag{B.1}$$

Since $1 \leq k \leq d-1$, the constant computed below is positive. Consequently, if $Q(M) = 0$, then M is a scalar matrix.

Step 2: evaluation of the constant. We evaluate (B.1) at the normalized trace-free matrix

$$M_0 := \frac{1}{\sqrt{2}} \operatorname{diag}(1, -1, 0, \dots, 0).$$

Then $\operatorname{Tr} M_0 = 0$ and $\|M_0\|_{\mathrm{HS}}^2 = 1$, hence

$$C(d,k) = Q(M_0).$$

Write $P_E = (p_{ab})_{a,b=1}^d$. Since $P_E = P_E^\mathsf{T} = P_E^2$, we have

$$\|P_E M_0 (I - P_E)\|_{\mathrm{HS}}^2 = \operatorname{Tr}\big(P_E M_0^2 P_E\big) - \operatorname{Tr}\big(P_E M_0 P_E M_0 P_E\big).$$

Using the fact that M_0 has only the diagonal entries $1/\sqrt{2}$ and $-1/\sqrt{2}$, this gives

$$\|P_E M_0 (I - P_E)\|_{\mathrm{HS}}^2 = \frac{1}{2}\big(p_{11} - p_{11}^2 + p_{22} - p_{22}^2\big) + p_{12}^2. \tag{B.2}$$

We now compute the required moments. First, by invariance,

$$\int_{G_{d,k}} P_E \, \mathrm{d}\pi_{d,k}(E) = \frac{k}{d} I,$$

and therefore

$$\int_{G_{d,k}} p_{ii} \, \mathrm{d}\pi_{d,k}(E) = \frac{k}{d}. \tag{B.3}$$

For the second moments, $O(d)$-invariance implies that there are constants $A_{d,k}$ and $B_{d,k}$ such that

$$\int_{G_{d,k}} p_{ab} p_{cd} \, \mathrm{d}\pi_{d,k}(E) = A_{d,k} \delta_{ab} \delta_{cd} + B_{d,k} \big(\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}\big). \tag{B.4}$$

The two constants are determined by the identities

$$\operatorname{Tr} P_E = k, \qquad \operatorname{Tr}(P_E^2) = \operatorname{Tr} P_E = k.$$

Indeed, summing (B.4) over a, c with $b = a$, $d = c$, we get

$$k^2 = \int (\operatorname{Tr} P_E)^2 \, \mathrm{d}\pi_{d,k}(E) = d^2 A_{d,k} + 2d B_{d,k}.$$

Similarly, summing (B.4) over a, b with $c = a$, $d = b$, we get

$$k = \int \operatorname{Tr}(P_E^2)\, \mathrm{d}\pi_{d,k}(E) = dA_{d,k} + d(d+1)B_{d,k}.$$

Solving these two equations yields

$$B_{d,k} = \frac{k(d-k)}{d(d-1)(d+2)}, \qquad A_{d,k} = \frac{k((d+1)k-2)}{d(d-1)(d+2)}.$$

Consequently,

$$\int_{G_{d,k}} p_{ii}^2\, \mathrm{d}\pi_{d,k}(E) = A_{d,k} + 2B_{d,k} = \frac{k(k+2)}{d(d+2)}, \tag{B.5}$$

and, for $i \neq j$,

$$\int_{G_{d,k}} p_{ij}^2\, \mathrm{d}\pi_{d,k}(E) = B_{d,k} = \frac{k(d-k)}{d(d-1)(d+2)}. \tag{B.6}$$

Inserting (B.3), (B.5), and (B.6) into (B.2), we obtain

$$\begin{aligned} C(d,k) &= \frac{1}{2} \cdot 2 \left(\frac{k}{d} - \frac{k(k+2)}{d(d+2)} \right) + \frac{k(d-k)}{d(d-1)(d+2)} \\ &= \frac{k(d-k)}{d(d+2)} + \frac{k(d-k)}{d(d-1)(d+2)} \\ &= \frac{k(d-k)}{(d-1)(d+2)}. \end{aligned}$$

Therefore

$$\int_{G_{d,k}} \|P_E M(I-P_E)\|_{\mathrm{HS}}^2\, \mathrm{d}\pi_{d,k}(E) = \frac{k(d-k)}{(d-1)(d+2)} \left\| M - \frac{\operatorname{Tr} M}{d} I \right\|_{\mathrm{HS}}^2. \tag{B.7}$$

This proves Lemma 2.6.

References

[1] N. Bonneel, J. Rabin, G. Peyré, and H. Pfister, Sliced and Radon Wasserstein barycenters of measures, *J. Math. Imaging Vision* **51** (2015), no. 1, 22–45. 2, 3

[2] N. Bonnotte, *Unidimensional and Evolution Methods for Optimal Transportation*, Ph.D. thesis, Université Paris-Sud, 2013. 3

[3] L. A. Caffarelli, Monotonicity properties of optimal transportation and the FKG and related inequalities, *Comm. Math. Phys.* **214** (2000), no. 3, 547–563. 12

[4] F. Cavalletti and A. Mondino, Sharp and rigid isoperimetric inequalities in metric-measure spaces with lower Ricci curvature bounds, *Invent. Math.* **208** (2017), no. 3, 803–849. 3

[5] P. G. Ciarlet, *Mathematical Elasticity. Vol. I. Three-Dimensional Elasticity*, Studies in Mathematics and its Applications, vol. 20, North-Holland, Amsterdam, 1988. 10

[6] K. J. Ciosmak, Leaves decompositions in Euclidean spaces, *J. Math. Pures Appl. (9)* **154** (2021), 212–244. 5

[7] T. A. Courtade, M. Fathi, and A. Pananjady, Quantitative stability of the entropy power inequality, *IEEE Trans. Inform. Theory* **64** (2018), no. 8, 5691–5703. 3

[8] B. Dacorogna, *Direct Methods in the Calculus of Variations*, 2nd ed., Applied Mathematical Sciences, vol. 78, Springer, New York, 2008. 10

[9] A. Delalande and Q. Mérigot, Quantitative stability of optimal transport maps under variations of the target measure, *Duke Math. J.* **172** (2023), no. 17, 3321–3357. 3

[10] R. Eldan, Thin shell implies spectral gap up to polylog via a stochastic localization scheme, *Geom. Funct. Anal.* **23** (2013), no. 2, 532–569. 5

[11] M. Fathi, E. Indrei, and M. Ledoux, Quantitative logarithmic Sobolev inequalities and stability estimates, *Discrete Contin. Dyn. Syst.* **36** (2016), no. 12, 6835–6853. 3

[12] M. Fathi, N. Gozlan, and M. Prod'homme, A proof of the Caffarelli contraction theorem via entropic regularization, *Calc. Var. Partial Differential Equations* **59** (2020), no. 3, Paper No. 96, 23 pp. 12

[13] A. Figalli, F. Maggi, and A. Pratelli, A mass transportation approach to quantitative isoperimetric inequalities, *Invent. Math.* **182** (2010), no. 1, 167–211. 3

[14] G. Friesecke, R. D. James, and S. Müller, A theorem on geometric rigidity and the derivation of nonlinear plate theory from three-dimensional elasticity, *Comm. Pure Appl. Math.* **55** (2002), no. 11, 1461–1506. 10

[15] A. Higuchi, Symmetric tensor spherical harmonics on the N-sphere and their application to the de Sitter group $SO(N,1)$, *J. Math. Phys.* **28** (1987), no. 7, 1553–1566. 21

[16] A. T. James and A. G. Constantine, Generalized Jacobi polynomials as spherical functions of the Grassmann manifold, *Proc. London Math. Soc. (3)* **29** (1974), 174–192. 21, 22

[17] S. Janson, *Gaussian Hilbert Spaces*, Cambridge Tracts in Mathematics, vol. 129, Cambridge University Press, Cambridge, 1997. 14

[18] R. Kannan, L. Lovász, and M. Simonovits, Isoperimetric problems for convex bodies and a localization lemma, *Discrete Comput. Geom.* **13** (1995), no. 3–4, 541–559. 5

[19] B. Klartag and J. Lehec, Affirmative resolution of Bourgain's slicing problem using Guan's bound, *Geom. Funct. Anal.* **35** (2025), 1147–1168. 5

[20] S. Kolouri, S. R. Park, M. Thorpe, D. Slepčev, and G. K. Rohde, Optimal mass transport: signal processing and machine-learning applications, *IEEE Signal Process. Mag.* **34** (2017), no. 4, 43–59. 3

[21] S. Kolouri, K. Nadjahi, U. Şimşekli, R. Badeau, and G. K. Rohde, Generalized sliced Wasserstein distances, in *Advances in Neural Information Processing Systems 32*, Curran Associates, Red Hook, NY, 2019. 3

[22] T. Le, M. Yamada, K. Fukumizu, and M. Cuturi, Tree-sliced variants of Wasserstein distances, in *Advances in Neural Information Processing Systems 32*, Curran Associates, Red Hook, NY, 2019. 3

[23] Y. T. Lee and S. Vempala, Eldan's stochastic localization and the KLS conjecture: isoperimetry, concentration and mixing, *Ann. of Math. (2)* **199** (2024), no. 3, 1043–1092. 5

[24] Q. Mérigot, A. Delalande, and F. Chazal, Quantitative stability of optimal transport maps and linearization of the 2-Wasserstein space, in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, Proc. Mach. Learn. Res., vol. 108, PMLR, 2020, pp. 3186–3196. 3

[25] K. Nadjahi, A. Durmus, L. Chizat, S. Kolouri, S. Shahrampour, and U. Şimşekli, Statistical and topological properties of sliced probability divergences, in *Advances in Neural Information Processing Systems 33*, Curran Associates, Red Hook, NY, 2020, pp. 20802–20812. 3

[26] K. Nadjahi, *Sliced-Wasserstein Distance for Large-Scale Machine Learning: Theory, Methodology and Extensions*, Ph.D. thesis, Institut Polytechnique de Paris, 2021. 3

[27] K. Nguyen, N. Ho, T. Pham, and H. Bui, Distributional sliced-Wasserstein and applications to generative modeling, in *International Conference on Learning Representations*, 2021. 3

[28] D. Nualart, *The Malliavin Calculus and Related Topics*, 2nd ed., Probability and Its Applications, Springer-Verlag, Berlin, 2006. 14

[29] J. Rabin, G. Peyré, J. Delon, and M. Bernot, Wasserstein barycenter and its application to texture mixing, in *Scale Space and Variational Methods in Computer Vision*, Lecture Notes in Comput. Sci., vol. 6667, Springer, Berlin, 2012, pp. 435–446. 3

[30] F. Santambrogio, *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*, Progress in Nonlinear Differential Equations and Their Applications, vol. 87, Birkhäuser/Springer, Cham, 2015. 2

[31] C. Villani, *Optimal Transport: Old and New*, Grundlehren der mathematischen Wissenschaften, vol. 338, Springer, Berlin, 2009. 2

Declaration. The authors declare that they have no conflict of interest and that the manuscript has no associated data. During the preparation of this manuscript, the authors used Kimi (Moonshot AI) and ChatGPT (OpenAI) for language polishing and stylistic refinement. All mathematical contents were thoroughly reviewed and verified by the authors.