A FINSLERIAN APPROACH FOR EMBEDDING DIRECTED DATA

Anonymous authors
Paper under double-blind review

ABSTRACT

Many datasets carry an intrinsic directionality: citations point backward in time, cells differentiate along lineages, and traffic follows preferred routes. Spectral embedding methods, including most of their extensions to directed graphs, discard this information: they symmetrize the data and map it into a Euclidean space where asymmetry cannot be represented. We instead model directed data as sampled from a Finsler manifold, whose distance depends on the direction of travel, and study the kernel operator built from this asymmetric distance. Through a moment expansion of this operator, we show that its symmetric and antisymmetric parts separate geometry from direction. As the bandwidth of the kernel vanishes, the symmetric part converges to a weighted Laplacian, recovering diffusion maps in the Riemannian case, while the antisymmetric part converges to a first-order transport operator that encodes the directionality. We prove that the corresponding graph operators, built from finitely many samples, converge uniformly and almost surely to these limits. For Randers metrics, this vector field is explicit and yields an embedding algorithm recovering both the manifold structure, from the spectrum of the symmetric part, and the underlying drift. We illustrate the approach on synthetic directed graphs and point-clouds.

1 INTRODUCTION

A citation never points to the future, a cell rarely un-differentiates, influence in a social network is rarely returned in equal measure, and traffic or commodity flows between two locations are rarely symmetric. These are just some of the many examples of *directed data*, i.e. data with asymmetric relationships between entities. This directionality can be given explicitly, as in directed graphs (Debaussart-Joniec et al., 2026; Chung, 2005; Cucuringu et al., 2020; Sevi et al., 2026), or arise from points in a continuous space, as in RNA velocity (Bergen et al., 2021) or lineage data (Zweig et al., 2026). For such data, *Finsler geometry*, an extension of Riemannian geometry allowing direction-dependent metrics, offers a natural mathematical language. Yet, Finsler geometry remains largely unexplored in machine learning or manifold learning.

Graph Laplacians and diffusion operators underlie many of the tools we rely on for structured data, including spectral clustering (Von Luxburg, 2007), dimensionality reduction (Donoho & Grimes, 2003), semi-supervised learning, and graph neural networks (Kipf & Welling, 2016). Their theoretical grounding rests on a convergence result: when data is sampled from a compact Riemannian manifold, the graph Laplacian converges, as the sample size grows and the kernel bandwidth shrinks, to the Laplace–Beltrami operator (Coifman & Lafon, 2006; Belkin & Niyogi, 2001; Lafon, 2004; Hein et al., 2007; Singer, 2006; Calder & Trillos, 2022; Peoples & Harlim, 2025). This theory is symmetric by construction, as the kernel is a function of a symmetric distance.

Methods for directed graphs generally restore symmetry somewhere along the way, through the stationary distribution of a random walk (Chung, 2005; Zhou et al., 2003), an explicit symmetrization of the adjacency matrix (Satuluri & Parthasarathy, 2011), an encoding of the direction in the complex phase of a Hermitian

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

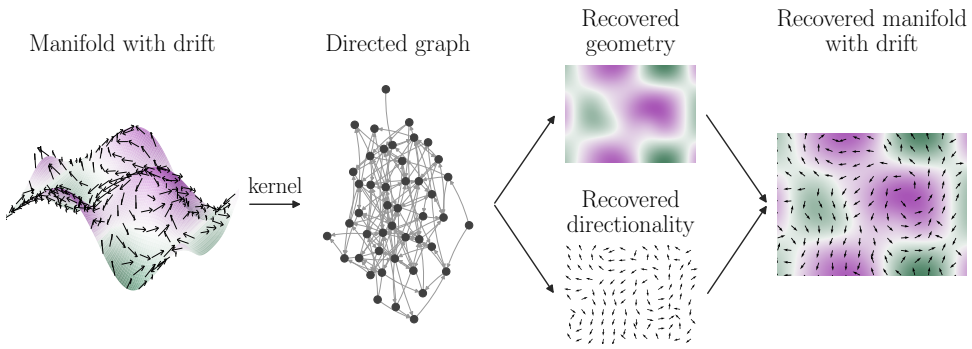


Figure 1: Recovering geometry and directionality from Finsler manifolds. A kernel built from the Finsler distance turns sampled points into a directed graph. The operators $\mathscr{L}^s$ and $\mathscr{L}^a$ of Theorem 2 recover the geometry and the directional structure, and are merged into a Randers metric on the embedding.

matrix (Cucuringu et al., 2020; Fanuel et al., 2018; Zhang et al., 2021), or vertex measures (Sevi et al., 2026; Debaussart-Joniec et al., 2026). Directed embedding methods (Chen et al., 2007; Ou et al., 2016) keep the asymmetry in the objective, but, like the spectral ones, return points of a Euclidean space, in which the directionality of the data is lost. Closest to our work, Perrault-Joncas & Meila (2011) and Yuan et al. (2022) take continuous limits of Laplacian-type operators built from an asymmetric kernel, and obtain a diffusion term together with an advection term. There, as in other uses of asymmetric kernels (Wu et al., 2010; He et al., 2023; Tao et al., 2024), the asymmetry is introduced by hand, without an underlying geometry.

Finsler geometry is classical in mathematics (Finsler, 1918; Yeon, 2015; Bao et al., 2012; Shen, 2001; Ohta, 2021), and arises wherever the cost of moving depends on the direction of travel (Zermelo, 1931; Markvorsen, 2016; Gahtan et al., 2026; Asanjarani, 2021; Melonakos et al., 2008; Chen et al., 2024; Pfeifer, 2019). In machine learning, it has been used to model anisotropic or asymmetric structures (Lopez et al., 2021; Weber et al., 2024; Pouplin et al., 2023; Zweig et al., 2026; Shaska, 2025), and Dagès et al. (2025); Dagès et al. (2026) generalize multidimensional scaling to a Finslerian embedding space. These methods learn direction-dependent representations, but the Finsler metric usually has a prescribed form (Weber et al., 2024; Dagès et al., 2025).

This work proposes a principled approach for representing directed data, grounded in Finsler geometry: we model the data as sampled from a Finsler manifold, so that the asymmetry comes from an arbitrary underlying geometry rather than from a hand-crafted kernel, and we recover this geometry, through a Randers approximation of the metric, in the embedding space (see Figure 1). Our contributions are as follows.

- **Continuum limits** (Section 3). Building on the convergence theory of graph Laplacians, we derive a moment expansion of the Finsler kernel operator. Its symmetric part converges to a weighted Laplacian of the Binet–Legendre metric, recovering diffusion maps in the Riemannian case, while its antisymmetric part converges to the derivative along the centroid field of the unit ball, which encodes the directionality.
- **Consistency** (Section 4). We prove that the corresponding graph operators, built from finitely many samples, converge uniformly and almost surely to these limits, with an explicit rate.
- **Directed embedding** (Section 5). We introduce the moment-Randers metric, a Randers approximation of the Finsler metric with the same limiting operators, and show that it can be estimated from the two graph operators alone. This yields an algorithm returning both an embedding and a Randers metric on it.
- **Experiments** (Section 6). We illustrate the approach on point-clouds sampled from Randers and Matsumoto manifolds, as well as on directed stochastic block models.

2 BACKGROUND AND SETTING

2.1 BASIC CONCEPTS

Notations are summarized in Tables 2 and 3, and Appendix B gives a self-contained introduction to Finsler geometry. We refer to Shen (2001) and Ohta (2021) for additional details.

Let $\mathscr{M}$ be an m-dimensional manifold. A *Finsler metric* is a map $\mathscr{F} : T\mathscr{M} \to [0,\infty)$, smooth and positive away from the zero section, that is positively homogeneous, $\mathscr{F}(x,tv) = t\mathscr{F}(x,v)$ for $t > 0$, sub-additive in v, and strongly convex. It plays on each tangent space $T_x\mathscr{M}$ the role that a norm plays on $\mathbb{R}^m$ and, exactly as in the Riemannian case, it defines the length of a curve and a distance $\mathrm{dst}_{\mathscr{F}}(x,y)$ as the infimum of the lengths of the curves joining x to y. Riemannian metrics are the special case where $\mathscr{F}(x,\cdot)$ is induced by an inner product. The essential difference is that homogeneity is only required for *positive* scalars: in general $\mathscr{F}(x,v) \neq \mathscr{F}(x,-v)$, so that the distance $\mathrm{dst}_{\mathscr{F}}$ satisfies the triangle inequality but not symmetry, since traveling from x to y need not cost as much as traveling back. The *reverse metric* $\mathscr{F}(x,-v)$ is again a Finsler metric, whose distance is $(x,y) \mapsto \mathrm{dst}_{\mathscr{F}}(y,x)$ (Lemma 20).

This work relies on a subclass of Finsler metrics and one canonical construction. *Randers metrics* (Randers, 1941) are a natural generalization of Riemannian metrics. For a Riemannian metric $\alpha(x,v) = \sqrt{v^\mathsf{T}\mathbf{A}(x)v}$ and a vector field b on $\mathscr{M}$, the Randers metric is defined as

$$\mathscr{F}(x,v) = \sqrt{v^\mathsf{T}\mathbf{A}(x)v} + b(x)^\mathsf{T}v, \qquad \|b(x)\|_{\mathbf{A}^{-1}} < 1. \tag{1}$$

The constraint on b ensures that $\mathscr{F}$ is positive. Its indicatrix $\{v \in T_x\mathscr{M} : \mathscr{F}(x,v) = 1\}$ is an ellipsoid whose center is shifted in a direction determined by $b(x)$, so that all the directional information is carried by the single vector field b. We write $\mathscr{B}_x = \{v \in T_x\mathscr{M} : \mathscr{F}(x,v) \leq 1\}$ for the unit ball of $\mathscr{F}$ in the tangent space $T_x\mathscr{M}$. The second moment of the uniform distribution on $\mathscr{B}_x$ yields a Riemannian metric canonically attached to $\mathscr{F}$: the *Binet–Legendre metric* (Matveev & Troyanov, 2012),

$$g^{ij}_{\mathrm{BL}}(x) = \frac{(m+2)}{\lambda_x(\mathscr{B}_x)} \int_{\mathscr{B}_x} v^i v^j \mathrm{d}\lambda_x(v), \tag{2}$$

which is smooth whenever $\mathscr{F}$ is smooth, and coincides with $\mathbf{A}$ when $\mathscr{F}$ is Riemannian. The first moment of the same distribution is the *centroid* of the unit ball,

$$\mathbf{c}(x) = \frac{1}{\lambda_x(\mathscr{B}_x)} \int_{\mathscr{B}_x} v \mathrm{d}\lambda_x(v) \in T_x\mathscr{M}, \tag{3}$$

which defines a vector field on $\mathscr{M}$, the *centroid field*. It vanishes whenever $\mathscr{F}$ is reversible, since the unit ball is then symmetric about the origin, and thus measures the asymmetry of $\mathscr{F}$ at first order; for Randers metrics, it is an explicit function of $\mathbf{A}$ and b (Proposition 43). The pair $(\mathbf{c}, g_{\mathrm{BL}})$, i.e. the first two moments of the unit ball, will carry respectively the direction and the geometry recovered by our operators.

Unlike a Riemannian manifold, a Finsler manifold carries no canonical volume. In this work, we use the *Busemann–Hausdorff measure* $\mathfrak{m}_{\mathrm{BH}}$ (Definition 17). In a chart, with $\mathrm{d}x$ the Lebesgue measure on $\mathscr{M}$ and λ_x the Lebesgue measure on $T_x\mathscr{M}$ in the coordinate basis $(\partial_i|_x)$, it reads

$$\mathrm{d}\mathfrak{m}_{\mathrm{BH}}(x) = \sigma_{\mathrm{BH}}(x)\mathrm{d}x, \qquad \sigma_{\mathrm{BH}}(x) = \frac{\mathrm{Vol}_{\mathrm{Euc}}(\mathbb{B}^m)}{\lambda_x(\mathscr{B}_x)}, \tag{4}$$

where $\mathrm{Vol}_{\mathrm{Euc}}(\mathbb{B}^m)$ is the volume of the Euclidean unit ball of $\mathbb{R}^m$. Although $\mathrm{d}x$ and λ_x are chart-dependent, $\mathfrak{m}_{\mathrm{BH}}$ is not. When $\mathscr{F}$ is Riemannian, $\mathfrak{m}_{\mathrm{BH}}$ coincides with the Riemannian volume. We assume that the data are N i.i.d. samples $X_1, \ldots, X_N$ from a probability measure $\mathbb{P}$ on $\mathscr{M}$, and denote by ρ its density with respect to $\mathfrak{m}_{\mathrm{BH}}$, so that $\mathrm{d}\mathbb{P}(x) = \rho(x)\mathrm{d}\mathfrak{m}_{\mathrm{BH}}(x)$.

2.2 Our setting

We make the following assumptions throughout:

- (A1) The kernel $K : \mathbb{R}_+ \to \mathbb{R}_+$ is a smooth, positive, monotone decreasing function with bounded derivative, and has a sub-exponential tail: there exist constants $C_K, \nu_K > 0$ such that for any $r \in \mathbb{R}_+$, $K(r) \leq C_K e^{-\nu_K r}$.
- (A2) The manifold $\mathcal{M}$ is smooth, connected, compact and *without boundary*[1], and the Finsler metric $\mathcal{F}$ is smooth and positive on the slit tangent bundle $T\mathcal{M} \setminus \{0\}$. Moreover, the sampling density ρ is smooth and bounded away from zero by a constant $\rho_{\min} > 0$.

Given K satisfying (A1), we define the kernel $W : \mathcal{M} \times \mathcal{M} \to \mathbb{R}_+$ and its out/in-degrees as:

$$W(x,y) = \frac{1}{\varepsilon^m} K\left(\frac{\mathrm{dst}_{\mathcal{F}}(x,y)}{\varepsilon}\right), \quad d(x) = \int_{\mathcal{M}} W(x,y)\mathrm{d}\mathbb{P}(y), \quad d'(x) = \int_{\mathcal{M}} W(y,x)\mathrm{d}\mathbb{P}(y),$$

where ε is the kernel bandwidth. We further assume that the degrees are bounded below: for any $x \in \mathcal{M}$, $d(x), d'(x) > d_{\min} > 0$. This is a standard assumption in the literature (Hein et al., 2007), and it is satisfied when the kernel is sub-exponential and ρ is positive. We study the kernel through its associated forward and backward transport operators $\mathcal{G}_{\mathcal{F}}$ and $\mathcal{G}'_{\mathcal{F}}$. For any $f : \mathcal{M} \to \mathbb{R}$ smooth enough, these are defined as:

$$\mathcal{G}_{\mathcal{F}}[f](x) = \int_{\mathcal{M}} W(x,y) f(y)\mathrm{d}\mathfrak{m}_{\mathrm{BH}}(y), \quad \mathcal{G}'_{\mathcal{F}}[f](x) = \int_{\mathcal{M}} W(y,x) f(y)\mathrm{d}\mathfrak{m}_{\mathrm{BH}}(y).$$

The backward operator is the transport operator of the reversed Finsler metric, and is the $L^2(\mathcal{M}, \mathfrak{m}_{\mathrm{BH}})$-adjoint of $\mathcal{G}_{\mathcal{F}}$, so that $\mathcal{G}_{\mathcal{F}}$ is self-adjoint precisely when $\mathcal{F}$ is reversible (Proposition 19, proved in Appendix D.1).

In practice, we only observe samples from $\mathbb{P}$, whose density ρ is unknown. To reduce its influence, we use a θ-normalization of the kernel W, similar to Coifman & Lafon (2006):

$$W^{(\theta)}(x,y) = \frac{W(x,y)}{q_\varepsilon(x)^\theta q_\varepsilon(y)^\theta}, \quad \mathcal{G}^{(\theta)}_{\mathcal{F}}[f](x) = \int_{\mathcal{M}} W^{(\theta)}(x,y) f(y)\mathrm{d}\mathfrak{m}_{\mathrm{BH}}(y),$$

where $q_\varepsilon(x) = (d'(x) + d(x))/2$ is a normalization factor, and $\theta \in [0,1]$. The backward operator is defined using $W^{(\theta)}(y,x)$ instead. In general, the forward and backward operators differ; Section 3 exploits this difference to split the operator into a symmetric and an antisymmetric part.

3 Continuum limits: separating geometry from direction

Our starting point is a moment expansion of the transport operator $\mathcal{G}_{\mathcal{F}}$, the Finslerian counterpart of the expansions used to derive continuum limits of graph Laplacians (Coifman & Lafon, 2006; Singer, 2006). The overall procedure is illustrated in Figure 1.

Theorem 1. *For any $f : \mathcal{M} \to \mathbb{R}$ smooth enough, we have*

$$\mathcal{G}_{\mathcal{F}}[f](x) = m_0 f(x) + \varepsilon \mathcal{A}_1[f](x) + \frac{\varepsilon^2}{2}\mathcal{A}_2[f](x) + \mathcal{O}(\varepsilon^3),$$

where $\mathcal{A}_1$ and $\mathcal{A}_2$ are first- and second-order differential operators, whose coefficients are explicitly given in Appendix D.2. The leading term m_0 is the zeroth moment of the kernel K, and does not depend on x.

[1] This assumption is standard in the study of convergence of graph Laplacians. When $\mathcal{M}$ has a boundary, the analysis still holds for points "far enough" from it (Coifman & Lafon, 2006).

We now split $\mathcal{G}_{\mathcal{F}}^{(\theta)}$ into its symmetric and antisymmetric parts $\mathcal{G}_{\mathcal{F}}^{(\theta,s)}$ and $\mathcal{G}_{\mathcal{F}}^{(\theta,a)}$, which is where the geometry and the direction separate:

$$\mathcal{G}_{\mathcal{F}}^{(\theta)} = \mathcal{G}_{\mathcal{F}}^{(\theta,s)} + \mathcal{G}_{\mathcal{F}}^{(\theta,a)}, \quad \text{where} \quad \mathcal{G}_{\mathcal{F}}^{(\theta,s)} = (\mathcal{G}_{\mathcal{F}}^{(\theta)} + \mathcal{G}_{\mathcal{F}}^{(\theta)'})/2 \quad \text{and} \quad \mathcal{G}_{\mathcal{F}}^{(\theta,a)} = (\mathcal{G}_{\mathcal{F}}^{(\theta)} - \mathcal{G}_{\mathcal{F}}^{(\theta)'})/2.$$

Define the associated transport operators $\mathcal{P}^{(\theta,s)}$ and $\mathcal{P}^{(\theta,a)}$ as

$$\mathcal{P}^{(\theta,\bullet)}[f] = \frac{1}{\mathcal{G}_{\mathcal{F}}^{(\theta,s)}[\rho]} \left(\mathcal{G}_{\mathcal{F}}^{(\theta,\bullet)}[\rho f] - f \mathcal{G}_{\mathcal{F}}^{(\theta,\bullet)}[\rho] \right),$$

for $\bullet \in \{s, a\}$. The following theorem gives their limits as $\varepsilon \to 0$, after rescaling by ε^{-2} and ε^{-1}, respectively.

Theorem 2. *The operators $\mathcal{P}^{(\theta,\bullet)}$ admit limiting operators $\mathcal{L}^{\bullet}$ as $\varepsilon \to 0$:*

$$\mathcal{L}^{s} f := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \mathcal{P}^{(\theta,s)}[f] = \frac{c_2}{\rho^{2(1-\theta)}} \operatorname{div}_{BH} \left(\rho^{2(1-\theta)} \nabla_{BL} f \right), \tag{5}$$

$$\mathcal{L}^{a} f := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \mathcal{P}^{(\theta,a)}[f] = \tilde{m}_1^i \partial_i f = c_1 \, \mathbf{c}^i \partial_i f, \tag{6}$$

where $\tilde{m}_1(x)$ is the normalized first moment of the kernel K over the unit ball $\mathcal{B}_x$ (Definition 21), $\mathbf{c}$ is the centroid field Equation (3), c_1 and c_2 are constants depending on the kernel and the dimension of the manifold, ∇_{BL} is the gradient with respect to the Binet–Legendre metric g_{BL} of $\mathcal{F}$, and div_{BH} is the divergence with respect to the Busemann–Hausdorff measure $\mathfrak{m}_{\mathrm{BH}}$. In particular, for any Finsler metric $\mathcal{F}$, $\mathcal{L}^s$ is an elliptic operator, self-adjoint on $L^2(\mathcal{M}, \rho^{2(1-\theta)}\mathfrak{m}_{\mathrm{BH}})$.

This result is proved in Appendix D.3, and the constants are given in Proposition 35; for instance, $c_1 = (m+1)\mu_1/(m\mu_0)$ and $c_2 = \mu_2/(2m\mu_0)$ with $\mu_n = \int_0^\infty K(r) r^{m+n-1} \mathrm{d}r$, see Section D.9 for more details. The symmetric limit is thus a weighted Laplacian, whose metric is the Binet–Legendre metric of $\mathcal{F}$ and whose reference measure is $\rho^{2(1-\theta)}\mathfrak{m}_{\mathrm{BH}}$. When $\theta = 1$, it does not depend on the sampling density ρ, which is a desirable property for the embedding algorithm. In contrast, $\theta = 0$ keeps the whole drift induced by ρ, and $\theta = 1/2$ halves it (see also Section D.4). The antisymmetric limit is the derivative along the centroid field: the first moment of the unit ball carries the direction, and the second moment the geometry. Note also that the finite-ε operator $\mathcal{P}^{(\theta,s)}$ is always diagonalizable with real eigenvalues.

The divergence div_{BH} is taken with respect to the Busemann–Hausdorff measure, and not with respect to the Riemannian volume of g_{BL}. The two coincide up to a constant factor for *Berwald* metrics, which are those whose connection depends on the position only and not on the direction (Definition 15). In that case, the symmetric limit simplifies as follows, proven in Appendix D.5.

Corollary 3 (Berwald and Riemannian metrics)**.** *If $\mathcal{F}$ is a Berwald metric, then*

$$\mathcal{L}^{s} f = c_2 \left(\Delta_{BL} f + 2(1-\theta) \langle \nabla \log \rho, \nabla f \rangle_{BL} \right), \tag{7}$$

where Δ_{BL} is the Laplace–Beltrami operator of g_{BL} and $\langle \nabla u, \nabla f \rangle_{BL} = g_{\mathrm{BL}}^{ij} \partial_i u \partial_j f$. If moreover $\mathcal{F}$ is Riemannian, i.e. $\mathcal{F}(x,v) = \sqrt{v^\top \mathbf{A}(x) v}$, then $g_{\mathrm{BL}} = \mathbf{A}$, $\mathcal{L}^a = 0$, and Equation (7) is the θ-normalized diffusion generator of Coifman & Lafon (2006).

4 Discrete operators and convergence guarantees

In the discrete setting, where the manifold is only accessed through N points $X_1, \ldots, X_N$ sampled i.i.d. from $\mathbb{P}$, the construction follows the same principle as the continuous one. From the empirical kernel matrix $\mathbf{W}_N(i,j) = W(X_i, X_j)$, we define the out- and in-degrees $\mathbf{D}_N(i,i) = \sum_j \mathbf{W}_N(i,j)$ and

$\mathbf{D}'_N(i,i) = \sum_j \mathbf{W}_N(j,i)$, and their half-sum $\mathbf{Q}_N = (\mathbf{D}_N + \mathbf{D}'_N)/2$. The θ-normalized matrix and its symmetric and antisymmetric parts are then

$$\mathbf{W}_N^{(\theta)} = \mathbf{Q}_N^{-\theta}\mathbf{W}_N\mathbf{Q}_N^{-\theta}, \qquad \mathbf{W}_N^{(\theta,\bullet)} = \tfrac{1}{2}\big(\mathbf{W}_N^{(\theta)} \pm (\mathbf{W}_N^{(\theta)})^\mathsf{T}\big),$$

with the sign $+$ for $\bullet = s$ and $-$ for $\bullet = a$. We write $\mathbf{D}_N^{(\theta,\bullet)}$ for the diagonal matrix of row sums of $\mathbf{W}_N^{(\theta,\bullet)}$, which discretizes $\mathscr{G}_{\mathscr{F}}^{(\theta,\bullet)}[\rho]$. Finally, for $\mathbf{f} = (f(X_1),\dots,f(X_N))^\mathsf{T} \in \mathbb{R}^N$, we define the discrete operators:

$$\mathbf{P}^{(\theta,\bullet)}[f] = \big(\mathbf{D}_N^{(\theta,s)}\big)^{-1}\Big(\mathbf{W}_N^{(\theta,\bullet)}\mathbf{f} - \mathbf{D}_N^{(\theta,\bullet)}\mathbf{f}\Big), \qquad \text{with} \qquad \mathbf{L}^s[f] = \frac{\mathbf{P}^{(\theta,s)}[f]}{\varepsilon^2}, \quad \mathbf{L}^a[f] = \frac{\mathbf{P}^{(\theta,a)}[f]}{\varepsilon}. \tag{8}$$

Both parts are normalized by the same $\mathbf{D}_N^{(\theta,s)}$, each is recentered by its own degree $\mathbf{D}_N^{(\theta,\bullet)}$, and each is rescaled by the order at which its limit appears in Theorem 2.

Theorem 4. *Let $f \in \mathscr{C}^3(\mathscr{M})$ and let the bandwidth $\varepsilon = \varepsilon(N)$ vary with the number of samples in such a way that $\varepsilon(N) \to 0$ and $N\varepsilon(N)^{m+2}/\log N \to \infty$ as $N \to \infty$. Then the discrete operators converge to the continuous ones, uniformly and almost surely: for $\bullet \in \{s,a\}$,*

$$\max_{1\le i\le N} \big|\mathbf{L}^\bullet[f](X_i) - \mathscr{L}^\bullet f(X_i)\big| \xrightarrow[N\to\infty]{a.s.} 0.$$

This result is proved in Appendix D.6, where the convergence rate is given explicitly.

5 Moment-Randers metric and directed embedding

Theorem 2 shows that the operators $\mathscr{L}^s$ and $\mathscr{L}^a$ depend on the Finsler metric $\mathscr{F}$ only through its Binet–Legendre metric, its Busemann–Hausdorff measure, and its centroid field. This motivates the definition of a Randers metric that has the same Binet–Legendre metric and centroid field as $\mathscr{F}$, and that can be estimated from the graph operators of Section 4. Results stated in this section are proved in Appendix D.8, where additional details are also provided.

5.1 Moment-Randers approximation

Definition 5. *(Moment-Randers metric) Let $\mathscr{F}$ be a Finsler metric on $\mathscr{M}$ with centroid field $\mathbf{c}$ and Binet–Legendre metric g_{BL}. Set $\mathbf{H}^{-1} = g_{\mathrm{BL}}^{-1} - (m+2)\mathbf{c}\mathbf{c}^\mathsf{T}$, and $\lambda = 1 - \|\mathbf{c}\|_{\mathbf{H}}^2$. If $\|\mathbf{c}(x)\|_{\mathbf{H}} < 1$ for all $x \in \mathscr{M}$, the* moment-Randers metric *of $\mathscr{F}$ is the Randers metric Equation (1) with data $(\mathbf{A}_R, b_R)$:*

$$\mathscr{F}_R(x,v) = \sqrt{v^\mathsf{T}\mathbf{A}_R(x)v} + b_R(x)^\mathsf{T}v, \qquad \mathbf{A}_R = \frac{\lambda\mathbf{H} + \mathbf{H}\mathbf{c}\mathbf{c}^\mathsf{T}\mathbf{H}}{\lambda^2}, \qquad b_R = -\frac{\mathbf{H}\mathbf{c}}{\lambda}.$$

Its drift satisfies $\|b_R\|_{\mathbf{A}_R^{-1}} = \|\mathbf{c}\|_{\mathbf{H}}$, so the admissibility constraint of Equation (1) is exactly the hypothesis above. $\mathscr{F}_R$ has the same Binet–Legendre metric and centroid field as $\mathscr{F}$, but can have a different Busemann–Hausdorff measure[2], and if $\mathscr{F}$ is a Randers metric, then $\mathscr{F}_R = \mathscr{F}$ as this is just the navigation map formulation (Section D.8). This means that $\mathscr{L}^a$ is identical for both metrics, that $\mathscr{L}^s$ only differs through the reference measure, and that the moment-Randers metric is a moment-matching Randers approximation of $\mathscr{F}$. The construction of $\mathbf{H}$ relies on g_{BL} and $\mathbf{c}$. We relate the well-definedness of $\mathbf{H}$ and of the moment-Randers metric to the norm of $\mathbf{c}$ in the Binet–Legendre metric, which, as shown below, is accessible from the data.

Proposition 6. *Let $\mathscr{F}$ be a Finsler metric and let $x \in \mathscr{M}$ be such that $\|\mathbf{c}(x)\|_{g_{\mathrm{BL}}}^2 < (m+2)^{-1}$. Then the matrix $\mathbf{H}(x)$ is well defined and positive definite, and*

$$\|\mathbf{c}(x)\|_{g_{\mathrm{BL}}}^2 = \frac{\|\mathbf{c}(x)\|_{\mathbf{H}}^2}{1 + (m+2)\|\mathbf{c}(x)\|_{\mathbf{H}}^2}, \qquad \textit{hence} \qquad \|\mathbf{c}(x)\|_{\mathbf{H}} < 1 \iff \|\mathbf{c}(x)\|_{g_{\mathrm{BL}}}^2 < (m+3)^{-1}.$$

[2] As detailed in Appendix D.8, the symmetric limits of both metrics have the same principal components.

5.2 Estimation through an embedding

The quantities $\mathbf{c}$ and g_{BL} cannot be observed directly. We therefore estimate them, through an embedding, from the operators $\mathscr{L}^s$ and $\mathscr{L}^a$, which are themselves approximated by the graph operators $\mathbf{L}^s$ and $\mathbf{L}^a$ (Theorem 4). Consider an embedding $\Psi(x) = (\psi_1(x), \dots, \psi_\ell(x))$ and its Jacobian at x, $J(x) \in \mathbb{R}^{\ell \times m}$, assumed of full column rank m. By Equation (6), the push-forward of $\mathbf{c}$ by Ψ is recovered from the antisymmetric operator as

$$V(x) := \mathscr{L}^a[\Psi](x) = c_1 J(x)\mathbf{c}(x).$$

Proposition 7. *The carré du champ operator Γ associated with $\mathscr{L}^s$,*

$$\Gamma(x)^{kl} = \frac{1}{2}\left(\mathscr{L}^s[\psi_k\psi_l](x) - \psi_k(x)\mathscr{L}^s[\psi_l](x) - \psi_l(x)\mathscr{L}^s[\psi_k](x)\right),$$

recovers the Binet–Legendre metric up to the embedding: $\Gamma(x) = c_2 J(x) g_{\mathrm{BL}}(x)^{-1} J(x)^\mathsf{T}$, where c_2 is the constant of Theorem 2.

Combining these results, the moment-Randers metric of $\mathscr{F}$ is well defined at x if and only if $\|\mathbf{c}(x)\|^2_{g_{\mathrm{BL}}} = \frac{c_2}{c_1^2}\|V(x)\|^2_{\Gamma^+} < (m+3)^{-1}$, where Γ^+ is the pseudo-inverse of Γ. This criterion only involves $\mathscr{L}^s$ and $\mathscr{L}^a$, and allows us to define the embedded moment-Randers metric $\hat{\mathscr{F}}_R$.

Definition 8. *The embedded moment-Randers metric $\hat{\mathscr{F}}_R$ is defined as the moment-Randers metric (Definition 5) associated with the estimated Binet–Legendre metric $\hat{g}_{BL}$ and the estimated centroid of the unit tangent ball $\hat{\mathbf{c}}$. Let*

$$\hat{g}_{BL}(x)^{-1} = \Gamma(x), \qquad \hat{\mathbf{c}}(x) = \frac{\sqrt{c_2}}{c_1} V(x),$$
$$\hat{\mathbf{H}}^{-1}(x) = \Gamma(x) - (m+2)\frac{c_2}{c_1^2} V(x)V(x)^\mathsf{T}, \qquad \hat{\lambda}(x) = 1 - \|\hat{\mathbf{c}}(x)\|^2_{\hat{\mathbf{H}}},$$

so that the embedded moment-Randers metric is given by

$$\hat{\mathscr{F}}_R(x,v) = \frac{1}{\hat{\lambda}(x)}\left(\sqrt{\hat{\lambda}(x)\|v\|^2_{\hat{\mathbf{H}}} + \langle \hat{\mathbf{c}}(x), v\rangle^2_{\hat{\mathbf{H}}}} - \langle \hat{\mathbf{c}}(x), v\rangle_{\hat{\mathbf{H}}}\right).$$

Proposition 9. *If $\|\hat{\mathbf{c}}(x)\|_{\hat{g}_{BL}} < 1/\sqrt{m+3}$ and J is of full column rank, then $\hat{\mathscr{F}}_R$ is related to the original moment-Randers metric $\mathscr{F}_R$ by*

$$\hat{\mathscr{F}}_R(x, J(x)v) = \frac{1}{\sqrt{c_2}}\mathscr{F}_R(x,v), \quad \forall v \in T_x\mathscr{M}.$$

In particular, $\hat{\mathscr{F}}_R$ is a valid Randers metric on $J(T_x\mathscr{M})$.

5.3 Embedding algorithm

Algorithm 1 summarizes the procedure. The embedding Ψ can be any classical embedding, e.g. Isomap (Tenenbaum et al., 2000) or the leading eigenvectors of $\mathbf{L}^s$, in which case J has full column rank for ℓ large enough. $\mathbf{V}_i$ and Γ_i are the discrete counterparts of V and Γ, obtained by replacing $\mathscr{L}^\bullet$ with $\mathbf{L}^\bullet$; here $Y^k \in \mathbb{R}^N$ is the k-th column of $\mathbf{Y}$. When the data is given as a weighted directed graph (Section 6), $\mathbf{W}_N$ is its adjacency matrix.

Algorithm 1 Finsler directed embedding

Require: Directed distances $\mathrm{dst}_{\mathscr{F}}(X_i, X_j)$ between samples, or a weighted directed graph with adjacency matrix $\mathbf{W}_N$; kernel K, bandwidth ε, normalization $\theta \in [0,1]$, intrinsic dimension m, embedding dimension ℓ.

Ensure: Embedded samples $Y_i \in \mathbb{R}^\ell$, drifts $\mathbf{V}_i$, and Randers metrics $\hat{\mathscr{F}}_R(Y_i, \cdot)$.

$\mathbf{W}_N(i,j) \leftarrow \varepsilon^{-m} K\big(\mathrm{dst}_{\mathscr{F}}(X_i, X_j)/\varepsilon\big)$ {skipped if $\mathbf{W}_N$ is given}
$\mathbf{Q}_N \leftarrow (\mathbf{D}_N + \mathbf{D}'_N)/2, \quad \mathbf{W}_N^{(\theta)} \leftarrow \mathbf{Q}_N^{-\theta} \mathbf{W}_N \mathbf{Q}_N^{-\theta}$
$\mathbf{W}_N^{(\theta,s)}, \mathbf{W}_N^{(\theta,a)} \leftarrow$ symmetric and antisymmetric parts of $\mathbf{W}_N^{(\theta)}$
Build $\mathbf{L}^s$ and $\mathbf{L}^a$ from Equation (8)
$\mathbf{Y} \leftarrow (\Psi(X_1), \ldots, \Psi(X_N))^\mathsf{T} \in \mathbb{R}^{N \times \ell}$ {e.g. leading eigenvectors of $\mathbf{L}^s$}
for $i = 1, \ldots, N$ **do**
  $\mathbf{V}_i \leftarrow (\mathbf{L}^a \mathbf{Y})_i, \quad \Gamma_i^{kl} \leftarrow \frac{1}{2}\big(\mathbf{L}^s[Y^k Y^l]_i - Y_i^k \mathbf{L}^s[Y^l]_i - Y_i^l \mathbf{L}^s[Y^k]_i\big)$
  $\hat{\mathbf{c}}_i \leftarrow \frac{\sqrt{c_2}}{c_1} \mathbf{V}_i, \quad \hat{\mathbf{H}}_i^{-1} \leftarrow \Gamma_i - (m+2)\hat{\mathbf{c}}_i \hat{\mathbf{c}}_i^\mathsf{T}$
  if $\|\hat{\mathbf{c}}_i\|^2_{\Gamma_i^+} < (m+3)^{-1}$ **then**
    Set $\hat{\mathscr{F}}_R(Y_i, \cdot)$ as in Definition 8
  end if
end for

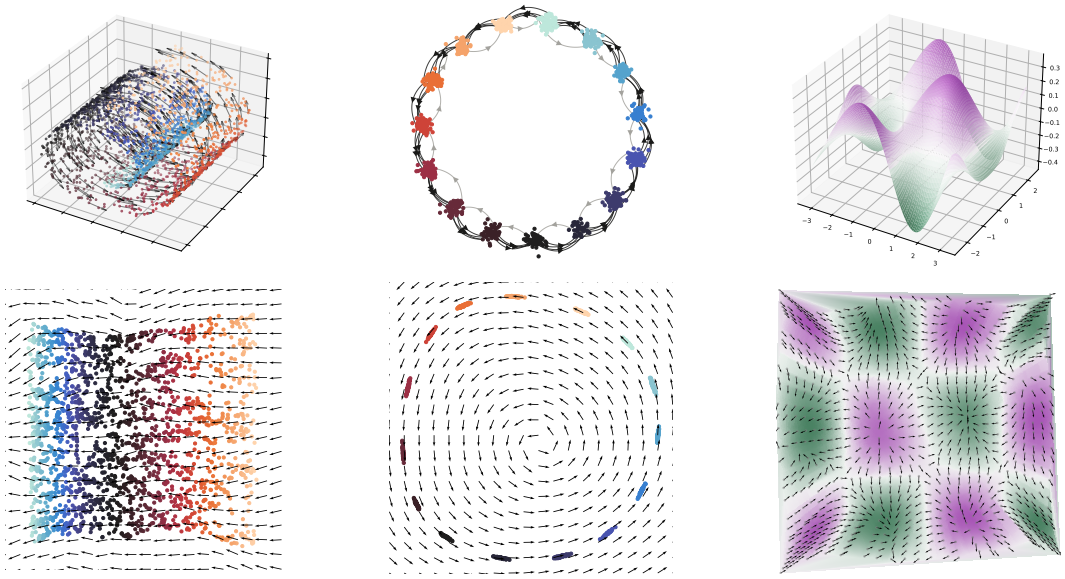
Table 1: Swiss roll. Relative error on $\|\hat{\mathbf{c}}\|^2_{\hat{g}_{BL}}$ and cosine between V_i and $c_1 J_i \mathbf{c}(X_i)$ (medians over samples), and admissible fraction; medians over 3 seeds. Admissible means $\|\hat{\mathbf{c}}\|^2_{\hat{g}_{BL}} < (m+3)^{-1}$.

	Rel. error $\|\hat{\mathbf{c}}\|^2_{\hat{g}_{BL}}$			$\cos(V_i, J_i\mathbf{c})$			Admissible		
$\|b\|_{\mathbf{A}^{-1}}$	$N = 1000$	2000	4000	$N = 1000$	2000	4000	$N = 1000$	2000	4000
0.1	0.17	0.14	0.13	1.00	1.00	1.00	1.00	1.00	1.00
0.3	0.15	0.13	0.12	1.00	1.00	1.00	1.00	1.00	1.00
0.5	0.13	0.11	0.10	1.00	1.00	1.00	0.98	0.99	0.99
0.7	0.11	0.10	0.09	1.00	1.00	1.00	0.89	0.91	0.91
0.9	0.09	0.08	0.07	1.00	1.00	1.00	0.46	0.47	0.50

6 Experiments

Swiss roll dataset. We first consider a Randers metric on the Swiss roll, a 2D manifold embedded in $\mathbb{R}^3$, whose Riemannian part is supported on the manifold and whose drift b is tangent to it (Appendix C.1). We sample $N = 2000$ points from the manifold and construct a k-NN graph with $k = 10$ neighbors. With Isomap as Ψ, Figure 2a shows that we recover both the geometry of the manifold and the directionality of the edges. As $\mathbf{c}$ and g_{BL} are known in closed form for a Randers metric (Proposition 43), Table 1 evaluates the estimated strength of the asymmetry for varying $\|b\|_{\mathbf{A}^{-1}}$ and N, on a radius graph and with J_i estimated by finite differences (Appendix C.1). The cosine between the recovered and true drifts is 1.00 in every configuration, and the relative error on $\|\hat{\mathbf{c}}\|^2_{\hat{g}_{BL}}$ stays between 0.07 and 0.17. It decreases with N at every drift strength, although slowly. The bandwidth ε is the median Finsler distance from a sample to its 10-th nearest neighbor, and the graph construction is detailed in Appendix C.1. The rejections at large $\|b\|_{\mathbf{A}^{-1}}$ are expected. All samples are admissible for the true metric, since $\mathscr{F}_R = \mathscr{F}$ for a Randers metric, but $\|\mathbf{c}\|^2_{g_{\mathrm{BL}}}$ approaches the threshold $(m+3)^{-1}$ as the drift grows, so that small estimation errors suffice to cross it.

Approximating Matsumoto metrics. The method relies on a Randers approximation of the Finsler metric. We investigate whether the directional embeddings recovered from a non-Randers metric still capture the essential geometric properties of the data. We consider the Matsumoto metric (Matsumoto, 1992) induced by movement on a height map h over $[-3.2, 3.2] \times [-2.4, 2.4]$ (Matsumoto, 1989; Chansri et al., 2018),



(a) 3D Swiss Roll Randers manifold. **(b)** DiSBM graph. **(c)** Matsumoto manifold.

Figure 2: Directional embeddings. Recovered directional embeddings on the three datasets. The top row shows the input data, and the bottom row shows the corresponding recovered embeddings and drifts.

$\mathcal{F} = \alpha^2/(\alpha - \beta)$, where α is the Riemannian metric of the graph of h and $\beta = \mathrm{d}h$ (Appendix C.2). Figure 2c shows the height map and the drift recovered with Isomap embeddings. The drift is aligned with the gradient of h, the defining feature of the Matsumoto metric, which suggests that our method captures the main directional structure of non-Randers metrics.

Embedding directed graphs. Finally, we apply our method to general directed graphs, without an underlying manifold. We generate directed graphs from a Directed Stochastic Block Model (DiSBM, Appendix C.3). Figure 2b shows the leading eigenvectors of $\mathbf{L}^s$ for $N = 1000$ nodes and 15 communities: nodes cluster by community, and the drift follows the dominant, here backward, cyclic flow.

7 Conclusion and Perspectives

We showed that Finsler geometry plays for directed graphs the role that Riemannian geometry plays for undirected ones. The symmetric and antisymmetric parts of the graph operators recover, respectively, the Binet–Legendre metric and the centroid field of the unit ball, which yields a moment-Randers approximation of the underlying metric and a directed embedding of the data. Experiments on Randers and Matsumoto manifolds and on directed stochastic block models show that both the geometry and the directionality are recovered. We hope that this work will open up new avenues for applied Finsler geometry in graph analysis.

AI use statement

In this work, we used generative AI tools to assist in the writing of proofs, to design or provide feedback on research methodology or experiments, and to help with translation. We have not used generative AI tools to generate synthetic datasets, help develop theoretical models or conceptual frameworks, formulate mathe-

matical claims, provide critical ingredients for proving mathematical claims, propose or refine hypotheses, implement methods, clean and reformat datasets, support qualitative and thematic data analysis, or interpret results. Additionally, we used generative AI tools to search for information and identify relevant literature. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work proposes a novel method for representation learning. As such, there are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

REPRODUCIBILITY STATEMENT

The code used to produce the results in this paper is available publicly at ANONYMOUS and is licensed under the MIT License. All datasets used in this work are synthetic and can be regenerated with the released code. The theoretical results in this work are accompanied by complete proofs in the appendix, and all assumptions are clearly stated. A primer is provided in the appendix to introduce the reader to the relevant concepts of Finsler geometry, and to help them check the definitions, notations and proofs used in the main text.

REFERENCES

Georgios Arvanitidis, Soren Hauberg, Philipp Hennig, and Michael Schober. Fast and robust shortest paths on manifolds learned from data. In *International Conference on Artificial Intelligence and Statistics*, 2019.

Azam Asanjarani. A Finsler geometrical programming for the nonlinear complementarity problem of traffic equilibrium. *Preprint arXiv:2109.01256*, 2021.

David Bao, Shiing-Shen Chern, and Zhongmin Shen. *An introduction to Riemann-Finsler geometry*. Springer Science & Business Media, 2012.

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems*, 2001.

Volker Bergen, Ruslan A Soldatov, Peter V Kharchenko, and Fabian J Theis. RNA velocity—current challenges and future perspectives. *Molecular Systems Biology*, 17(8):MSB202110282, 2021.

Jeff Calder and Nicolas Garcia Trillos. Improved spectral convergence rates for graph Laplacians on ε-graphs and k-NN graphs. *Applied and Computational Harmonic Analysis*, 60:123–175, 2022.

P Chansri, Pattrawut Chansangiam, and Sorin V Sabau. The geometry on the slope of a mountain. *Preprint arXiv:1811.02123*, 2018.

Da Chen, Jean-Marie Mirebeau, Huazhong Shu, and Laurent D Cohen. A region-based Randers geodesic approach for image segmentation. *International Journal of Computer Vision*, 132(2):349–391, 2024.

Mo Chen, Qiong Yang, and Xiaoou Tang. Directed graph embedding. In *International Joint Conference on Artificial Intelligence*, 2007.

Fan Chung. Laplacians and the Cheeger inequality for directed graphs. *Annals of Combinatorics*, 9(1):1–19, 2005.

Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006.

Mihai Cucuringu, Huan Li, He Sun, and Luca Zanetti. Hermitian matrices for clustering directed graphs: insights and applications. In *International Conference on Artificial Intelligence and Statistics*, 2020.

Thomas Dagès, Simon Weber, Ya-Wei Eileen Lin, Ronen Talmon, Daniel Cremers, Michael Lindenbaum, Alfred M. Bruckstein, and Ron Kimmel. Finsler multi-dimensional scaling: Manifold learning for asymmetric dimensionality reduction and embedding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025.

Thomas Dagès, Simon Weber, Daniel Cremers, and Ron Kimmel. Harnessing data asymmetry: Manifold learning in the Finsler world. *Preprint arXiv:2603.11396*, 2026.

Gwendal Debaussart-Joniec, Harry Sevi, Matthieu Jonckheere, and Argyris Kalogeratos. Parametrized power-iteration clustering for directed graphs. In *International Conference on Machine Learning*, 2026.

David L Donoho and Carrie Grimes. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences*, 2003.

Michaël Fanuel, Carlos M. Alaíz, Ángela Fernández, and Johan A.K. Suykens. Magnetic eigenmaps for the visualization of directed networks. *Applied and Computational Harmonic Analysis*, 44(1):189–199, 2018.

Paul Finsler. *Über Kurven und Flächen in allgemeinen Räumen*. Philos. Fak., Georg-August-Univ., 1918.

Barak Gahtan, Jacob Shpund, and Alex M Bronstein. Differentiable Randers-Finsler eikonal solvers. *Computer Graphics Forum*, pp. e70489, 2026.

Mingzhen He, Fan He, Lei Shi, Xiaolin Huang, and Johan A.K. Suykens. Learning with asymmetric kernels: Least squares and feature interpretation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):10044–10054, 2023.

Matthias Hein, Jean-Yves Audibert, and Ulrike von Luxburg. Graph Laplacians and their convergence on random neighborhood graphs. *Journal of Machine Learning Research*, 8(6), 2007.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *Preprint arXiv:1609.02907*, 2016.

Stéphane S Lafon. *Diffusion maps and geometric harmonics*. Yale University, 2004.

John M. Lee. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer, 2 edition, 2012.

Federico Lopez, Beatrice Pozzetti, Steve Trettel, Michael Strube, and Anna Wienhard. Symmetric spaces for graph embeddings: A Finsler-Riemannian approach. In *International Conference on Machine Learning*, 2021.

Steen Markvorsen. A Finsler geodesic spray paradigm for wildfire spread modelling. *Nonlinear Analysis: Real World Applications*, 28:208–228, 2016.

Makoto Matsumoto. A slope of a mountain is a Finsler surface with respect to a time measure. *Journal of Mathematics of Kyoto University*, 29(1):17–25, 1989.

Makoto Matsumoto. Theory of Finsler spaces with (α, β)-metric. *Reports on Mathematical Physics*, 31(1): 43–83, 1992.

Vladimir S Matveev and Marc Troyanov. The Binet–Legendre metric in Finsler geometry. *Geometry & Topology*, 16(4):2135–2170, 2012.

John Melonakos, Eric Pichon, Sigurd Angenent, and Allen Tannenbaum. Finsler active contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3), 2008.

Shin-ichi Ohta. *Comparison Finsler Geometry*. Springer Monographs in Mathematics. Springer, 2021.

Mingdong Ou, Peng Cui, Jian Pei, Ziwei Zhang, and Wenwu Zhu. Asymmetric transitivity preserving graph embedding. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

J. Wilson Peoples and John Harlim. Spectral convergence of symmetrized graph Laplacian on manifolds with boundary. *Foundations of Data Science*, 8(0):119–167, 2025.

Dominique Perrault-Joncas and Marina Meila. Directed graph embedding: an algorithm based on continuous limits of Laplacian-type operators. *Advances in Neural Information Processing Systems*, 24, 2011.

Christian Pfeifer. Finsler spacetime geometry in physics. *International Journal of Geometric Methods in Modern Physics*, 16(supp02):1941004, 2019.

Alison Pouplin, David Eklund, Carl Henrik Ek, and Søren Hauberg. Identifying latent distances with Finslerian geometry. *Preprint arXiv:2212.10010*, 2023.

Gunnar Randers. On an asymmetrical metric in the four-space of general relativity. *Physical Review*, 59(2): 195, 1941.

Frederik Möbius Rygaard and Søren Hauberg. GEORCE: A fast new control algorithm for computing geodesics. *Preprint arXiv:2505.05961*, 2025.

Venu Satuluri and Srinivasan Parthasarathy. Symmetrizations for clustering directed graphs. In *International Conference on Extending Database Technology*, 2011.

Harry Sevi, Gwendal Debaussart-Joniec, Malik Hacini, Matthieu Jonckheere, and Argyris Kalogeratos. Generalized Dirichlet energy and graph Laplacians for clustering directed and undirected graphs. *Transactions on Machine Learning Research*, 2026.

Tony Shaska. Geometric learning and Finsler dissimilarity in weighted projective spaces. *Preprint arXiv:2507.00001*, 2025.

Yi-Bing Shen and Zhongmin Shen. *Introduction to modern Finsler geometry*. World Scientific Publishing Company, 2016.

Zhongmin Shen. *Lectures on Finsler geometry*. World Scientific, 2001.

Amit Singer. From graph to manifold Laplacian: The convergence rate. *Applied and Computational Harmonic Analysis*, 21(1):128–134, 2006. Special Issue: Diffusion Maps and Wavelets.

Stefan Sommer, Tom Fletcher, and Xavier Pennec. 1 - introduction to differential and Riemannian geometry. In Xavier Pennec, Stefan Sommer, and Tom Fletcher (eds.), *Riemannian Geometric Statistics in Medical Image Analysis*, pp. 3–37. Academic Press, 2020.

Qinghua Tao, Francesco Tonin, Alex Lambert, Yingyi Chen, Panagiotis Patrinos, and Johan A.K. Suykens. Learning in feature spaces via coupled covariances: asymmetric kernel SVD and Nyström method. In *International Conference on Machine Learning*, 2024.

Joshua B Tenenbaum, Vin de Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

Roman Vershynin. *High-dimensional probability*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2019.

Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

Simon Weber, Thomas Dagès, Maolin Gao, and Daniel Cremers. Finsler-Laplace-Beltrami operators with application to shape analysis. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

Wei Wu, Jun Xu, Hang Li, and Satoshi Oyama. Asymmetric kernel learning. Technical report, Microsoft Research, 2010.

Won Dae Yeon. On the History of the Birth of Finsler Geometry at Göttingen. *Journal for History of Mathematics*, 2015.

Amber Yuan, Jeff Calder, and Braxton Osting. A continuum limit for the PageRank algorithm. *European Journal of Applied Mathematics*, 33(3):472–504, 2022.

Ernst Zermelo. Über das Navigationsproblem bei ruhender oder veränderlicher Windverteilung. *Journal of Applied Mathematics and Mechanics*, 11(2):114–124, 1931.

Xitong Zhang, Yixuan He, Nathan Brugnone, Michael Perlmutter, and Matthew Hirn. MagNet: A neural network for directed graphs. In *Advances in Neural Information Processing Systems*, 2021.

Dengyong Zhou, Olivier Bousquet, Thomas Lal, Jason Weston, and Bernhard Schölkopf. Learning with local and global consistency. *Advances in Neural Information Processing Systems*, 16, 2003.

Aaron Zweig, Mingxuan Zhang, David A Knowles, and Elham Azizi. Learning lineage-guided geodesics with finsler geometry. *Preprint arXiv:2603.16708*, 2026.

APPENDIX CONTENTS

A Table of notations **14**

B A Primer of Riemannian and Finsler-Randers Geometry **14**

C Additional experimental results **19**

C.1 Swiss Roll dataset 20

C.2 Matsumoto Metric approximation 20

C.3 Directed Stochastic Block Model 21

D Proofs **23**

D.1 Forward and backward transport operators 23

D.2 Proof of Theorem 1 24

D.3 Proof of Theorem 2 30

D.4 Links between the moments of the kernel and the unit ball 40

D.5 Proof of Corollary 3 42

D.6 Proof of Theorem 4 43

D.7 Computation of the moments in the Randers setting 50

D.8 The moment-Randers approximation 52

D.9 Computation of the kernel moments 58

A TABLE OF NOTATIONS

We provide in Tables 2 and 3 a summary of the main notations used throughout the paper.

B A PRIMER OF RIEMANNIAN AND FINSLER-RANDERS GEOMETRY

In this section we provide a brief introduction to Riemannian geometry and its extension to Finsler geometry. For a more complete and formal introduction to those subjects, we refer the reader to Sommer et al. (2020) for Riemannian geometry, and to Ohta (2021) for a comprehensive treatment of Finsler geometry and Randers metrics. Throughout this paper, we use the Einstein summation convention, where an index repeated once as a superscript and once as a subscript is summed over, e.g. $\partial_i f\, v^i = \sum_{i=1}^m \partial_i f\, v^i$.

Riemannian geometry. Riemannian manifolds can be described in multiple ways. Generally, they are defined as smooth manifolds (i.e. topological spaces that are locally homeomorphic to Euclidean spaces, with smoothly compatible charts) endowed with a smoothly varying inner product on the tangent space at each point, called the Riemannian metric. This inner product allows one to define distances and angles on the manifold, see Figure 3b for an illustration. One may think of a Riemannian manifold as a curved space in which the local cost of crossing a region is governed by the metric. More formally, given a smooth manifold

Table 2: Notation table for the continuous setting. This table summarizes the notations attached to the manifold, the Finsler metric and the integral operators, with references to the equations where they are defined.

Notation	Description
$\mathscr{M}$	Compact m-dimensional manifold embedded in $\mathbb{R}^D$
$\|w\|_M$, $\langle w, w'\rangle_M$	$\|w\|_M^2 = w^\mathsf{T} M w$ and $\langle w, w'\rangle_M = w^\mathsf{T} M w'$, for $M \succ 0$
$\mathscr{F}$, $\mathrm{dst}_\mathscr{F}(\cdot,\cdot)$	Finsler metric and its associated distance
$\overleftarrow{\mathscr{F}}$	Reversed Finsler metric
$\alpha, \beta, \mathbf{A}, b$	Randers metric parameters
$\lambda_\mathscr{F}$, $i_\mathscr{M}$	Reversibility constant $\sup_{x,v} \mathscr{F}(x,v)/\mathscr{F}(x,-v)$ and injectivity radius of $\mathscr{M}$ (A2)
$\mathfrak{m}_{\mathrm{BH}}$, σ_{BH}	Busemann–Hausdorff measure on $\mathscr{M}$ and its density in a chart (Equation (4))
λ_x	Lebesgue measure on $T_x\mathscr{M}$, in the coordinate basis
dist, Vol	Riemannian distance and volume induced on $\mathscr{M}$ by the ambient Euclidean norm
$\mathscr{B}_x$, $\mathscr{B}_x(r)$	Finsler unit ball $\{\mathscr{F}(x,\cdot) \le 1\}$ and open ball $\{\mathscr{F}(x,\cdot) < r\}$ in the tangent space $T_x\mathscr{M}$
$B^+_\mathscr{F}(x,r)$	Forward metric ball $\{y \in \mathscr{M} : \mathrm{dst}_\mathscr{F}(x,y) < r\}$ on the manifold $\mathscr{M}$
$\mathbb{B}^m$	Euclidean unit ball of $\mathbb{R}^m$
$\mathscr{I}_x$	Indicatrix $\{v \in T_x\mathscr{M} : \mathscr{F}(x,v) = 1\}$, boundary of $\mathscr{B}_x$
$\mathbf{c}(x)$, $\mathbf{S}(x)$	Centroid and raw second moment of the unit ball $\mathscr{B}_x$ (Equations (22) and (23))
g_{BL}	Binet–Legendre metric of $\mathscr{F}$, with $g^{ij}_{\mathrm{BL}} = (m+2)\mathbf{S}^{ij}$ (Equation (10))
$\mathbb{P}$, ρ	Sampling distribution on $\mathscr{M}$ and its density $\rho = \mathrm{d}\mathbb{P}/\mathrm{d}\mathfrak{m}_{\mathrm{BH}}$
K	Radial kernel profile on $\mathbb{R}_+$
W, $W^{(\theta)}$	Directed kernel on $\mathscr{M}\times\mathscr{M}$ and its θ-normalization
d, d', q_ε	Out- and in-degrees of W, and their half-sum $q_\varepsilon = (d+d')/2$
$\mathscr{G}_\mathscr{F}$, $\mathscr{G}'_\mathscr{F}$	Integral operators built from W; $\mathscr{G}'_\mathscr{F} = \mathscr{G}_{\overleftarrow{\mathscr{F}}} = \mathscr{G}^*_\mathscr{F}$ (Proposition 19)
$\mathscr{G}^{(\theta)}_\mathscr{F}$, $\mathscr{G}^{(\theta,s)}_\mathscr{F}$, $\mathscr{G}^{(\theta,a)}_\mathscr{F}$	θ-normalized integral operator, and its symmetric and antisymmetric parts
$\mathscr{P}^{(\theta,s)}$, $\mathscr{P}^{(\theta,a)}$	Continuous transport operators built from $\mathscr{G}^{(\theta,s)}_\mathscr{F}$ and $\mathscr{G}^{(\theta,a)}_\mathscr{F}$ (Section 3)
$\mathscr{L}^s$, $\mathscr{L}^a$	Their limits as $\varepsilon \to 0$: a weighted Laplacian and a drift along $\mathbf{c}$ (Theorem 2)
$\exp^\mathscr{F}$	Finsler exponential map of $\mathscr{F}$
G^k, N^i_j	Geodesic spray coefficients and nonlinear connection of $\mathscr{F}$ (Definition 27)
$\tilde{\Gamma}^k_{ij}$, $\mathrm{Ric}^\mathscr{F}$	Formal Christoffel symbols and Ricci curvature of $\mathscr{F}$
τ, S	Distortion and S-curvature of $\mathscr{F}$ (Definition 18)

$\mathscr{M}$, for any point $x \in \mathscr{M}$ we can define the *tangent space* $T_x\mathscr{M}$ as the vector space of all tangent vectors at x. A Riemannian manifold is then a pair $(\mathscr{M}, \mathbf{A})$, where $\mathbf{A}$ is a smoothly varying inner product on the tangent space at each point,

$$\langle\cdot,\cdot\rangle_{\mathbf{A}(x)} : T_x\mathscr{M} \times T_x\mathscr{M} \to \mathbb{R}, \qquad (v,w) \mapsto v^\mathsf{T}\mathbf{A}(x)w.$$

In coordinates, components of tangent vectors carry upper indices, $v = (v^i)$, while components of covectors, such as the partial derivatives $\partial_i f$ of a function, carry lower indices. We write $\mathbf{A}_{ij}$ for the entries of $\mathbf{A}$, so that $\langle v, w\rangle_\mathbf{A} = \mathbf{A}_{ij}v^i w^j$, and $\mathbf{A}^{ij} = (\mathbf{A}^{-1})_{ij}$ for those of its inverse, so that $\mathbf{A}^{ik}\mathbf{A}_{kj} = \delta^i_j$. The same convention applies to any metric. This inner product induces a natural local norm, $\|v\|_{\mathbf{A}(x)} = \sqrt{v^\mathsf{T}\mathbf{A}(x)v}$ for $v \in T_x\mathscr{M}$, which in turn defines the length and energy of a curve $\gamma : [0,1] \to \mathscr{M}$:

$$L(\gamma) = \int_0^1 \|\dot\gamma(t)\|_{\mathbf{A}(\gamma(t))}\mathrm{d}t \quad \text{and} \quad E(\gamma) = \int_0^1 \|\dot\gamma(t)\|^2_{\mathbf{A}(\gamma(t))}\mathrm{d}t.$$

This lets us define the geodesic distance between two points as the length of the shortest path between them, which is also (up to reparametrization) the minimizer of the energy between these two points:

$$\forall (a,b) \in \mathscr{M}^2, \mathrm{dst}(a,b) = \min_{\substack{\gamma:[0,1]\to\mathscr{M} \\ \gamma(0)=a, \gamma(1)=b}} L(\gamma) \tag{9}$$

Table 3: Notation table: discrete setting, hyper-parameters and constants. This table summarizes the notations attached to the sampled data, the matrices built from it, the hyper-parameters, and the moments and constants of the kernel.

Notation	Description
$X_1, \dots, X_N$	A sequence of N i.i.d. samples from $\mathbb{P}$ on $\mathscr{M}$
$\mathbf{W}_N$, $\mathbf{Q}_N$	Empirical kernel matrix and half-sum of degrees $\mathbf{Q}_N = (\mathbf{D}_N + \mathbf{D}'_N)/2$ (Section 4)
$\mathbf{D}_N$, $\mathbf{D}'_N$	Diagonal matrices of the out- and in-degrees of $\mathbf{W}_N$
$\mathbf{W}_N^{(\theta)}$, $\mathbf{W}_N^{(\theta,\bullet)}$	θ-normalized matrix and its symmetric ($\bullet = s$) and antisymmetric ($\bullet = a$) parts
$\mathbf{D}_N^{(\theta,\bullet)}$	Diagonal matrix of the row sums of $\mathbf{W}_N^{(\theta,\bullet)}$, discretizing $\mathscr{G}_{\mathscr{F}}^{(\theta,\bullet)}[\rho]$
$\mathbf{P}^{(\theta,\bullet)}$	Discrete transport operators, discretizing $\mathscr{P}^{(\theta,\bullet)}$ (Equation (8))
$\mathbf{L}^s$, $\mathbf{L}^a$	Rescaled discrete operators $\mathbf{P}^{(\theta,s)}/\varepsilon^2$ and $\mathbf{P}^{(\theta,a)}/\varepsilon$ (Equation (8))
$\hat{q}_N$	Empirical counterpart of the normalization q_ε
Ψ, Y_i	Spectral embedding $\Psi = (\psi_1, \dots, \psi_\ell)$ of $\mathbf{L}^s$, and embedded points $Y_i = \Psi(X_i)$
$\mathbf{H}$, $\mathscr{F}_R$	Moment-Randers matrix of $\mathscr{F}$ and its moment-Randers metric (Definition 5)
Γ, V	Estimated Binet–Legendre metric and drift, computed from $\mathscr{L}^s, \mathscr{L}^a$ (Proposition 7)
$\hat{\mathbf{c}}$, $\hat{\mathbf{H}}$, $\hat{\mathscr{F}}_R$	Empirical centroid, matrix, and moment-Randers metric estimated from the embedding
ε	Kernel bandwidth parameter
θ	Normalization exponent of the kernel, $\theta \in [0,1]$
D	Ambient dimension of the manifold $\mathscr{M}$
m	Intrinsic dimension of the manifold $\mathscr{M}$
ℓ	Dimension of the embedding, i.e. number of eigenvectors of $\mathbf{L}^s$ retained
m_n, Λ^k, s_n, Υ	Moments of the kernel K (Definition 21)
$\tilde{m}_n$	Normalized moments of the kernel, $\tilde{m}_n = m_n/m_0$ (Proposition 35)
μ_n	Raw moments of the radial profile, $\mu_n = \int_0^{+\infty} K(r) r^{m+n-1} \mathrm{d}r$
C_K, ν_K	Sub-exponential constants of the kernel: $K(r) \le C_K e^{-\nu_K r}$ (A1)
c_1	Constant $(m+1)\mu_1/(m\mu_0)$ appearing in $\mathscr{L}^a$ (Theorem 2)
c_2	Constant $\mu_2/(2m\mu_0)$ appearing in $\mathscr{L}^s$ and Γ (Theorem 2 and Proposition 7)

Finsler geometry. Finsler manifolds are a natural extension of Riemannian manifolds that relax two of their defining constraints: the local cost need not be a quadratic form, and it need not be symmetric. This lets Finsler geometry model *local asymmetry* (the cost of crossing a region can depend on the direction of travel) while keeping a formulation close to that of Riemannian geometry, as illustrated in Figure 3c.

Definition 10 (Finsler metric)**.** *A Finsler metric on a smooth manifold $\mathscr{M}$ is a function $\mathscr{F} : T\mathscr{M} \to [0,\infty)$ such that for each $x \in \mathscr{M}$ and $v \in T_x\mathscr{M}$, the following conditions hold:*

- ***Positive homogeneity**: $\mathscr{F}(x, tv) = t\mathscr{F}(x,v)$ for all $t > 0$ and $v \in T_x\mathscr{M}$.*
- ***Strong convexity**: The Hessian of $\mathscr{F}^2$ with respect to the velocity variable is positive definite for all nonzero $v \in T_x\mathscr{M}$.*
- ***Triangle inequality**: $\mathscr{F}(x, v+w) \le \mathscr{F}(x,v) + \mathscr{F}(x,w)$ for all $v, w \in T_x\mathscr{M}$.*

Since homogeneity is only required for positive scalars t, these metrics induce local distances that depend on direction: we generally do not have $\mathscr{F}(x,v) = \mathscr{F}(x,-v)$.

Remark 11. *In some textbooks, one may also encounter the notation $\mathscr{F}(v)$, dropping the explicit dependence on the base point x. Throughout this work, we will keep the explicit dependence on x on the Finsler-based quantities.*

Definition 12 (Fundamental tensor)**.** *The fundamental tensor g of a Finsler metric $\mathscr{F}$ is given by*

$$g_{ij}(x,v) = \frac{1}{2} \frac{\partial^2 \mathscr{F}^2}{\partial v^i \partial v^j}(x,v),$$

where $(x, v) \in T\mathcal{M}$ and v^i are the components of v in a local coordinate system. The fundamental tensor is a smoothly varying, positive definite bilinear form on the tangent space at each point, and it generalizes the notion of a Riemannian metric to Finsler geometry.

A general Finsler tangent indicatrix $\{v \in T_x\mathcal{M} : \mathcal{F}(x, v) = 1\}$ need not be an ellipsoid at all, as Finsler metrics are only required to be strongly convex, not quadratic. Nonetheless, the fundamental tensor always yields the best local Riemannian (quadratic) approximation of $\mathcal{F}$ at (x, v).

Similarly to Riemannian geometry, shortest paths under a Finsler metric are also minimizers of the energy of the curve. Using Euler's homogeneous function theorem, one can show that the Finsler local energy satisfies

$$\mathcal{F}^2(x, v) = v^\mathsf{T} g(x, v) v = \|v\|^2_{g(x,v)}.$$

Consequently, we can define the length and energy of curves, and thus geodesic distances, on Finsler manifolds by replacing $\mathbf{A}(\gamma_t)$ with $g(\gamma_t, \dot{\gamma}_t)$ in the definitions above. This lets us use the same class of algorithms as in the Riemannian case to solve the shortest-path problem, at marginal additional cost (Rygaard & Hauberg, 2025; Arvanitidis et al., 2019).

Balls. Several kinds of balls appear in this work. For $x \in \mathcal{M}$ and $r > 0$,

$$\mathcal{B}_x = \{v \in T_x\mathcal{M} : \mathcal{F}(x, v) \leq 1\},\ \mathcal{B}_x(r) = \{v \in T_x\mathcal{M} : \mathcal{F}(x, v) < r\},\ B^+_{\mathcal{F}}(x, r) = \{y \in \mathcal{M} : \mathrm{dst}_{\mathcal{F}}(x, y) < r\}.$$

The first two live in the tangent space $T_x\mathcal{M}$: $\mathcal{B}_x$ is the (closed) unit tangent ball, whose boundary is the indicatrix $\mathcal{I}_x = \{v \in T_x\mathcal{M} : \mathcal{F}(x, v) = 1\}$, and $\mathcal{B}_x(r)$ is the open tangent ball of radius r, which coincides with $r\mathcal{B}_x$ up to its boundary. The third one is the forward metric ball and lives on the manifold $\mathcal{M}$. Since $\mathrm{dst}_{\mathcal{F}}$ is not symmetric, it differs in general from the backward ball. Below the injectivity radius, $\exp^{\mathcal{F}}_x$ maps $\mathcal{B}_x(r)$ onto $B^+_{\mathcal{F}}(x, r)$. Finally, $\mathbb{B}^m$ denotes the Euclidean unit ball of $\mathbb{R}^m$, and $\mathrm{Vol}_{\mathrm{Euc}}(\mathbb{B}^m)$ its volume. Unlike $\lambda_x(\mathcal{B}_x)$, it does not depend on x.

Randers metrics. A Randers metric (Randers, 1941) is a special type of Finsler metric that can be expressed as the sum of a Riemannian metric and a one-form.

Definition 13 (Randers metric)**.** *A Randers metric on a smooth manifold $\mathcal{M}$ is a Finsler metric $\mathcal{F}$ that can be expressed as*

$$\mathcal{F}(x, v) = \alpha(x, v) + \beta(x, v),$$

where α is a Riemannian metric and β is a 1-form, such that $\alpha(x, v)^2 = v^\mathsf{T}\mathbf{A}(x)v$, with $\mathbf{A} : \mathcal{M} \to \mathcal{S}^D_{++}$ a smooth map from the manifold to the space of positive definite matrices, and $\beta(x, v) = b(x)^\mathsf{T} v$, with $b : \mathcal{M} \to \mathbb{R}^D$ a smooth vector field on the manifold verifying $\|b(x)\|_{\mathbf{A}^{-1}} < 1$.

In coordinates, $\beta(x, v) = b_i(x)v^i$, so that b carries a lower index as a covector. Its index is raised with $\mathbf{A}$, $b^i = \mathbf{A}^{ij}b_j$, and $\|b\|^2_{\mathbf{A}^{-1}} = \mathbf{A}^{ij}b_i b_j = b_i b^i$.

Randers metrics are a particular case of Finsler metrics that are generally easier to work with due to their specific structure: for such metrics we have an explicit form of g and g^{-1}, enabling efficient implementation. Moreover, the Finsler indicatrix of a Randers metric is always an ellipsoid, simply not centered at the origin, obtained by translating the α-indicatrix by a vector determined by $b(x)$.

Example 14. *Perhaps the simplest example of a Randers metric is the Euclidean space $\mathbb{R}^2$ with the standard Euclidean metric $\alpha(x, v) = \|v\|_2$ and a constant vector field $b(x) = (b_1, b_2)$ with $\|b\|_2 < 1$. In this case, the Randers metric is given by*

$$\mathcal{F}(x, v) = \|v\|_2 + b^\mathsf{T} v = \sqrt{v_1^2 + v_2^2} + b_1 v_1 + b_2 v_2.$$

The indicatrix of this Randers metric is an ellipse that is shifted in the direction of the vector b. One may verify that, in this context, geodesics are straight lines, but the distance between two points depends on the direction of travel.

Berwald metrics. Among Finsler metrics, an important intermediate class, strictly more general than Riemannian metrics but with enough rigidity to retain some of their linear structure, is given by Berwald metrics.

Definition 15 (Berwald metric)**.** *A Finsler metric $\mathscr{F}$ is a* Berwald metric *if its formal Christoffel symbols $\tilde{\Gamma}^k_{ij}(x,v)$ (Ohta, 2021) do not depend on the direction v, i.e. $\tilde{\Gamma}^k_{ij}(x,v) = \Gamma^k_{ij}(x)$ for some coefficients Γ^k_{ij} on $\mathscr{M}$.*

Every Riemannian metric is trivially Berwald, since its fundamental tensor $g_{ij}(x,v) = \mathbf{A}_{ij}(x)$ does not depend on v to begin with. In the Randers case, $\mathscr{F} = \alpha + \beta$ is Berwald if and only if β is parallel with respect to the Levi–Civita connection of α (Ohta, 2021), a condition on b that constant vector fields such as the one in the previous example trivially satisfy, but that a direction field varying freely across $\mathscr{M}$, as we consider in our experiments, generally will not.

Binet–Legendre metric. The fundamental tensor $g(x,v)$ of a Finsler metric depends on the direction v. A direction-free Riemannian metric is obtained by averaging over the unit tangent ball $\mathscr{B}_x$ (Matveev & Troyanov, 2012).

Definition 16 (Binet–Legendre metric)**.** *The Binet–Legendre metric g_{BL} of $\mathscr{F}$ is the Riemannian metric whose inverse is*

$$g^{ij}_{\mathrm{BL}}(x) = \frac{m+2}{\lambda_x(\mathscr{B}_x)} \int_{\mathscr{B}_x} v^i v^j \, \mathrm{d}\lambda_x(v) = (m+2)\mathbf{S}^{ij}(x). \tag{10}$$

The factor $m+2$ ensures that $g_{\mathrm{BL}} = \mathbf{A}$ when $\mathscr{F}(x,v) = \sqrt{v^{\mathsf{T}}\mathbf{A}(x)v}$ is Riemannian (Appendix D.5). Here g^{ij}_{BL} denotes the entries of the inverse of g_{BL}, following the convention above. By contrast, the second moment $\mathbf{S}^{ij}$ of $\mathscr{B}_x$, like the kernel moment m_2^{ij} (Definition 21), is defined directly with upper indices, and is not the inverse of a matrix with lower indices. The metric g_{BL} is what the symmetric limit $\mathscr{L}^s$ sees (Theorem 2), while the antisymmetric limit $\mathscr{L}^a$ sees the centroid $\mathbf{c}(x)$ of $\mathscr{B}_x$.

Busemann–Hausdorff measure. Finsler manifolds carry no canonical volume; the two usual choices are the Busemann–Hausdorff and Holmes–Thompson measures, which coincide in the Riemannian case (Shen, 2001). We use the former.

Definition 17 (Busemann–Hausdorff measure)**.** *The Busemann–Hausdorff measure $\mathfrak{m}_{\mathrm{BH}}$ is the measure on $\mathscr{M}$ which reads, in a chart, $\mathrm{d}\mathfrak{m}_{\mathrm{BH}}(x) = \sigma_{\mathrm{BH}}(x)\mathrm{d}x$, with density*

$$\sigma_{\mathrm{BH}}(x) = \frac{\mathrm{Vol}_{\mathrm{Euc}}(\mathbb{B}^m)}{\lambda_x(\mathscr{B}_x)}, \tag{11}$$

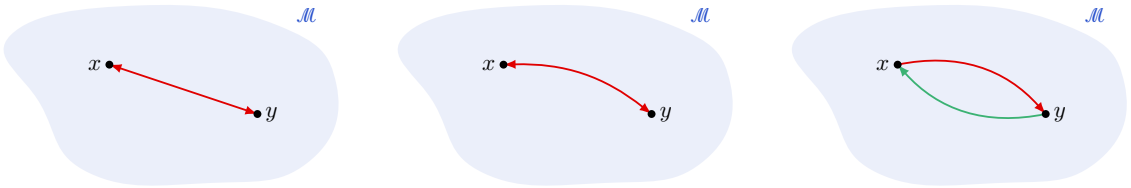
where $\lambda_x(\mathscr{B}_x)$ is the Lebesgue volume of $\mathscr{B}_x$ in the coordinate basis $(\partial_i|_x)$. Under a change of chart, σ_{BH} and $\mathrm{d}x$ are multiplied by inverse factors, so that $\mathfrak{m}_{\mathrm{BH}}$ does not depend on the chart.

Under $\mathfrak{m}_{\mathrm{BH}}$, every unit tangent ball has the volume of the Euclidean unit ball. This is what makes the zeroth kernel moment m_0 constant (Proposition 35), so that the degree normalization behaves as in the Riemannian case.

Distortion and S-curvature. The Jacobian of $\exp^{\mathscr{F}}_x$ with respect to $\mathfrak{m}_{\mathrm{BH}}$ (Lemma 25) involves the mismatch between $\mathfrak{m}_{\mathrm{BH}}$ and the direction-dependent volume $\sqrt{\det g(x,v)}\mathrm{d}x$, which is measured by the following two quantities.

Definition 18 (Distortion and S-curvature)**.** *For $z \in \mathscr{M}$ and $u \in T_z\mathscr{M} \setminus \{0\}$, the distortion of $\mathscr{F}$ with respect to $\mathfrak{m}_{\mathrm{BH}}$ is*

$$\tau(z,u) = \log\left(\frac{\sqrt{\det g_{ij}(z,u)}}{\sigma_{\mathrm{BH}}(z)}\right),$$



(a) Euclidean distance through ambient space does not take into account the geometry of the manifold.

(b) Riemannian distance follows a symmetric geodesic on the manifold. It captures the intrinsic geometry of the manifold without any directional information.

(c) Finsler distance is direction-dependent, with different paths for forward and reverse directions. It captures both the intrinsic geometry of the manifold and the directional information.

Figure 3: Comparison of distance metrics on a manifold. Each panel shows the same manifold $\mathcal{M}$ with two points x and y, but different notions of distance between them. (a) Euclidean distance is the straight line in ambient space. (b) Riemannian distance follows a symmetric geodesic on the manifold. (c) Finsler distance is direction-dependent, with different paths for forward (orange) and reverse (purple) directions.

and the S-curvature is its derivative along the geodesic γ with $\gamma(0) = z$ and $\dot\gamma(0) = u$, i.e. $S(z,u) = \frac{\mathrm{d}}{\mathrm{d}t}\tau(\gamma(t),\dot\gamma(t))\big|_{t=0}$.

The S-curvature vanishes for Berwald metrics, and in particular for Riemannian ones (Shen & Shen, 2016, Proposition 4.3), but not for general Randers metrics. It produces the terms s_0 and s_1 of Definition 21, which are removed by the centering of $\mathcal{P}^{(\theta,\bullet)}$.

Consequences of assumption (A2). The proofs use three uniform properties of $(\mathcal{M},\mathcal{F})$, which follow from the compactness of $\mathcal{M}$ and the smoothness and positivity of $\mathcal{F}$ on $T\mathcal{M}\setminus\{0\}$.

(i) *Bounded asymmetry.* The reversibility constant $\lambda_{\mathcal{F}} = \sup_{x,v\neq 0}\mathcal{F}(x,v)/\mathcal{F}(x,-v)$ is finite, as the supremum of a continuous function on the compact unit sphere bundle. Hence $\lambda_{\mathcal{F}}^{-1}\mathrm{dst}_{\mathcal{F}}(y,x) \leq \mathrm{dst}_{\mathcal{F}}(x,y) \leq \lambda_{\mathcal{F}}\mathrm{dst}_{\mathcal{F}}(y,x)$.

(ii) *Completeness.* $\mathcal{M}$ is forward and backward complete, so that any two points are joined by a minimizing geodesic and $\exp^{\mathcal{F}}$ is defined on all of $T\mathcal{M}$ (Bao et al., 2012).

(iii) *Injectivity radius.* The injectivity radius is bounded below by some $i_{\mathcal{M}} > 0$, so that $\exp^{\mathcal{F}}_x$ is a diffeomorphism from $\mathcal{B}_x(i_{\mathcal{M}})$ onto its image for every x.

C Additional Experimental Results

We provide in this section additional details on the experiments presented in Section 6. We show how we construct the Randers metric and the drift for the Swiss Roll dataset, the Matsumoto metric approximation, and the Directed Stochastic Block Model (DiSBM).

On the illustrations. The presented method recovers the drift field as a pointwise quantity at each embedded sample. In the figures, we display the drift field as a vector field on a grid of points in the embedding space, obtained by interpolating the drift vectors at the embedded samples with a neural network.

C.1 SWISS ROLL DATASET

We consider the Swiss Roll dataset $(X_i)_i$ generated using *sklearn* with $N = 2000$ points and a noise of 0.1. We consider this dataset as sampled from a Randers manifold. The Riemannian metric is given by:

$$\mathbf{A}_{ii}^{-1}(z) = \sum_k \exp\left(-\frac{\|z - z_k\|^2}{h}\right) + \xi, \quad \text{and} \quad \mathbf{A}_{ij}^{-1} = 0 \text{ for } i \neq j,$$

where $\xi = 10^{-3}$ and h is chosen as the average distance to the 5th nearest neighbor.

We can then construct the drift as tangential to the dataset. Let $\varphi_i = \operatorname{atan2}(X_{i,z}, X_{i,x})$ be the angle of the point X_i in the xz-plane. The initial tangent vector is given by $b(X_i) = (-\sin(\varphi_i), 0, \cos(\varphi_i))$, which is then normalized so that $\|b(X_i)\|_{\mathbf{A}^{-1}} = 0.9$.

The directed graph is computed as a KNN graph with $k = 10$ neighbors, and the edge weights, i.e. geodesic distances between nodes, are computed using the *GEORCE* algorithm (Rygaard & Hauberg, 2025). The ε in the Gaussian kernel is chosen as the standard deviation of the geodesic distances between the nodes. The points are then embedded in $\mathbb{R}^2$ using the Isomap algorithm. The embedding is then used to estimate the Randers metric and the drift following the method described in the paper.

Quantitative evaluation. For Table 1, we keep the same metric and vary the drift norm $\beta = \|b\|_{\mathbf{A}^{-1}}$ and the number of points, with 3 seeds per configuration. Here, we use radius graphs rather than k-NN graphs. With a strong drift, the kernel extends further along the drift than across it, so that a k-NN graph would truncate it unevenly and bias its moments, and hence the limit of Theorem 2. The bandwidth ε is the median Finsler distance from a sample to its 10-th nearest neighbor. The graph is then truncated to the ball $\mathrm{dst}_{\mathscr{F}}(X_i, X_j) < 3\varepsilon$, outside of which the kernel is negligible. As above, the edge weights are $K(\mathrm{dst}_{\mathscr{F}}(X_i, X_j)/\varepsilon)$ with the Gaussian kernel, where $\mathrm{dst}_{\mathscr{F}}$ is computed with *GEORCE*. In other words, the graph keeps every pair on which the kernel is non-negligible, and the geometry is carried by the weights.

The ground truth $\mathbf{c}$ and g_{BL} is given by Proposition 43. The tangent plane of the roll is estimated by a local PCA on the 15 nearest neighbors. The Jacobian J_i of Isomap at X_i is estimated by least squares on the same neighbors, $Y_j - Y_i \approx J_i(X_j - X_i)$ in these tangent coordinates. It is only needed for the direction of the drift, since $\|\hat{\mathbf{c}}\|_{\hat{g}_{BL}}$ does not depend on the embedding. Table 1 reports the cosine between V_i and its limit $c_1 J_i \mathbf{c}(X_i)$: the direction of the drift is recovered at every strength and sample size.

C.2 MATSUMOTO METRIC APPROXIMATION

We focus on the case of a graph sampled from a Matsumoto metric (Matsumoto, 1992), more precisely on the metric induced by movement on a height map (Matsumoto, 1989; Chansri et al., 2018). We define on $[-3.2, 3.2] \times [-2.4, 2.4]$ the height function h:

$$h(x,y) = 0.4\sin(x)\cos(y) + 0.15\exp\left(-((x+1)^2 + (y-0.5)^2)\right) - 0.15\exp\left(-((x-1)^2 + (y+0.5)^2)\right).$$

The Matsumoto metric is then given by:

$$\begin{aligned}
\alpha((x,y),v)^2 &= (1+\partial_x h(x,y)^2)v_x^2 + 2\partial_x h(x,y)\partial_y h(x,y)v_x v_y + (1+\partial_y h(x,y)^2)v_y^2, \\
\beta((x,y),v) &= \partial_x h(x,y)v_x + \partial_y h(x,y)v_y, \\
\mathscr{F}((x,y),v) &= \frac{\alpha((x,y),v)^2}{\alpha((x,y),v) - \beta((x,y),v)}.
\end{aligned}$$

We then sample $N = 3000$ points uniformly on $[-3.2, 3.2] \times [-2.4, 2.4]$ that yield the samples $(x_i, y_i, h(x_i, y_i))_i$. Similarly to the Swiss Roll dataset, we construct a directed graph from the sampled points using a KNN graph with $k = 10$ neighbors, and where the edge weights are computed using the *GEORCE* algorithm (Rygaard & Hauberg, 2025). The ε in the Gaussian kernel is chosen as the standard deviation of the geodesic distances between the nodes. The samples are then embedded in $\mathbb{R}^2$ using the Isomap algorithm. The embedding is then used to estimate the Randers metric and the drift following the method described in the paper.

C.3 Directed Stochastic Block Model

Formal definition. We start by giving a formal definition of the DiSBM. Let N be the number of nodes in the graph, B the number of communities, and $n \in \mathbb{N}^B$ the number of nodes in each community. We denote by $\Pi \in [0,1]^{B \times B}$ the block matrix that defines the connection probabilities between communities, such that $\sum_j \Pi_{i,j} = 1$ for every i. We consider in particular the case of directed graphs that are generated from a circular block model, where the communities are arranged in a circle and the connection probabilities are defined as follows:

$$\Pi_{i,j} = \begin{cases} p & \text{if } j = i+1 \mod B, \\ q & \text{if } j = i-1 \mod B, \\ r & \text{if } j = i, \\ 0 & \text{otherwise,} \end{cases}$$

where $p, q, r \in [0,1]$ and $p+q+r = 1$. The probabilities p, q, r are the probabilities of connecting to the next, previous, and same community, respectively. The adjacency matrix $\mathbf{Q} \in \{0,1\}^{N \times N}$ of the graph is then generated by sampling each entry $\mathbf{Q}_{i,j}$ independently according to a Bernoulli distribution with parameter $\Pi_{c(i),c(j)}$, where $c(i)$ is the community label of node i. The model is illustrated in Figure 4 for $B = 3$.

Experiment setting. We consider a DiSBM with $B = 15$ communities, $N = 1000$, $r = 0.1$, $p = 0.4$ and $q = 0.5$. Note that $p > q$ is not necessary: as soon as $p \neq q$, the flow between communities has a preferred orientation, forward if $p > q$ and backward if $q > p$. The community labels are drawn uniformly at random, and the adjacency matrix is generated according to the DiSBM model. The ε in the kernel is chosen as 1. The points are then embedded in $\mathbb{R}^2$ using the first two non-trivial eigenvectors of $\mathbf{L}^s$, with $\theta = 1$. The embedding is then used to estimate the Randers metric and the drift following the method described in the paper.

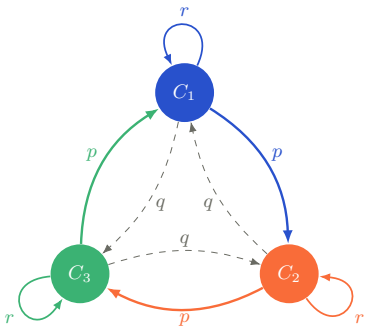
Direction of the drift. Although DiSBM has no underlying manifold, the direction of the flow between communities is known: forward if $p > q$, backward if $q > p$, and none if $p = q$. We measure whether the recovered drift follows it as $p - q$ varies. The N, B and r parameters remains as in the previous paragraph. Let $Y_i \in \mathbb{R}^2$ be the embedding of node i and $c(i)$ its community. The center of community k is

$$\mu_k = \frac{1}{|C_k|} \sum_{j \in C_k} Y_j,$$

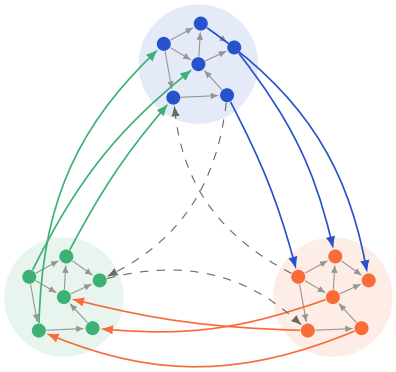
where $C_k = \{j : c(j) = k\}$ is the set of its nodes. The forward direction at community k is the tangent to the cycle of communities, $t_k = \mu_{k+1} - \mu_{k-1}$, with indices taken modulo B. The chord $\mu_{k+1} - \mu_k$ would be off by half the angle between two consecutive communities. The *signed cosine* of node i is the cosine between its drift $\mathbf{V}_i$ and the forward direction at its community,

$$s_i = \frac{\langle \mathbf{V}_i, t_{c(i)} \rangle}{\|\mathbf{V}_i\| \, \|t_{c(i)}\|} \in [-1, 1],$$

which is close to 1 when the drift points forward, close to -1 when it points backward, and spread around 0 when it has no preferred direction. It is signed because it is measured against the forward direction whatever

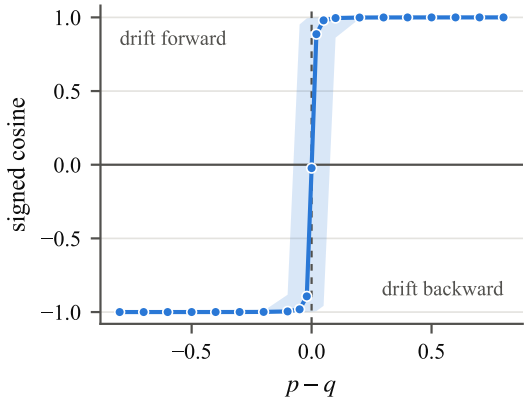


(a) Block model. Each community sends its edges to the next community with probability p, to the previous one with probability q, and to itself with probability r.

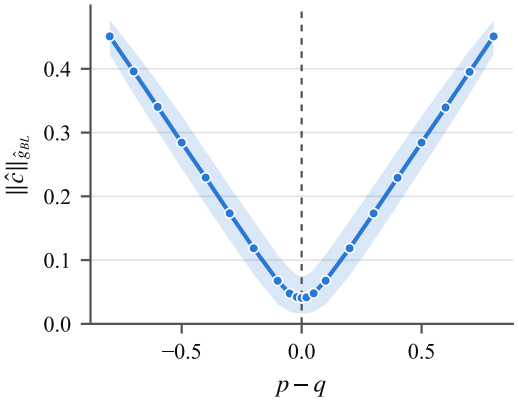


(b) Sampled graph. Intra-community edges (gray) have no preferred orientation, while edges between communities mostly follow the cycle (colored by source), edges going against the cycle are less frequent (dashed).

Figure 4: Cyclic Directed Stochastic Block Model. Illustration with $B = 3$ communities and $p > q$. (a) Connection probabilities of the block matrix Π. (b) A realization of the adjacency matrix $\mathbf{Q}$.



(a) Signed cosine according to the direction of the flow.



(b) Estimated strength of the asymmetry.

Figure 5: Embedding of the DiSBM and its dynamics. Direction of the drift on the DiSBM as a function of $p-q$, with $N = 1000$ nodes, $B = 15$ communities and $r = 0.1$. Lines are medians over the nodes of 5 random graphs, and shaded bands their 10–90% range.

the actual flow, so that it changes sign when the flow is reversed. We also report the estimated strength of the asymmetry, $\|\hat{\mathbf{c}}_i\|_{\hat{g}_{BL}}$. Figure 5 shows that the drift points in the direction of the flow as soon as $p \neq q$: the median signed cosine is already ± 0.89 for $|p-q| = 0.02$, and above 0.99 from $|p-q| = 0.1$ on. When $p = q$, the drift has no preferred direction. The strength of the asymmetry is symmetric in $p-q$ and grows linearly with $|p-q|$.

D PROOFS

This appendix gathers the proofs of the results stated in the main text.

D.1 FORWARD AND BACKWARD TRANSPORT OPERATORS

Proposition 19 (Relation between the forward and backward transport operators)**.** *The backward transport operator $\mathcal{G}'_{\mathcal{F}}$ is equal to the transport operator of the reversed Finsler metric $\overleftarrow{\mathcal{F}}$, and is the $L^2(\mathcal{M}, \mathfrak{m}_{\mathrm{BH}})$-adjoint of $\mathcal{G}_{\mathcal{F}}$, i.e. $\mathcal{G}'_{\mathcal{F}} = (\mathcal{G}_{\mathcal{F}})^* = \mathcal{G}_{\overleftarrow{\mathcal{F}}}$.*

Before proving Proposition 19, we prove the following lemma, which allows us to formally work with the reversed Finsler metric $\overleftarrow{\mathcal{F}}$ and its associated distance.

Lemma 20 (Finsler metric and its reverse)**.** *The reversed Finsler metric $\overleftarrow{\mathcal{F}}$ is a Finsler metric on $\mathcal{M}$. Moreover, its geodesics are the reverse of the geodesics of $\mathcal{F}$, and the geodesic distance $\mathrm{dst}_{\overleftarrow{\mathcal{F}}}$ is the reverse of the geodesic distance $\mathrm{dst}_{\mathcal{F}}$.*

Proof of Lemma 20. First, we show that $\overleftarrow{\mathcal{F}}$ is indeed a Finsler metric. By definition, $\overleftarrow{\mathcal{F}}(x,v) = \mathcal{F}(x,-v)$ for all $x \in \mathcal{M}$ and $v \in T_x\mathcal{M}$. Hence $\overleftarrow{\mathcal{F}}(x,v) > 0$ for all $v \neq 0$, and $\overleftarrow{\mathcal{F}}(x,0) = 0$. For any $t > 0$, we have $\overleftarrow{\mathcal{F}}(x,tv) = \mathcal{F}(x,-tv) = t\mathcal{F}(x,-v) = t\overleftarrow{\mathcal{F}}(x,v)$, satisfying the positive homogeneity property. Finally, the Hessian of $\overleftarrow{\mathcal{F}}^2$ with respect to v is the same as the Hessian of $\mathcal{F}^2$ with respect to $-v$, which is positive definite by the definition of a Finsler metric. Hence, $\overleftarrow{\mathcal{F}}$ satisfies all three properties of a Finsler metric.

For the second claim, let $\gamma : [0,T] \to \mathcal{M}$ be the geodesic of $\mathcal{F}$ connecting two points $x, y \in \mathcal{M}$, so that $\gamma(0) = x$ and $\gamma(T) = y$. Then, the reverse curve $\sigma(t) = \gamma(T-t)$ is a curve connecting y to x. We have that $\dot{\sigma}(t) = -\dot{\gamma}(T-t)$, so that

$$\begin{aligned}
\int_0^T \overleftarrow{\mathcal{F}}(\sigma(t), \dot{\sigma}(t))\mathrm{d}t &= \int_0^T \mathcal{F}(\sigma(t), -\dot{\sigma}(t))\mathrm{d}t \\
&= \int_0^T \mathcal{F}(\gamma(T-t), \dot{\gamma}(T-t))\mathrm{d}t && \text{by the definition of } \sigma \\
&= \int_0^T \mathcal{F}(\gamma(s), \dot{\gamma}(s))\mathrm{d}s && \text{by change of variable } s = T-t \\
&= \mathrm{dst}_{\mathcal{F}}(x,y).
\end{aligned}$$

Since γ is the geodesic of $\mathcal{F}$, it minimizes the integral of $\mathcal{F}$ along curves connecting x and y. Therefore, σ minimizes the integral of $\overleftarrow{\mathcal{F}}$ along curves connecting y and x, and is thus the geodesic of $\overleftarrow{\mathcal{F}}$ connecting y to x. This shows that the geodesics of $\overleftarrow{\mathcal{F}}$ are the reverse of the geodesics of $\mathcal{F}$, and that the geodesic distance $\mathrm{dst}_{\overleftarrow{\mathcal{F}}}(x,y) = \mathrm{dst}_{\mathcal{F}}(y,x)$. □

Applying this lemma, we can now prove Proposition 19.

Proof of Proposition 19. By Lemma 20, $\mathrm{dst}_{\mathcal{F}}(x,y) = \mathrm{dst}_{\overleftarrow{\mathcal{F}}}(y,x)$ for all $x, y \in \mathcal{M}$, so that, denoting by $W_{\overleftarrow{\mathcal{F}}}$ the kernel built from $\overleftarrow{\mathcal{F}}$ in place of $\mathcal{F}$,

$$W(x,y) = \frac{1}{\varepsilon^m} K\left(\frac{\mathrm{dst}_{\mathcal{F}}(x,y)}{\varepsilon}\right) = \frac{1}{\varepsilon^m} K\left(\frac{\mathrm{dst}_{\overleftarrow{\mathcal{F}}}(y,x)}{\varepsilon}\right) = W_{\overleftarrow{\mathcal{F}}}(y,x).$$

Hence $\mathcal{G}'_{\mathcal{F}} f(x) = \int_{\mathcal{M}} W(y,x) f(y) \mathrm{d}\mathfrak{m}_{\mathrm{BH}}(y) = \int_{\mathcal{M}} W_{\overleftarrow{\mathcal{F}}}(x,y) f(y) \mathrm{d}\mathfrak{m}_{\mathrm{BH}}(y) = \mathcal{G}_{\overleftarrow{\mathcal{F}}} f(x)$.

For the adjoint claim, by Fubini's theorem,

$$\begin{aligned}\langle \mathcal{G}_{\mathcal{F}} f, h \rangle &= \iint_{\mathcal{M}\times\mathcal{M}} W(x,y) f(y) h(x) \mathrm{d}\mathfrak{m}_{\mathrm{BH}}(y) \mathrm{d}\mathfrak{m}_{\mathrm{BH}}(x) \\ &= \int_{\mathcal{M}} f(y) \left(\int_{\mathcal{M}} W(x,y) h(x) \mathrm{d}\mathfrak{m}_{\mathrm{BH}}(x) \right) \mathrm{d}\mathfrak{m}_{\mathrm{BH}}(y) = \langle f, \mathcal{G}'_{\mathcal{F}} h \rangle. \end{aligned}$$ □

D.2 Proof of Theorem 1

Despite its simple statement, the proof of Theorem 1 is rather dense, and requires a number of technical lemmas. For ease of reading, we first give a proof sketch, and defer the technical lemmas to the paragraphs below. We strongly encourage the interested reader to read Ohta (2021) and Shen & Shen (2016) for a comprehensive treatment of Finsler geometry, and Coifman & Lafon (2006) for a similar proof in the Riemannian case.

Proof sketch. The proof of Theorem 1 relies on the expansion of the integral operator $\mathcal{G}_{\mathcal{F}}$. To do so, we expand the function f around a point x using the Finsler exponential map $\exp_x^{\mathcal{F}}$. This expansion will be used to expand the integral operator $\mathcal{G}_{\mathcal{F}}$ in terms of the derivatives of f at x, and the moments of the kernel K. This expansion is possible since we assume that the kernel K is sub-exponential, and thus that the integral can be localized around x for small ε. The change of variable through $\exp_x^{\mathcal{F}}$ brings a Jacobian, which we expand in terms of the density σ_{BH}, the S-curvature and the Ricci curvature of $\mathcal{F}$ (Lemma 25). Once this expansion is obtained, we can then use Proposition 19 and the fact that $\mathcal{G}'_{\mathcal{F}} = \mathcal{G}_{\overleftarrow{\mathcal{F}}}$ to obtain the expansion of $\mathcal{G}'_{\mathcal{F}}$. Finally, one can note the following identity:

$$\mathcal{G}_{\mathcal{F}}^{(\theta)}[f] = \frac{1}{{q_\varepsilon}^\theta} \mathcal{G}_{\mathcal{F}} \left[\frac{f}{{q_\varepsilon}^\theta} \right],$$

so that the expansion of $\mathcal{G}_{\mathcal{F}}^{(\theta)}$ can be easily obtained from the expansion of $\mathcal{G}_{\mathcal{F}}$. The final result is obtained by taking the symmetric and antisymmetric parts of the expansion of $\mathcal{G}_{\mathcal{F}}^{(\theta)}$, and applying the proper normalization.

We give here the form of the differential operators that appear in the expansion of $\mathcal{G}_{\mathcal{F}}$. To do so, we start by defining the *kernel moments* of K as follows:

Definition 21 (Kernel moments)**.** *Let $x \in \mathcal{M}$ and $n \in \mathbb{N}$. Using the formal Christoffel symbols $\tilde{\Gamma}^k_{ij}$, the Ricci curvature $Ric^{\mathcal{F}}$ and the S-curvature S (Definition 18) of the Finsler metric $\mathcal{F}$, $\dot{S}$ denoting the derivative of S along the geodesic flow, we define*

$$\begin{aligned} m_n(x) &= \sigma_{\mathrm{BH}}(x) \int_{T_x\mathcal{M}} K(\mathcal{F}(x,v)) v^{\otimes n} \mathrm{d}\lambda_x(v), \\ \Lambda^k(x) &= \sigma_{\mathrm{BH}}(x) \int_{T_x\mathcal{M}} K(\mathcal{F}(x,v)) \tilde{\Gamma}^k_{ij}(x,v) v^i v^j \mathrm{d}\lambda_x(v), \\ s_n(x) &= \sigma_{\mathrm{BH}}(x) \int_{T_x\mathcal{M}} K(\mathcal{F}(x,v)) S(x,v) v^{\otimes n} \mathrm{d}\lambda_x(v), \\ \Upsilon(x) &= \sigma_{\mathrm{BH}}(x) \int_{T_x\mathcal{M}} K(\mathcal{F}(x,v)) \\ &\qquad \times \left(S(x,v)^2 - \dot{S}(x,v) - \frac{1}{3} Ric^{\mathcal{F}}(x,v) \right) \mathrm{d}\lambda_x(v), \end{aligned}$$

where σ_{BH} is the density of the Busemann–Hausdorff measure (Definition 17).

With this definition at hand, we can clarify the definition of the differential operators $\mathscr{A}_1$ and $\mathscr{A}_2$ in Theorem 1:

$$\mathscr{A}_1[f](x) = m_1^i(x)\partial_i f(x) - s_0(x)f,$$
$$\mathscr{A}_2[f](x) = m_2^{ij}(x)\partial_i\partial_j f(x) - (\Lambda^k(x) + 2s_1^k(x))\partial_k f(x) + \Upsilon(x)f(x).$$

Remark 22. *This expansion is a generalization of the moment expansion of the kernel in the Riemannian setting, see e.g. Coifman & Lafon (2006, Lemma 8). Several differences are to be noted:*

- *Due to the non-reversibility of the metric, m_1 does not vanish in general.*
- *The second-order term contains the term $\Lambda^k(x)$, which simplifies in the Riemannian setting, and the curvature term $\Upsilon(x)$, which only multiplies f and reduces to an average of the Ricci curvature in the Riemannian setting.*
- *The S-curvature brings the terms s_0 and s_1, among which a zeroth-order term $-\varepsilon s_0 f$ at first order in ε, and contributes to Υ. These contributions vanish in the Riemannian setting, and more generally for Berwald metrics, whose S-curvature with respect to $\mathfrak{m}_{\mathrm{BH}}$ is identically zero (Shen & Shen, 2016, Proposition 4.3).*

We now turn to the technical lemmas that are used in the proof of Theorem 1.

Main lemmas

We start by giving the expansion of a function f around a point x using the Finsler exponential map $\exp_x^{\mathscr{F}}$. This expansion will be used to expand the integral operator $\mathscr{G}_{\mathscr{F}}$ in terms of the derivatives of f at x, and the moments of the kernel K. It is to be contrasted with the expansion using the Riemannian exponential map $\exp_x^{\alpha}$ (Coifman & Lafon, 2006).

Lemma 23 (Function expansion using Finsler metric)**.** *Let $x \in \mathscr{M}$ and $v \in T_x\mathscr{M}$ be a unit tangent vector at x. Let also $\varepsilon \in \mathbb{R}_*^+$. Then*

$$f(\exp_x^{\mathscr{F}}(\varepsilon v)) = f(x) + \varepsilon\partial_i f(x)v^i + \frac{\varepsilon^2}{2}\left(\partial_j\partial_i f(x) - \partial_k f(x)\tilde{\Gamma}_{ij}^k(x,v)\right)v^j v^i + O(\varepsilon^3), \tag{12}$$

where $\tilde{\Gamma}_{ij}^k$ is the Christoffel symbol of the Finsler metric $\mathscr{F}$[3]*.*

Proof. Let $\gamma(\varepsilon) = \exp_x^{\mathscr{F}}(\varepsilon v)$, and $\phi(\varepsilon) = f(\gamma(\varepsilon))$. We have that $\phi : \mathbb{R} \to \mathbb{R}$ and $\phi(0) = f(x)$. Taylor expansion of ϕ around 0 yields:

$$\phi(\varepsilon) = \phi(0) + \phi'(0)\varepsilon + \frac{1}{2}\phi''(0)\varepsilon^2 + O(\varepsilon^3)$$

Now we need to express $\phi'(0)$ and $\phi''(0)$ in terms of the derivatives of f at x.

For the *first derivative*, by the chain rule,

$$\phi'(0) = \frac{d}{d\varepsilon}f(\gamma(\varepsilon))\bigg|_{\varepsilon=0} = \partial_i f(\gamma(\varepsilon))\dot{\gamma}^i(\varepsilon)\bigg|_{\varepsilon=0} = \partial_i f(x)v^i$$

[3]The Christoffel symbols of a Finsler metric are also called *formal Christoffel symbols* (Ohta, 2021)

For the *second derivative*, by the product rule,

$$\begin{aligned}\phi''(\varepsilon) &= \frac{d}{d\varepsilon}(df(\gamma(\varepsilon))\dot\gamma(\varepsilon)) \\ &= \underbrace{\frac{d}{d\varepsilon}\Big(df(\gamma(\varepsilon))\Big)\dot\gamma^i(\varepsilon)}_{\text{(A)}} + \underbrace{df(\gamma(\varepsilon))\ddot\gamma^i(\varepsilon)}_{\text{(B)}}.\end{aligned}$$

Using the chain rule again, the first term (A) gives

$$\frac{d}{d\varepsilon}\Big(df(\gamma(\varepsilon))\Big)\dot\gamma^i(\varepsilon)\Big|_{\varepsilon=0} = \partial_j\partial_i f(\gamma(\varepsilon))\dot\gamma^j(\varepsilon)\dot\gamma^i(\varepsilon)\Big|_{\varepsilon=0} = \partial_j\partial_i f(x)v^j v^i.$$

For the second term (B), since γ is a geodesic under the Finsler metric $\mathscr{F}$, the geodesic equation (Ohta, 2021, Chap. 3.3) gives

$$\ddot\gamma^i(\varepsilon) = -\tilde G^i(\gamma(\varepsilon),\dot\gamma(\varepsilon)) = -\tilde\Gamma^i_{jk}(x,\dot\gamma(\varepsilon))\dot\gamma(\varepsilon)^j\dot\gamma(\varepsilon)^k$$

where $\tilde G^i$ is the geodesic spray coefficient of the Finsler metric $\mathscr{F}$, and $\tilde\Gamma$ is the Christoffel symbol of the Finsler metric $\mathscr{F}$. Consequently,

$$df(\gamma(\varepsilon))\ddot\gamma^k(\varepsilon)\Big|_{\varepsilon=0} = -\partial_k f(x)\tilde\Gamma^k_{ij}(x,v)v^i v^j,$$

where we changed the index from i to k. Hence, the second derivative of ϕ at 0 is

$$\phi''(0) = (\partial_j\partial_i f(x) - \partial_k f(x)\tilde\Gamma^k_{ij}(x,v))v^j v^i.$$

The Taylor expansion of f around x is thus

$$f(\exp^{\mathscr{F}}_x(\varepsilon v)) = f(x) + \varepsilon\partial_i f(x)v^i + \frac{\varepsilon^2}{2}\left(\partial_j\partial_i f(x) - \partial_k f(x)\tilde\Gamma^k_{ij}(x,v)\right)v^j v^i + O(\varepsilon^3).$$

This concludes the proof. □

To use the previous expansion, we need to localize the integral operator $\mathscr{G}_{\mathscr{F}}$ around a point $x\in\mathscr{M}$, so that the integral is instead taken over the tangent space $T_x\mathscr{M}$, which can be identified with $\mathbb{R}^m$. This is possible thanks to the decay of the kernel K, which allows us to restrict the integral to a small neighborhood of x.

Lemma 24 (Integral localization)**.** *Let K satisfy (A1) and $x\in\mathscr{M}$. Then, for any $f\in L_1(\mathscr{M})$ smooth enough,*

$$\mathscr{G}_{\mathscr{F}}[f](x) = \int_{\mathscr{B}_x(r_0/\varepsilon)} K\left(\mathscr{F}(x,v)\right) f(\exp^{\mathscr{F}}_x(\varepsilon v))\mathscr{J}_x(\varepsilon v)\mathrm{d}\lambda_x(v) + O(\varepsilon^3),$$

where $\mathscr{J}_x$ is the Jacobian of the change of variable through the exponential map, and $0<r_0<i_{\mathscr{M}}$ is a constant smaller than the injectivity radius of $\mathscr{M}$, which does not depend on x.

Proof. We start by splitting the integral into two parts, one over the ball $B^+_{\mathscr{F}}(x,r_0) = \{y\in\mathscr{M} : \mathrm{dst}_{\mathscr{F}}(x,y) < r_0\}$ and one over its complement.

On the far region. Since $\mathscr{M}$ is compact, f is bounded, and we denote $M_f = \sup_{y\in\mathscr{M}}|f(y)| < \infty$. By definition, it holds that $\mathrm{dst}_{\mathscr{F}}(x,y)\geq r_0$ for all $y\in\mathscr{M}\setminus B^+_{\mathscr{F}}(x,r_0)$. Using the sub-exponential property of the kernel (A1), it holds:

$$\begin{aligned}\varepsilon^{-m}\int_{\mathscr{M}\setminus B^+_{\mathscr{F}}(x,r_0)} K\left(\frac{\mathrm{dst}_{\mathscr{F}}(x,y)}{\varepsilon}\right) f(y)\mathrm{d}\mathfrak{m}_{\mathrm{BH}}(y) &\leq \varepsilon^{-m}M_f\int_{\mathscr{M}\setminus B^+_{\mathscr{F}}(x,r_0)} C_K\exp\left(-\frac{\nu_K r_0}{\varepsilon}\right)\mathrm{d}\mathfrak{m}_{\mathrm{BH}}(y) \\ &\leq M_f C_K\varepsilon^{-m}\exp\left(-\frac{\nu_K r_0}{\varepsilon}\right)\mathfrak{m}_{\mathrm{BH}}(\mathscr{M}),\end{aligned}$$

which is $O(\varepsilon^N)$ for any $N>0$, and in particular $O(\varepsilon^3)$.

On the near region. Because $r_0 < i_{\mathscr{M}}$, the exponential map $\exp_x^{\mathscr{F}}$ is a diffeomorphism from the tangent ball $\mathscr{B}_x(r_0) \subset T_x\mathscr{M}$ onto the forward metric ball $B^+_{\mathscr{F}}(x,r_0) \subset \mathscr{M}$ (it is C^1 at the origin and smooth elsewhere), so that we can do the change of variable $y = \exp_x^{\mathscr{F}}(\varepsilon v)$, with $v \in \mathscr{B}_x(r_0/\varepsilon)$. For such v, $\varepsilon\mathscr{F}(x,v) < r_0 < i_{\mathscr{M}}$, so that the Finslerian distance writes $\mathrm{dst}_{\mathscr{F}}(x, \exp_x^{\mathscr{F}}(\varepsilon v)) = \varepsilon\mathscr{F}(x,v)$. The integral over the near region can then be rewritten as:

$$\varepsilon^{-m}\int_{B^+_{\mathscr{F}}(x,r_0)} K\left(\frac{\mathrm{dst}_{\mathscr{F}}(x,y)}{\varepsilon}\right) f(y)\mathrm{d}\mathfrak{m}_{\mathrm{BH}}(y) = \int_{\mathscr{B}_x(r_0/\varepsilon)} K\left(\mathscr{F}(x,v)\right) f(\exp_x^{\mathscr{F}}(\varepsilon v))\mathscr{J}_x(\varepsilon v)\mathrm{d}\lambda_x(v),$$

where $\mathscr{J}_x(u)$ is the Jacobian of the change of variable $y = \exp_x^{\mathscr{F}}(u)$, i.e. $\mathrm{d}\mathfrak{m}_{\mathrm{BH}}(y) = \mathscr{J}_x(u)\mathrm{d}\lambda_x(u)$. The factor ε^{-m} is absorbed in the change of variable $u = \varepsilon v$. Finally, we can combine the two regions to get the desired result. □

Lemma 25 (Jacobian expansion)**.** *Let $x \in \mathscr{M}$, $v \in T_x\mathscr{M} \setminus \{0\}$, and $\varepsilon > 0$ with $\varepsilon\mathscr{F}(x,v) < i_{\mathscr{M}}$. Then the Jacobian $\mathscr{J}_x$ of Lemma 24 satisfies*

$$\mathscr{J}_x(\varepsilon v) = \sigma_{\mathrm{BH}}(x)\left(1 - \varepsilon S(x,v) + \frac{\varepsilon^2}{2}\left(S(x,v)^2 - \dot{S}(x,v) - \frac{1}{3}\mathit{Ric}^{\mathscr{F}}(v)\right) + \mathcal{O}(\varepsilon^3)\right),$$

where S and $\dot{S}$ are the S-curvature (Definition 18) and its derivative along the geodesic flow, and where the constant in $\mathcal{O}$ depends neither on x nor on v.

Proof. By definition of the Jacobian, $\mathscr{J}_x$ has the form:

$$\mathscr{J}_x(\varepsilon v) = \underbrace{\sigma_{\mathrm{BH}}(\exp_x^{\mathscr{F}}(\varepsilon v))}_{\text{change of measure}(\star)} \times \underbrace{\det(d(\exp_x^{\mathscr{F}})_{\varepsilon v})}_{\text{change of variable}(\star\star)},$$

where $\sigma_{\mathrm{BH}}(\exp_x^{\mathscr{F}}(\varepsilon v))$ comes from the change of measure from the Busemann–Hausdorff measure to the Lebesgue measure on $T_x\mathscr{M}$, and $\det(d(\exp_x^{\mathscr{F}})_{\varepsilon v})$ comes from the change of variable $y = \exp_x^{\mathscr{F}}(\varepsilon v)$. This proof is split in two parts, the expansion of $\sigma_{\mathrm{BH}}(\exp_x^{\mathscr{F}}(\varepsilon v))$ and the expansion of $\det(d(\exp_x^{\mathscr{F}})_{\varepsilon v})$.

Expansion of $(\star)$. By definition of the distortion (Definition 18), we have, for every $u \in T_z\mathscr{M} \setminus \{0\}$, that $\sigma_{\mathrm{BH}}(z) = \exp(-\tau(z,u))\sqrt{\det g(z,u)}$. The left-hand side does not depend on u, so that we may choose the velocity of the geodesic γ, with $\gamma(0) = x$ and $\dot\gamma(0) = v$, at both of its ends:

$$\sigma_{\mathrm{BH}}(x) = \exp(-\tau(x,v))\sqrt{\det g(x,v)}, \qquad \sigma_{\mathrm{BH}}(\exp_x^{\mathscr{F}}(\varepsilon v)) = \exp(-\tau(\gamma(\varepsilon),\dot\gamma(\varepsilon)))\sqrt{\det g(\gamma(\varepsilon),\dot\gamma(\varepsilon))}.$$

Taking the ratio of these two identities, we have that

$$\sigma_{\mathrm{BH}}(\exp_x^{\mathscr{F}}(\varepsilon v)) = \sigma_{\mathrm{BH}}(x)\frac{\sqrt{\det g(\gamma(\varepsilon),\dot\gamma(\varepsilon))}}{\sqrt{\det g(x,v)}}\exp\left(\tau(x,v) - \tau(\gamma(\varepsilon),\dot\gamma(\varepsilon))\right).$$

By definition of the S-curvature (Definition 18), and since $t \mapsto \gamma(s+t)$ is the geodesic with initial conditions $(\gamma(s),\dot\gamma(s))$, the first two derivatives of $t \mapsto \tau(\gamma(t),\dot\gamma(t))$ are $S(\gamma(t),\dot\gamma(t))$ and $\dot S(\gamma(t),\dot\gamma(t))$, so that

$$\tau(\gamma(\varepsilon),\dot\gamma(\varepsilon)) - \tau(x,v) = \varepsilon S(x,v) + \frac{\varepsilon^2}{2}\dot S(x,v) + O(\varepsilon^3).$$

Hence, we have that

$$\sigma_{\mathrm{BH}}(\exp_x^{\mathscr{F}}(\varepsilon v)) = \sigma_{\mathrm{BH}}(x)\frac{\sqrt{\det g(\gamma(\varepsilon),\dot\gamma(\varepsilon))}}{\sqrt{\det g(x,v)}}\left(1 - \varepsilon S(x,v) + \frac{\varepsilon^2}{2}\left(S(x,v)^2 - \dot S(x,v)\right) + O(\varepsilon^3)\right).$$

Expansion of $(\star\star)$. Let $(e_1,\dots,e_m)$ be a basis of $T_x\mathscr{M}$ which is orthonormal for $g(x,v)$, and such that $e_m = v/\mathscr{F}(x,v)$. For each i, we consider the variation of geodesics $\gamma_i(t,s) = \exp_x^{\mathscr{F}}(t(v+se_i))$, and we define

$$J_i(t) = \partial_s\gamma_i(t,0) = d(\exp_x^{\mathscr{F}})_{tv}(te_i),$$

which is a Jacobi field along the geodesic $\gamma(t) = \exp_x^{\mathscr{F}}(tv)$ (Ohta, 2021, Prop. 5.5). For the sake of completeness, we check here its initial conditions $J_i(0)$ and $\dot{J}_i(0)$. First,

$$J_i(0) = \partial_s\gamma_i(0,0) = \partial_s x = 0.$$

Second, $\partial_t\gamma_i(0,s) = v+se_i$, and by the symmetry of second derivatives[4], we have that $\partial_t\partial_s\gamma_i(0,0) = \partial_s\partial_t\gamma_i(0,0)$. Thus, we get $\dot{J}_i(0) = e_i$. As a Jacobi field, J_i satisfies the Jacobi equation (Ohta, 2021, Def. 5.1):

$$\ddot{J}_i(t) + R(t)J_i(t) = 0$$

where R is the curvature tensor of the Finsler metric $\mathscr{F}$, evaluated along the geodesic $\gamma(t)$. By definition, for $t=0$, we have that $R(0)J_i(0) = 0$. Hence, the Jacobi equation at $t=0$ gives $\ddot{J}_i(0) = 0$. Differentiating the Jacobi equation gives

$$\dddot{J}_i(t) + \dot{R}(t)J_i(t) + R(t)\dot{J}_i(t) = 0,$$

which, evaluated at $t=0$, gives $\dddot{J}_i(0) = -R(0)\dot{J}_i(0) = -R(0)e_i$. To write these relations in matrix form, we express the $J_i(t)$ in the basis $(E_1(t),\dots,E_m(t))$ of $T_{\gamma(t)}\mathscr{M}$ obtained by parallel transport of $(e_1,\dots,e_m)$ along γ, for the Chern connection with reference vector $\dot\gamma$. This basis remains orthonormal for $g(\gamma(t),\dot\gamma(t))$, and the covariant derivatives along γ are the ordinary derivatives of the components (Ohta, 2021). Letting $\mathbf{J}(t)$ be the matrix of the components of $J_1(t),\dots,J_m(t)$ in this basis, and $\mathbf{R}(t)$ the matrix of $R(t)$, we have that $\mathbf{J}(0) = \ddot{\mathbf{J}}(0) = 0$, $\dot{\mathbf{J}}(0) = I_m$ and $\dddot{\mathbf{J}}(0) = -\mathbf{R}(0)$. Hence, the Taylor expansion of $\mathbf{J}$ around 0 is given by

$$\mathbf{J}(t) = tI_m - \frac{t^3}{6}\mathbf{R}(0) + O(t^4).$$

Since $J_i(\varepsilon)/\varepsilon = d(\exp_x^{\mathscr{F}})_{\varepsilon v}e_i$, the matrix $\mathbf{J}(\varepsilon)/\varepsilon$ represents $d(\exp_x^{\mathscr{F}})_{\varepsilon v}$ in the orthonormal bases (e_i) of $T_x\mathscr{M}$ and $(E_i(\varepsilon))$ of $T_{\gamma(\varepsilon)}\mathscr{M}$. Its determinant is thus

$$\begin{aligned}\det\left(\frac{1}{\varepsilon}\mathbf{J}(\varepsilon)\right) &= \det\left(I_m - \frac{\varepsilon^2}{6}\mathbf{R}(0) + O(\varepsilon^3)\right)\\ &= 1 - \frac{\varepsilon^2}{6}\operatorname{tr}(\mathbf{R}(0)) + O(\varepsilon^3)\\ &= 1 - \frac{\varepsilon^2}{6}\mathrm{Ric}^{\mathscr{F}}(v) + O(\varepsilon^3),\end{aligned}$$

where $\mathrm{Ric}^{\mathscr{F}}(v)$ is the Ricci curvature of $\mathscr{F}$ in the direction v, i.e. the trace of R_v in a basis which is orthonormal for $g(x,v)$. Since $e_m = v/\mathscr{F}(x,v)$ and $R_v v = 0$, only the $m-1$ transverse directions contribute to this trace.

Going back to the original basis, the determinant in $(\star\star)$ is taken in the coordinate bases of $T_x\mathscr{M}$ and $T_{\gamma(\varepsilon)}\mathscr{M}$, and not in the orthonormal ones. Let $\mathbf{P}$ and $\mathbf{Q}$ be the matrices whose columns are the coordinates of $e_1,\dots,e_m$ and of $E_1(\varepsilon),\dots,E_m(\varepsilon)$. The matrix of $d(\exp_x^{\mathscr{F}})_{\varepsilon v}$ in the coordinate bases is $\mathbf{Q}(\mathbf{J}(\varepsilon)/\varepsilon)\mathbf{P}^{-1}$, so that

$$\det\left(d(\exp_x^{\mathscr{F}})_{\varepsilon v}\right) = \frac{\det\mathbf{Q}}{\det\mathbf{P}}\det\left(\frac{1}{\varepsilon}\mathbf{J}(\varepsilon)\right).$$

[4]While the exponential map is only C^1 at 0, the Jacobi fields admit a C^∞ extension, see (Bao et al., 2012, Lemma 5.4.1).

The orthonormality of the two bases reads $\mathbf{P}^\mathsf{T} g(x,v)\mathbf{P} = I_m$ and $\mathbf{Q}^\mathsf{T} g(\gamma(\varepsilon),\dot\gamma(\varepsilon))\mathbf{Q} = I_m$, so that $\det \mathbf{P} = \det g(x,v)^{-1/2}$ and $\det \mathbf{Q} = \det g(\gamma(\varepsilon),\dot\gamma(\varepsilon))^{-1/2}$. Both are positive when (e_i) is positively oriented, since the determinant of the coordinates of $(E_i(t))$ is then continuous in t, never vanishes, and equals $\det \mathbf{P}$ at $t=0$. Hence

$$\det\left(d(\exp_x^{\mathscr{F}})_{\varepsilon v}\right) = \frac{\sqrt{\det g(x,v)}}{\sqrt{\det g(\gamma(\varepsilon),\dot\gamma(\varepsilon))}}\left(1 - \frac{\varepsilon^2}{6}\mathrm{Ric}^{\mathscr{F}}(v) + O(\varepsilon^3)\right).$$

Conclusion. Multiplying the expansions of $(\star)$ and $(\star\star)$, the two ratios of $\sqrt{\det g}$ cancel exactly, and

$$\mathscr{J}_x(\varepsilon v) = \sigma_{\mathrm{BH}}(x)\left(1 - \varepsilon S(x,v) + \frac{\varepsilon^2}{2}\left(S(x,v)^2 - \dot S(x,v)\right) + O(\varepsilon^3)\right)\left(1 - \frac{\varepsilon^2}{6}\mathrm{Ric}^{\mathscr{F}}(v) + O(\varepsilon^3)\right).$$

Expanding the product, the cross term $\varepsilon S(x,v)\cdot\frac{\varepsilon^2}{6}\mathrm{Ric}^{\mathscr{F}}(v)$ is of order ε^3, so that

$$\mathscr{J}_x(\varepsilon v) = \sigma_{\mathrm{BH}}(x)\left(1 - \varepsilon S(x,v) + \frac{\varepsilon^2}{2}\left(S(x,v)^2 - \dot S(x,v) - \frac{1}{3}\mathrm{Ric}^{\mathscr{F}}(v)\right) + \mathcal{O}(\varepsilon^3)\right).$$

This concludes the proof. □

Proof of the moment expansion of the kernel

We can now prove the moment expansion of the kernel of Theorem 1. The proof is a direct application of the previous lemmas. We recall that it states that, for any $f : \mathscr{M} \to \mathbb{R}$ smooth enough,

$$\mathscr{G}_{\mathscr{F}}[f](x) = m_0 f + \varepsilon \mathscr{A}_1[f](x) + \frac{\varepsilon^2}{2}\mathscr{A}_2[f](x) + \mathcal{O}(\varepsilon^3).$$

Proof. The proof utilizes Lemma 24 and both expansions of $f(\exp_x^{\mathscr{F}}(\varepsilon v))$ and $\mathscr{J}_x(\varepsilon v)$ in ε (Lemmas 23 and 25). Putting those together, it holds that:

$$\begin{aligned}
\mathscr{G}_{\mathscr{F}}[f](x) &= \int_{\mathscr{B}_x(r_0/\varepsilon)} K(\mathscr{F}(x,v)) f(\exp_x^{\mathscr{F}}(\varepsilon v)) \mathscr{J}_x(\varepsilon v)\mathrm{d}\lambda_x(v) + O(\varepsilon^3) \\
&= \sigma_{\mathrm{BH}}(x)\int_{\mathscr{B}_x(r_0/\varepsilon)} K(\mathscr{F}(x,v))\Big[f(x) + \varepsilon\partial_i f(x) v^i \\
&\quad + \frac{\varepsilon^2}{2}\left(\partial_j\partial_i f(x) - \partial_k f(x)\tilde\Gamma^k_{ij}(x,v)\right)v^j v^i + O(\varepsilon^3)\Big] \\
&\quad \times \left[1 - \varepsilon S(x,v) + \frac{\varepsilon^2}{2}\left(S(x,v)^2 - \dot S(x,v) - \frac{1}{3}\mathrm{Ric}^{\mathscr{F}}(x,v)\right) + O(\varepsilon^3)\right]\mathrm{d}\lambda_x(v) + O(\varepsilon^3).
\end{aligned}$$

Expanding the product and keeping the terms up to order ε^2, the integrand is $K(\mathscr{F}(x,v))$ times

$$\begin{aligned}
&f(x) + \varepsilon\left(\partial_i f(x) v^i - S(x,v) f(x)\right) \\
&\quad + \frac{\varepsilon^2}{2}\Big[\partial_i\partial_j f(x) v^i v^j - \partial_k f(x)\tilde\Gamma^k_{ij}(x,v) v^i v^j - 2S(x,v)\partial_k f(x) v^k \\
&\qquad + \left(S(x,v)^2 - \dot S(x,v) - \frac{1}{3}\mathrm{Ric}^{\mathscr{F}}(x,v)\right) f(x)\Big],
\end{aligned}$$

up to terms of order ε^3 and higher. The moments of Definition 21 are integrals over the whole tangent space $T_x\mathscr{M}$, whereas this integral is over $\mathscr{B}_x(r_0/\varepsilon)$. As in Lemma 24, the sub-exponential decay of K (A1) makes

the integral over $T_x\mathscr{M} \setminus \mathscr{B}_x(r_0/\varepsilon)$ exponentially small, so that the domain can be extended to $T_x\mathscr{M}$ up to $O(\varepsilon^3)$. By collecting the terms of order 0, ε and ε^2, and including back the density $\sigma_{\mathrm{BH}}(x)$, we recover the definitions of the moments (Definition 21) and thus

$$\mathscr{G}_{\mathscr{F}}[f](x) = m_0(x)f(x) + \varepsilon\left(m_1^i(x)\partial_i f(x) - s_0(x)f(x)\right) + \frac{\varepsilon^2}{2}\left(m_2^{ij}(x)\partial_i\partial_j f(x) - \left(\Lambda^k(x) + 2s_1^k(x)\right)\partial_k f(x) + \Upsilon(x)f(x)\right) + \mathcal{O}(\varepsilon^3).$$

Hence to recover the operators $\mathscr{A}_1$ and $\mathscr{A}_2$ of Theorem 1, we define:

$$\mathscr{A}_1[f](x) = m_1^i(x)\partial_i f(x) - s_0(x)f(x),$$
$$\mathscr{A}_2[f](x) = m_2^{ij}(x)\partial_i\partial_j f(x) - \left(\Lambda^k(x) + 2s_1^k(x)\right)\partial_k f(x) + \Upsilon(x)f(x).$$

□

D.3 Proof of Theorem 2

This subsection is devoted to the proof of Theorem 2, which builds upon the previous expansion of the kernel. To do so, we first prove some properties of the reverse Finsler metric and its nonlinear connection.

Lemmas on the reverse Finsler metric and nonlinear connection

Some of the following properties are stated in Ohta (2021, Section 2.5), but are not proved there. We provide here proofs for completeness.

Proposition 26 (Geodesic spray of the reverse Finsler metric)**.** *Let $\overleftarrow{\mathscr{F}}(x,v) = \mathscr{F}(x,-v)$, and let G^k and $\overleftarrow{G}^k$ be the geodesic spray coefficients of $\mathscr{F}$ and $\overleftarrow{\mathscr{F}}$, respectively. Then, we have that $\overleftarrow{G}^k(x,v) = G^k(x,-v)$ for all $v \in T_x\mathscr{M}$.*

Proof. We first recall the relevant notations. For $x \in \mathscr{M}$ and $v \in T_x\mathscr{M}$, let $g_{ij}(x,v)$ and $\overleftarrow{g}_{ij}(x,v)$ be the fundamental tensors of $\mathscr{F}$ and $\overleftarrow{\mathscr{F}}$, respectively. By Definition 12, we have that:

$$g_{ij}(x,v) = \frac{1}{2}\frac{\partial^2 \mathscr{F}^2(x,v)}{\partial v^i \partial v^j} \quad \text{and} \quad \overleftarrow{g}_{ij}(x,v) = \frac{1}{2}\frac{\partial^2 \overleftarrow{\mathscr{F}}^2(x,v)}{\partial v^i \partial v^j}.$$

And the geodesic spray coefficients (Ohta, 2021) are given by:

$$G^k(x,v) = \frac{1}{4}g^{kl}(x,v)\left(\frac{\partial^2 \mathscr{F}^2(x,v)}{\partial x^i \partial v^l}v^i - \frac{\partial \mathscr{F}^2(x,v)}{\partial x^l}\right),$$
$$\overleftarrow{G}^k(x,v) = \frac{1}{4}\overleftarrow{g}^{kl}(x,v)\left(\frac{\partial^2 \overleftarrow{\mathscr{F}}^2(x,v)}{\partial x^i \partial v^l}v^i - \frac{\partial \overleftarrow{\mathscr{F}}^2(x,v)}{\partial x^l}\right).$$

To start, we show that the fundamental tensor of the reverse Finsler metric is related to the fundamental tensor of the original Finsler metric by $\overleftarrow{g}_{ij}(x,v) = g_{ij}(x,-v)$. Indeed:

$$\overleftarrow{g}_{ij}(x,v) = \frac{1}{2}\frac{\partial^2 \overleftarrow{\mathscr{F}}^2(x,v)}{\partial v^i \partial v^j} = \frac{1}{2}\frac{\partial^2 \mathscr{F}^2(x,-v)}{\partial v^i \partial v^j} = g_{ij}(x,-v).$$

This implies that $\overleftarrow{g}^{kl}(x,v) = g^{kl}(x,-v)$. Similarly,

$$\frac{\partial^2 \overleftarrow{\mathscr{F}}^2(x,v)}{\partial x^i \partial v^l} = -\frac{\partial^2 \mathscr{F}^2(x,-v)}{\partial x^i \partial v^l} \quad \text{and} \quad \frac{\partial \overleftarrow{\mathscr{F}}^2(x,v)}{\partial x^l} = \frac{\partial \mathscr{F}^2(x,-v)}{\partial x^l}.$$

Putting everything together, we have that:

$$\begin{aligned}\overleftarrow{G}^k(x,v) &= \frac{1}{4}g^{kl}(x,-v)\left(-\frac{\partial^2\mathscr{F}^2(x,-v)}{\partial x^i\partial v^l}v^i - \frac{\partial\mathscr{F}^2(x,-v)}{\partial x^l}\right)\\ &= \frac{1}{4}g^{kl}(x,-v)\left(\frac{\partial^2\mathscr{F}^2(x,-v)}{\partial x^i\partial v^l}(-v)^i - \frac{\partial\mathscr{F}^2(x,-v)}{\partial x^l}\right)\\ &= G^k(x,-v).\end{aligned}$$

This concludes the proof. □

Next, we define the nonlinear connection of a Finsler metric, which is a key quantity in Finsler geometry. It is used to define the curvature tensor and the Ricci curvature of a Finsler metric, which appeared in Theorem 1.

Definition 27 (Nonlinear connection)**.** *The nonlinear connection of a Finsler metric $\mathscr{F}$ is defined as:*

$$N^i_j(x,v) = \frac{\partial G^i}{\partial v^j}(x,v).$$

Lemma 28 (Nonlinear connection under reversal)**.** *Let $\overleftarrow{N}{}^i_j$ and N^i_j be the nonlinear connections of $\overleftarrow{\mathscr{F}}$ and $\mathscr{F}$ respectively. Then $\overleftarrow{N}{}^i_j(x,v) = -N^i_j(x,-v)$.*

Proof. Using Proposition 26 ($\overleftarrow{G}{}^i(x,v) = G^i(x,-v)$) and the chain rule,

$$\overleftarrow{N}{}^i_j(x,v) = \frac{\partial\overleftarrow{G}{}^i}{\partial v^j}(x,v) = \frac{\partial}{\partial v^j}\left[G^i(x,-v)\right] = -\frac{\partial G^i}{\partial v^j}(x,-v) = -N^i_j(x,-v).$$ □

This allows us to prove the following property of the Ricci curvature of the reverse Finsler metric.

Proposition 29 (Ricci curvature of the reverse Finsler metric)**.** *Let $\mathrm{Ric}^{\overleftarrow{\mathscr{F}}}$ and $\mathrm{Ric}^{\mathscr{F}}$ be the Ricci curvatures of $\overleftarrow{\mathscr{F}}$ and $\mathscr{F}$ respectively. Then $\mathrm{Ric}^{\overleftarrow{\mathscr{F}}}(x,v) = \mathrm{Ric}^{\mathscr{F}}(x,-v)$.*

Proof. We first introduce a few auxiliary quantities. The curvature tensor of a Finsler metric $\mathscr{F}$ is defined as

$$R^i_j(x,v) = \frac{\partial G^i}{\partial x^j}(x,v) - \sum_{k=1}^m\left(\frac{\partial N^i_j}{\partial x^k}(x,v)v^k - \frac{\partial N^i_j}{\partial v^k}(x,v)G^k(x,v)\right) - \sum_{k=1}^m N^i_k(x,v)N^k_j(x,v),$$

inducing a linear endomorphism $R_v(x,\cdot): T_x\mathscr{M}\to T_x\mathscr{M}$ defined as $R_v(x,w) = \sum_{i,j=1}^m R^i_j(x,v)w^j\frac{\partial}{\partial x^i}\big|_x$. The Ricci curvature is the trace of this endomorphism, for $v\neq 0$:

$$\mathrm{Ric}^{\mathscr{F}}(x,v) = \sum_{i=1}^{m-1} g(x,v)_{kl}R_v(x,e_i)^k e^l_i,$$

where $\mathscr{B} = \{v/\mathscr{F}(x,v)\}\cup\{e_i\}_{i=1}^{m-1}\subset T_x\mathscr{M}$ is an orthonormal basis with respect to the fundamental tensor $g(x,v)$ of $\mathscr{F}$. We similarly define $\overleftarrow{R}{}^i_j(x,v)$, $\overleftarrow{R}_v(x,w)$ and $\mathrm{Ric}^{\overleftarrow{\mathscr{F}}}(x,v)$ for the reverse Finsler metric $\overleftarrow{\mathscr{F}}$.

By Lemma 28, $\overleftarrow{N}{}^i_j(x,v) = -N^i_j(x,-v)$. Differentiating this identity in x^k (which passes through unchanged, as in the proof of Proposition 26) and in v^k (which picks up an extra sign, as in the proof of Lemma 28 itself) gives

$$\frac{\partial\overleftarrow{N}{}^i_j}{\partial x^k}(x,v) = -\frac{\partial N^i_j}{\partial x^k}(x,-v), \qquad \frac{\partial\overleftarrow{N}{}^i_j}{\partial v^k}(x,v) = \frac{\partial N^i_j}{\partial v^k}(x,-v).$$

Together with $\overleftarrow{G}^k(x,v) = G^k(x,-v)$ (Proposition 26), we get

$$\frac{\partial \overleftarrow{N}^i_j}{\partial x^k}(x,v)v^k - \frac{\partial \overleftarrow{N}^i_j}{\partial v^k}(x,v)\overleftarrow{G}^k(x,v) = -\frac{\partial N^i_j}{\partial x^k}(x,-v)v^k - \frac{\partial N^i_j}{\partial v^k}(x,-v)G^k(x,-v)$$
$$= \frac{\partial N^i_j}{\partial x^k}(x,-v)(-v)^k - \frac{\partial N^i_j}{\partial v^k}(x,-v)G^k(x,-v),$$

and $\sum_k \overleftarrow{N}^i_k(x,v)\overleftarrow{N}^k_j(x,v) = \sum_k N^i_k(x,-v)N^k_j(x,-v)$ (product of two sign-flipped terms). By Proposition 26:

$$\frac{\partial \overleftarrow{G}^i}{\partial x^j}(x,v) = \frac{\partial G^i}{\partial x^j}(x,-v).$$

Putting everything together, we have that

$$\overleftarrow{R}^i_j(x,v) = \frac{\partial G^i}{\partial x^j}(x,-v) - \sum_{k=1}^{m}\left(\frac{\partial N^i_j}{\partial x^k}(x,-v)(-v)^k - \frac{\partial N^i_j}{\partial v^k}(x,-v)G^k(x,-v)\right)$$
$$-\sum_{k=1}^{m} N^i_k(x,-v)N^k_j(x,-v) = R^i_j(x,-v),$$

so that $\overleftarrow{R}_v(x,w) = R_{-v}(x,w)$. Since the fundamental tensor of $\overleftarrow{\mathscr{F}}$ is $\overleftarrow{g}(x,v) = g(x,-v)$ (Proposition 26's proof), a basis orthonormal for $g(x,-v)$ is also orthonormal for $\overleftarrow{g}(x,v)$, so we may use the same basis $\mathscr{B}$ for both. Hence

$$\mathrm{Ric}^{\overleftarrow{\mathscr{F}}}(x,v) = \sum_{i=1}^{m-1} \overleftarrow{g}(x,v)_{kl}\overleftarrow{R}_v(x,e_i)^k e^l_i = \sum_{i=1}^{m-1} g(x,-v)_{kl}R_{-v}(x,e_i)^k e^l_i = \mathrm{Ric}^{\mathscr{F}}(x,-v). \qquad \square$$

Finally, we need the behavior of the S-curvature under reversal. Since $-\mathscr{B}_x$ and $\mathscr{B}_x$ have the same volume[5], $\mathscr{F}$ and $\overleftarrow{\mathscr{F}}$ share the Busemann–Hausdorff measure, and we denote by $\overleftarrow{\tau}$, $\overleftarrow{S}$ and $\overleftarrow{\dot{S}}$ the distortion, the S-curvature and its derivative along geodesics of $\overleftarrow{\mathscr{F}}$, with respect to $\mathfrak{m}_{\mathrm{BH}}$.

Lemma 30 (S-curvature under reversal)**.** *For every $x \in \mathscr{M}$ and $v \in T_x\mathscr{M} \setminus \{0\}$, we have that $\overleftarrow{\tau}(x,v) = \tau(x,-v)$, $\overleftarrow{S}(x,v) = -S(x,-v)$ and $\overleftarrow{\dot{S}}(x,v) = \dot{S}(x,-v)$.*

Proof. The distortion only involves the fundamental tensor and the density σ_{BH}. Since $\overleftarrow{g}_{ij}(x,v) = g_{ij}(x,-v)$ (see the proof of Proposition 26) and σ_{BH} is shared by $\mathscr{F}$ and $\overleftarrow{\mathscr{F}}$, we get $\overleftarrow{\tau}(x,v) = \tau(x,-v)$.

Let γ be the geodesic of $\mathscr{F}$ with $\gamma(0) = x$ and $\dot{\gamma}(0) = -v$, defined on some interval $(-\delta,\delta)$, and let $\bar{\gamma}(t) = \gamma(-t)$, so that $\bar{\gamma}(0) = x$ and $\dot{\bar{\gamma}}(0) = v$. Using the geodesic equation of $\mathscr{F}$ and Proposition 26, we have that

$$\ddot{\bar{\gamma}}^k(t) = \ddot{\gamma}^k(-t) = -2G^k\big(\gamma(-t),\dot{\gamma}(-t)\big) = -2G^k\big(\bar{\gamma}(t),-\dot{\bar{\gamma}}(t)\big) = -2\overleftarrow{G}^k\big(\bar{\gamma}(t),\dot{\bar{\gamma}}(t)\big),$$

so that $\bar{\gamma}$ is the geodesic of $\overleftarrow{\mathscr{F}}$ with initial conditions (x,v). Hence, for $t \in (-\delta,\delta)$,

$$\overleftarrow{\tau}\big(\bar{\gamma}(t),\dot{\bar{\gamma}}(t)\big) = \tau\big(\bar{\gamma}(t),-\dot{\bar{\gamma}}(t)\big) = \tau\big(\gamma(-t),\dot{\gamma}(-t)\big).$$

Differentiating once and twice at $t = 0$ gives $\overleftarrow{S}(x,v) = -S(x,-v)$ and $\overleftarrow{\dot{S}}(x,v) = \dot{S}(x,-v)$. $\square$

[5]The unit ball of $\overleftarrow{\mathscr{F}}$ at x is $-\mathscr{B}_x$, and the linear map $v \mapsto -v$ has determinant ± 1, so that it preserves the Lebesgue measure λ_x. Hence $\lambda_x(-\mathscr{B}_x) = \lambda_x(\mathscr{B}_x)$ and, by Equation (11), $\mathscr{F}$ and $\overleftarrow{\mathscr{F}}$ have the same density σ_{BH}.

Lemma 31. *The moments of the kernel for the reverse Finsler metric are related to the moments of the kernel for the Finsler metric. In particular, for $n \in \mathbb{N}$ we have that:*

$$\overleftarrow{m}_n(x) = (-1)^n m_n(x), \qquad \overleftarrow{\Lambda}^k(x) = \Lambda^k(x),$$
$$\overleftarrow{s}_n(x) = (-1)^{n+1} s_n(x), \qquad \overleftarrow{\Upsilon}(x) = \Upsilon(x).$$

Proof. Since $\mathscr{F}$ and $\overleftarrow{\mathscr{F}}$ share the density σ_{BH}, which does not depend on v, it factors out of all the integrals below, and we omit it. First, we show the relation between $\overleftarrow{m}_n$ and m_n. By the change of variable $u = -v$,

$$\begin{aligned}
\overleftarrow{m}_n(x) &= \int_{T_x\mathscr{M}} K(\overleftarrow{\mathscr{F}}(x,v)) v^{\otimes n} \mathrm{d}\lambda_x(v) \\
&= \int_{T_x\mathscr{M}} K(\mathscr{F}(x,-u))(-u)^{\otimes n} \mathrm{d}\lambda_x(u) \\
&= (-1)^n \int_{T_x\mathscr{M}} K(\mathscr{F}(x,u)) u^{\otimes n} \mathrm{d}\lambda_x(u) \\
&= (-1)^n m_n(x).
\end{aligned}$$

The change of variable $u = -v$ has Jacobian determinant $\det(-I_m) = (-1)^m$, whose absolute value is 1, so that it preserves λ_x, and the sign factor $(-1)^n$ comes only from the $v^{\otimes n}$ term. This argument is also used below for the other terms.

It remains to treat the $\overleftarrow{\Lambda}^k(x)$, $\overleftarrow{s}_n(x)$ and $\overleftarrow{\Upsilon}(x)$ terms. For $\overleftarrow{\Lambda}^k(x)$, recall that $\overleftarrow{\tilde{\Gamma}}^k_{ij}(v) v^i v^j = 2\overleftarrow{G}^k(x,v)$ and $\tilde{\Gamma}^k_{ij}(x,v) v^i v^j = 2G^k(x,v)$. Using $\overleftarrow{G}^k(x,v) = G^k(x,-v)$ (Proposition 26),

$$\begin{aligned}
\overleftarrow{\Lambda}^k(x) &= 2\int_{T_x\mathscr{M}} K(\overleftarrow{\mathscr{F}}(x,v)) \overleftarrow{G}^k(x,v) \mathrm{d}\lambda_x(v) \\
&= 2\int_{T_x\mathscr{M}} K(\mathscr{F}(x,-v)) G^k(x,-v) \mathrm{d}\lambda_x(v) \\
&= 2\int_{T_x\mathscr{M}} K(\mathscr{F}(x,u)) G^k(x,u) \mathrm{d}\lambda_x(u) \qquad \text{by change of variable } u = -v \\
&= \Lambda^k(x).
\end{aligned}$$

For the S-curvature terms, using Lemma 30 and the same change of variable,

$$\begin{aligned}
\overleftarrow{s}_n(x) &= \int_{T_x\mathscr{M}} K(\overleftarrow{\mathscr{F}}(x,v)) \overleftarrow{S}(x,v) v^{\otimes n} \mathrm{d}\lambda_x(v) = -\int_{T_x\mathscr{M}} K(\mathscr{F}(x,-v)) S(x,-v) v^{\otimes n} \mathrm{d}\lambda_x(v) \\
&= (-1)^{n+1} \int_{T_x\mathscr{M}} K(\mathscr{F}(x,u)) S(x,u) u^{\otimes n} \mathrm{d}\lambda_x(u) = (-1)^{n+1} s_n(x).
\end{aligned}$$

Finally, for $\overleftarrow{\Upsilon}(x)$, Lemma 30 gives $\overleftarrow{S}(x,v)^2 - \overleftarrow{\dot{S}}(x,v) = S(x,-v)^2 - \dot{S}(x,-v)$, and Proposition 29 gives $\mathrm{Ric}^{\overleftarrow{\mathscr{F}}}(x,v) = \mathrm{Ric}^{\mathscr{F}}(x,-v)$, so that, by the change of variable $u=-v$,

$$\begin{aligned}
\overleftarrow{\Upsilon}(x) &= \int_{T_x\mathscr{M}} K(\overleftarrow{\mathscr{F}}(x,v)) \left(\overleftarrow{S}(x,v)^2 - \overleftarrow{\dot{S}}(x,v) - \frac{1}{3}\mathrm{Ric}^{\overleftarrow{\mathscr{F}}}(x,v) \right) \mathrm{d}\lambda_x(v) \\
&= \int_{T_x\mathscr{M}} K(\mathscr{F}(x,-v)) \left(S(x,-v)^2 - \dot{S}(x,-v) - \frac{1}{3}\mathrm{Ric}^{\mathscr{F}}(x,-v) \right) \mathrm{d}\lambda_x(v) \\
&= \int_{T_x\mathscr{M}} K(\mathscr{F}(x,u)) \left(S(x,u)^2 - \dot{S}(x,u) - \frac{1}{3}\mathrm{Ric}^{\mathscr{F}}(x,u) \right) \mathrm{d}\lambda_x(u) \\
&= \Upsilon(x).
\end{aligned}$$

Putting everything together, we have shown that $\overleftarrow{m}_n(x) = (-1)^n m_n(x)$, $\overleftarrow{\Lambda}^k(x) = \Lambda^k(x)$, $\overleftarrow{s}_n(x) = (-1)^{n+1} s_n(x)$ and $\overleftarrow{\Upsilon}(x) = \Upsilon(x)$, which concludes the proof. □

Recap. The goal of this subsection was to study the behavior of the reverse Finsler metric $\overleftarrow{\mathscr{F}}$ and its associated transport operator $\mathscr{G}_{\overleftarrow{\mathscr{F}}}$. We have shown that the moments of the kernel for the reverse Finsler metric are related to the moments of the kernel for the Finsler metric, with $\overleftarrow{m}_n(x) = (-1)^n m_n(x)$, $\overleftarrow{\Lambda}^k(x) = \Lambda^k(x)$, $\overleftarrow{s}_n(x) = (-1)^{n+1} s_n(x)$ and $\overleftarrow{\Upsilon}(x) = \Upsilon(x)$. This allows us to derive the moment expansion of the transport operator $\mathscr{G}_{\overleftarrow{\mathscr{F}}}$ in terms of the moments of the kernel for $\mathscr{F}$. In particular, we have the following proposition.

Proposition 32. *The transport operator $\mathscr{G}_{\overleftarrow{\mathscr{F}}}$ associated with the reverse Finsler metric $\overleftarrow{\mathscr{F}}$ has the following moment expansion:*

$$\mathscr{G}_{\overleftarrow{\mathscr{F}}}[f] = m_0 f - \varepsilon \left(m_1^i \partial_i f - s_0 f \right) + \frac{\varepsilon^2}{2} \left(m_2^{ij} \partial_i \partial_j f - \left(\Lambda^k + 2 s_1^k \right) \partial_k f + \Upsilon f \right) + O(\varepsilon^3).$$

This follows from Theorem 1 applied to $\overleftarrow{\mathscr{F}}$, which shares the density σ_{BH} with $\mathscr{F}$, together with Lemma 31. Compared with the expansion of $\mathscr{G}_{\mathscr{F}}$, the terms of odd order in ε change sign under reversal, while the terms of even order do not.

Proof of the moment expansion of the symmetric and anti-symmetric transport operators

Using Proposition 32, the moment expansions of the symmetric and antisymmetric parts $\mathscr{G}^s_{\mathscr{F}} = (\mathscr{G}_{\mathscr{F}} + \mathscr{G}_{\overleftarrow{\mathscr{F}}})/2$ and $\mathscr{G}^a_{\mathscr{F}} = (\mathscr{G}_{\mathscr{F}} - \mathscr{G}_{\overleftarrow{\mathscr{F}}})/2$ of the transport operator follow directly: up to order ε^2, the symmetric part retains the even-order terms and the antisymmetric part the odd-order ones. This yields

$$\begin{aligned}
\mathscr{G}^s_{\mathscr{F}}[f] &= m_0 f + \frac{\varepsilon^2}{2} \left(m_2^{ij} \partial_i \partial_j f - \left(\Lambda^k + 2 s_1^k \right) \partial_k f + \Upsilon f \right) + O(\varepsilon^3) \quad \text{and} \\
\mathscr{G}^a_{\mathscr{F}}[f] &= \varepsilon \left(m_1^i \partial_i f - s_0 f \right) + O(\varepsilon^3).
\end{aligned}$$

It is important to note that, in practice, one does not observe the transport operator $\mathscr{G}_{\mathscr{F}}[f]$ directly, as points are rarely sampled uniformly from the manifold. Instead, one typically observes $\mathscr{G}_{\mathscr{F}}[\rho f]$ applied to the density ρ of the sampling distribution. Looking at the expansion of $\mathscr{G}_{\mathscr{F}}[\rho f]$, we have:

$$\begin{aligned}
\mathscr{G}_{\mathscr{F}}[\rho f] = m_0 \rho f &+ \varepsilon \left(m_1^i \partial_i (\rho f) - s_0 \rho f \right) \\
&+ \frac{\varepsilon^2}{2} \left(m_2^{ij} \partial_i \partial_j (\rho f) - \left(\Lambda^k + 2 s_1^k \right) \partial_k (\rho f) + \Upsilon \rho f \right) + O(\varepsilon^3)
\end{aligned}$$

so that every term is 'polluted' by the density ρ. This calls for a normalization, which we now introduce.

θ-NORMALIZED TRANSPORT OPERATOR

Our goal is now to define a θ-normalized transport operator $\mathcal{G}_{\mathcal{F}}^{(\theta)}$ that mitigates the impact of the sampling density ρ on the limit operator. We define $\mathcal{G}_{\mathcal{F}}^{(\theta)}$ as:

$$\mathcal{G}_{\mathcal{F}}^{(\theta)}[f](x) = \frac{1}{q_\varepsilon(x)^\theta}\mathcal{G}_{\mathcal{F}}\left[\frac{f}{{q_\varepsilon}^\theta}\right](x),$$

where $q_\varepsilon(x)$ is a normalization factor. Specifically, we take $q_\varepsilon = (d'+d)/2$. This choice allows us to clean up the expansion of $\mathcal{G}_{\mathcal{F}}^{(\theta)}[\rho f]$ and analyze its behavior as $\varepsilon \to 0$. Since q_ε is invariant under reversal, the normalization $q_\varepsilon(x)^\theta q_\varepsilon(y)^\theta$ is symmetric in (x,y), and the symmetric and antisymmetric parts of $\mathcal{G}_{\mathcal{F}}^{(\theta)}$ are obtained from those of $\mathcal{G}_{\mathcal{F}}$ as

$$\mathcal{G}_{\mathcal{F}}^{(\theta,s)}[f] = \frac{1}{{q_\varepsilon}^\theta}\mathcal{G}_{\mathcal{F}}^{s}\left[\frac{f}{{q_\varepsilon}^\theta}\right] \quad \text{and} \quad \mathcal{G}_{\mathcal{F}}^{(\theta,a)}[f] = \frac{1}{{q_\varepsilon}^\theta}\mathcal{G}_{\mathcal{F}}^{a}\left[\frac{f}{{q_\varepsilon}^\theta}\right].$$

We use q_ε rather than d' or d alone because its expansion has no term of order ε, which makes it a cleaner proxy for the sampling density ρ. Indeed, the expansions of d' and d read

$$\begin{aligned} d = \mathcal{G}_{\mathcal{F}}[\rho] &= m_0\rho + \varepsilon\left(m_1^i\partial_i\rho - s_0\rho\right) \\ &\quad + \frac{\varepsilon^2}{2}\left(m_2^{ij}\partial_i\partial_j\rho - \left(\Lambda^k + 2s_1^k\right)\partial_k\rho + \Upsilon\rho\right) + O(\varepsilon^3), \\ d' = \mathcal{G}_{\overleftarrow{\mathcal{F}}}[\rho] &= m_0\rho - \varepsilon\left(m_1^i\partial_i\rho - s_0\rho\right) \\ &\quad + \frac{\varepsilon^2}{2}\left(m_2^{ij}\partial_i\partial_j\rho - \left(\Lambda^k + 2s_1^k\right)\partial_k\rho + \Upsilon\rho\right) + O(\varepsilon^3). \end{aligned}$$

Hence $q_\varepsilon = m_0\rho + O(\varepsilon^2)$, whereas d' and d each carry a term of order ε. Moreover, $m_0 = m\mu_0\mathrm{Vol}_{\mathrm{Euc}}(\mathbb{B}^m)$ does not depend on x (Section D.4), so that q_ε is proportional to ρ up to $O(\varepsilon^2)$. We now define $\phi^{(\theta)}(x) = \rho(x)/q_\varepsilon(x)^\theta$, and look at the expansion of the observable quantity $\mathcal{G}_{\mathcal{F}}^{(\theta)}[\rho f]$:

$$\begin{aligned} \mathcal{G}_{\mathcal{F}}^{(\theta)}[\rho f] &= \frac{1}{{q_\varepsilon}^\theta}\mathcal{G}_{\mathcal{F}}[\phi^{(\theta)} f] \\ &= \frac{1}{{q_\varepsilon}^\theta}\Big[m_0\phi^{(\theta)} f + \varepsilon\left(m_1^i\partial_i(\phi^{(\theta)} f) - s_0\phi^{(\theta)} f\right) \\ &\quad + \frac{\varepsilon^2}{2}\left(m_2^{ij}\partial_i\partial_j(\phi^{(\theta)} f) - \left(\Lambda^k + 2s_1^k\right)\partial_k(\phi^{(\theta)} f) + \Upsilon\phi^{(\theta)} f\right) + O(\varepsilon^3)\Big]. \end{aligned}$$

Expanding the derivatives with the Leibniz rule gives

$$\begin{aligned} \partial_i(\phi^{(\theta)} f) &= \phi^{(\theta)}\partial_i f + f\partial_i\phi^{(\theta)}, \\ m_2^{ij}\partial_i\partial_j(\phi^{(\theta)} f) &= m_2^{ij}\phi^{(\theta)}\partial_i\partial_j f + 2m_2^{ij}\partial_i\phi^{(\theta)}\partial_j f + m_2^{ij} f\partial_i\partial_j\phi^{(\theta)}, \end{aligned}$$

where we used the fact that $m_2^{ij} = m_2^{ji}$. Plugging these into the expansion of $\mathcal{G}_{\mathcal{F}}^{(\theta)}[\rho f]$, we obtain:

$$\begin{aligned} \mathcal{G}_{\mathcal{F}}^{(\theta)}[\rho f] = \frac{1}{{q_\varepsilon}^\theta}\Bigg[&m_0\phi^{(\theta)} f + \varepsilon\left(m_1^i\left(\phi^{(\theta)}\partial_i f + f\partial_i\phi^{(\theta)}\right) - s_0\phi^{(\theta)} f\right) \\ &+ \frac{\varepsilon^2}{2}\Big(m_2^{ij}(\phi^{(\theta)}\partial_i\partial_j f + 2\partial_i\phi^{(\theta)}\partial_j f + f\partial_i\partial_j\phi^{(\theta)}) - \left(\Lambda^k + 2s_1^k\right)\left(\phi^{(\theta)}\partial_k f + f\partial_k\phi^{(\theta)}\right) \\ &+ \Upsilon\phi^{(\theta)} f\Big) + O(\varepsilon^3)\Bigg]. \end{aligned} \tag{13}$$

We then define the symmetric and antisymmetric normalized transport operators $\mathscr{P}^{(\theta,s)}$ and $\mathscr{P}^{(\theta,a)}$ as

$$\mathscr{P}^{(\theta,\bullet)}[f] = \frac{1}{d^{(\theta,s)}}\left(\mathscr{G}^{(\theta,\bullet)}[\rho f] - f\mathscr{G}^{(\theta,\bullet)}[\rho]\right) \tag{14}$$

for $\bullet \in \{s,a\}$, where $d^{(\theta,s)} = \mathscr{G}^{(\theta,s)}_{\mathscr{F}}[\rho]$ is the symmetric normalized degree defined below. This expression centers the operators $\mathscr{G}$ around their respective degrees, so that $\mathscr{P}$ vanishes on constant functions.

Expansion for the degrees of the θ-normalized operator. The in- and out-degrees $d'^{(\theta)}$ and $d^{(\theta)}$ of the θ-normalized kernel expand as

$$d^{(\theta)} = \mathscr{G}^{(\theta)}_{\mathscr{F}}[\rho] = \frac{1}{{q_\varepsilon}^\theta}\Big(m_0\phi^{(\theta)} + \varepsilon\left(m_1^i\partial_i\phi^{(\theta)} - s_0\phi^{(\theta)}\right) + \frac{\varepsilon^2}{2}\left(m_2^{ij}\partial_i\partial_j\phi^{(\theta)} - \left(\Lambda^k + 2s_1^k\right)\partial_k\phi^{(\theta)} + \Upsilon\phi^{(\theta)}\right)\Big) + O(\varepsilon^3),$$

and $d'^{(\theta)} = \mathscr{G}^{(\theta)}_{\overleftarrow{\mathscr{F}}}[\rho]$ has the same expansion, with the opposite sign in front of the term of order ε. By definition, the degrees of the symmetric and antisymmetric normalized kernels are

$$d^{(\theta,s)} = \mathscr{G}^{(\theta,s)}_{\mathscr{F}}[\rho] = \frac{1}{2}\left(\mathscr{G}^{(\theta)}_{\mathscr{F}}[\rho] + \mathscr{G}^{(\theta)}_{\overleftarrow{\mathscr{F}}}[\rho]\right) = \frac{d^{(\theta)} + d'^{(\theta)}}{2},$$
$$d^{(\theta,a)} = \mathscr{G}^{(\theta,a)}_{\mathscr{F}}[\rho] = \frac{1}{2}\left(\mathscr{G}^{(\theta)}_{\mathscr{F}}[\rho] - \mathscr{G}^{(\theta)}_{\overleftarrow{\mathscr{F}}}[\rho]\right) = \frac{d^{(\theta)} - d'^{(\theta)}}{2}.$$

Plugging the expansions of $d'^{(\theta)}$ and $d^{(\theta)}$ into the above expressions, we have that:

$$d^{(\theta,s)} = \frac{1}{{q_\varepsilon}^\theta}\left(m_0\phi^{(\theta)} + \frac{\varepsilon^2}{2}\left(m_2^{ij}\partial_i\partial_j\phi^{(\theta)} - \left(\Lambda^k + 2s_1^k\right)\partial_k\phi^{(\theta)} + \Upsilon\phi^{(\theta)}\right)\right) + O(\varepsilon^3), \tag{15}$$
$$d^{(\theta,a)} = \frac{1}{{q_\varepsilon}^\theta}\varepsilon\left(m_1^i\partial_i\phi^{(\theta)} - s_0\phi^{(\theta)}\right) + O(\varepsilon^3). \tag{16}$$

Using the fact that $1/(1+x) = 1 - x + O(x^2)$, and noting that the bracket of order ε^2 in Equation (15) is $\mathscr{A}_2[\phi^{(\theta)}]$ (Definition 21), we can write the following expansion for $1/d^{(\theta,s)}$:

$$\begin{aligned}\frac{1}{d^{(\theta,s)}} &= \left(\frac{m_0\phi^{(\theta)}}{{q_\varepsilon}^\theta}\right)^{-1}\left(1 + \frac{\varepsilon^2}{2m_0\phi^{(\theta)}}\mathscr{A}_2[\phi^{(\theta)}] + O(\varepsilon^3)\right)^{-1} \\ &= \frac{{q_\varepsilon}^\theta}{m_0\phi^{(\theta)}}\left(1 - \frac{\varepsilon^2}{2m_0\phi^{(\theta)}}\mathscr{A}_2[\phi^{(\theta)}] + O(\varepsilon^3)\right) \\ &= \frac{{q_\varepsilon}^\theta}{m_0\phi^{(\theta)}} + O(\varepsilon^2)\end{aligned} \tag{17}$$

We now expand the symmetric and antisymmetric normalized transport operators $\mathscr{P}^{(\theta,s)}$ and $\mathscr{P}^{(\theta,a)}$ of Equation (14).

Expansion for the symmetric part. By putting together Equation (13) and Equation (15), we have that:

$$\begin{aligned}\mathcal{G}_{\mathcal{F}}^{(\theta,s)}[\rho f] - f d^{(\theta,s)} &= \frac{\varepsilon^2}{2{q_\varepsilon}^\theta}\Bigg(m_2^{ij}\left(\phi^{(\theta)}\partial_i\partial_j f + 2\partial_i\phi^{(\theta)}\partial_j f + f\partial_i\partial_j\phi^{(\theta)} - f\partial_i\partial_j\phi^{(\theta)}\right)\\ &\quad - \left(\Lambda^k + 2s_1^k\right)\left(\phi^{(\theta)}\partial_k f + f\partial_k\phi^{(\theta)} - f\partial_k\phi^{(\theta)}\right)\\ &\quad + \Upsilon\left(\phi^{(\theta)} f - \phi^{(\theta)} f\right)\Bigg) + O(\varepsilon^3)\\ &= \frac{\varepsilon^2}{2{q_\varepsilon}^\theta}\left(m_2^{ij}\left(\phi^{(\theta)}\partial_i\partial_j f + 2\partial_i\phi^{(\theta)}\partial_j f\right) - \phi^{(\theta)}\left(\Lambda^k + 2s_1^k\right)\partial_k f\right) + O(\varepsilon^3).\end{aligned}$$

Hence, by Equation (17), we have that:

$$\begin{aligned}\mathcal{P}^{(\theta,s)}[f] &= \frac{1}{d^{(\theta,s)}}\left(\mathcal{G}_{\mathcal{F}}^{(\theta,s)}[\rho f] - f d^{(\theta,s)}\right)\\ &= \frac{{q_\varepsilon}^\theta}{m_0\phi^{(\theta)}}\frac{\varepsilon^2}{2{q_\varepsilon}^\theta}\left(m_2^{ij}\left(\phi^{(\theta)}\partial_i\partial_j f + 2\partial_i\phi^{(\theta)}\partial_j f\right) - \phi^{(\theta)}\left(\Lambda^k + 2s_1^k\right)\partial_k f\right) + O(\varepsilon^3)\\ &= \frac{\varepsilon^2}{2m_0}\left(m_2^{ij}\left(\partial_i\partial_j f + 2\partial_i\log\phi^{(\theta)}\partial_j f\right) - \left(\Lambda^k + 2s_1^k\right)\partial_k f\right) + O(\varepsilon^3).\end{aligned}$$

where we used $\partial_i \log\phi^{(\theta)} = \partial_i\phi^{(\theta)}/\phi^{(\theta)}$. Note that the curvature term Υ, which only multiplies f, is removed by the centering.

Expansion for the anti-symmetric part. By putting together Equation (13) and Equation (16), we have that:

$$\begin{aligned}\mathcal{G}_{\mathcal{F}}^{(\theta,a)}[\rho f] - f d^{(\theta,a)} &= \frac{\varepsilon}{{q_\varepsilon}^\theta}\left(m_1^i\left(\phi^{(\theta)}\partial_i f + f\partial_i\phi^{(\theta)}\right) - s_0\phi^{(\theta)} f - f\left(m_1^i\partial_i\phi^{(\theta)} - s_0\phi^{(\theta)}\right)\right) + O(\varepsilon^3)\\ &= \frac{\varepsilon}{{q_\varepsilon}^\theta} m_1^i\phi^{(\theta)}\partial_i f + O(\varepsilon^3).\end{aligned}$$

In particular, the term s_0 brought by the S-curvature only multiplies f, and is removed by the centering. Hence, by Equation (17), we have that:

$$\begin{aligned}\mathcal{P}^{(\theta,a)}[f] &= \frac{1}{d^{(\theta,s)}}\left(\mathcal{G}_{\mathcal{F}}^{(\theta,a)}[\rho f] - f d^{(\theta,a)}\right)\\ &= \frac{{q_\varepsilon}^\theta}{m_0\phi^{(\theta)}}\frac{\varepsilon}{{q_\varepsilon}^\theta} m_1^i\phi^{(\theta)}\partial_i f + O(\varepsilon^3)\\ &= \frac{\varepsilon}{m_0} m_1^i\partial_i f + O(\varepsilon^3).\end{aligned}$$

Consequently, if we introduce the normalized quantities $\tilde{m}_n = m_n/m_0$, $\tilde{\Lambda}^k = \Lambda^k/m_0$ and $\tilde{s}_1^k = s_1^k/m_0$, we have that:

$$\begin{aligned}\mathcal{P}^{(\theta,s)}[f] &= \frac{\varepsilon^2}{2}\left(\tilde{m}_2^{ij}\left(\partial_i\partial_j f + 2\partial_i\log\phi^{(\theta)}\partial_j f\right) - \left(\tilde{\Lambda}^k + 2\tilde{s}_1^k\right)\partial_k f\right) + O(\varepsilon^3),\\ \mathcal{P}^{(\theta,a)}[f] &= \varepsilon\tilde{m}_1^i\partial_i f + O(\varepsilon^3).\end{aligned}$$

Impact of the normalization parameter. The antisymmetric normalized transport operator $\mathscr{P}^{(\theta,a)}$ depends neither on the sampling density ρ nor on the normalization parameter θ. We thus focus on the impact of the normalization parameter θ on the symmetric normalized transport operator $\mathscr{P}^{(\theta,s)}$. This normalization appears through the term $\partial_i \log \phi^{(\theta)}$, by definition of $\phi^{(\theta)} = \rho/{q_\varepsilon}^\theta$. We have that:

$$\partial_i \log \phi^{(\theta)} = \partial_i \log \left(\frac{\rho}{{q_\varepsilon}^\theta} \right) = \partial_i \log \rho - \theta \partial_i \log q_\varepsilon.$$

Moreover, we have seen that $q_\varepsilon = m_0 \rho + O(\varepsilon^2)$, where $m_0 = m\mu_0 \mathrm{Vol}_{\mathrm{Euc}}(\mathbb{B}^m)$ does not depend on x (Section D.4), so that $\partial_i \log q_\varepsilon = \partial_i \log \rho + O(\varepsilon^2)$. Hence, we have that:

$$\partial_i \log \phi^{(\theta)} = (1-\theta)\partial_i \log \rho + O(\varepsilon^2).$$

The leading-order term in the expansion of $\mathscr{P}^{(\theta,s)}$ is thus

$$\mathscr{P}^{(\theta,s)}[f] = \frac{\varepsilon^2}{2} \left(\tilde{m}_2^{ij} \left(\partial_i \partial_j f + 2(1-\theta) \partial_i \log \rho \partial_j f \right) - \left(\tilde{\Lambda}^k + 2\tilde{s}_1^k \right) \partial_k f \right) + O(\varepsilon^3).$$

In particular, for $\theta = 1$ the sampling density disappears from the limit, exactly as in the Riemannian setting (Coifman & Lafon, 2006). Taking the limits for both operators when $\varepsilon \to 0$ yields $\mathscr{L}^a f = \tilde{m}_1^i \partial_i f$, and

$$\mathscr{L}^s f = \frac{1}{2} \tilde{m}_2^{ij} \partial_i \partial_j f + (1-\theta) \tilde{m}_2^{ij} \partial_i \log \rho \partial_j f - \frac{1}{2} \left(\tilde{\Lambda}^k + 2\tilde{s}_1^k \right) \partial_k f. \tag{18}$$

Divergence form of the symmetric limit. It remains to write Equation (18) in divergence form. This relies on the following identities between the coefficients of Theorem 1.

Lemma 33 (Divergence identities). *For every $x \in \mathscr{M}$, the kernel moments of Definition 21 satisfy*

$$s_0 = -\frac{1}{2} \operatorname{div}_{BH}(m_1) \qquad \text{and} \qquad \Lambda^k + 2s_1^k = -\frac{1}{\sigma_{\mathrm{BH}}} \partial_j \left(\sigma_{\mathrm{BH}} m_2^{jk} \right),$$

where $\operatorname{div}_{BH} X = \sigma_{\mathrm{BH}}^{-1} \partial_i (\sigma_{\mathrm{BH}} X^i)$. Equivalently, the operators of Theorem 1 read

$$\mathscr{A}_1 f = m_1^i \partial_i f + \frac{1}{2} \operatorname{div}_{BH}(m_1) f, \qquad \mathscr{A}_2 f = \frac{1}{\sigma_{\mathrm{BH}}} \partial_j \left(\sigma_{\mathrm{BH}} m_2^{jk} \partial_k f \right) + \Upsilon f,$$

so that $\mathscr{A}_1$ is skew-adjoint and $\mathscr{A}_2$ is self-adjoint on $L^2(\mathscr{M}, \mathfrak{m}_{\mathrm{BH}})$.

Proof. We work in a chart, write $\sigma = \sigma_{\mathrm{BH}}$, and $M_n(x) = \int_{T_x\mathscr{M}} K(\mathscr{F}(x,v)) v^{\otimes n} \mathrm{d}\lambda_x(v)$, so that $m_n = \sigma M_n$. The proof relies on two classical facts about the geodesic spray coefficients G^k of $\mathscr{F}$ (Proposition 26), whose geodesics satisfy $\ddot{\gamma}^k = -2G^k(\gamma, \dot{\gamma})$. First, the S-curvature reads (Shen, 2001)

$$S(x,v) = \frac{\partial G^l}{\partial v^l}(x,v) - v^l \partial_l \log \sigma(x). \tag{19}$$

Second, $\mathscr{F}$ is constant along geodesics, so that differentiating $t \mapsto \mathscr{F}(\gamma(t), \dot{\gamma}(t))$ at $t = 0$ gives

$$v^j \frac{\partial \mathscr{F}}{\partial x^j}(x,v) = 2G^l(x,v) \frac{\partial \mathscr{F}}{\partial v^l}(x,v). \tag{20}$$

Let P be a polynomial in v. An integration by parts in v, followed by Equation (20), gives

$$\begin{aligned}
\int_{T_x\mathscr{M}} K(\mathscr{F}) \frac{\partial G^l}{\partial v^l} P \mathrm{d}\lambda_x(v) &= -\int_{T_x\mathscr{M}} K'(\mathscr{F}) \frac{\partial \mathscr{F}}{\partial v^l} G^l P \mathrm{d}\lambda_x(v) - \int_{T_x\mathscr{M}} K(\mathscr{F}) G^l \frac{\partial P}{\partial v^l} \mathrm{d}\lambda_x(v) \\
&= -\frac{1}{2} \int_{T_x\mathscr{M}} K'(\mathscr{F}) v^j \frac{\partial \mathscr{F}}{\partial x^j} P \mathrm{d}\lambda_x(v) - \int_{T_x\mathscr{M}} K(\mathscr{F}) G^l \frac{\partial P}{\partial v^l} \mathrm{d}\lambda_x(v) \\
&= -\frac{1}{2} \partial_j \int_{T_x\mathscr{M}} K(\mathscr{F}) v^j P \mathrm{d}\lambda_x(v) - \int_{T_x\mathscr{M}} K(\mathscr{F}) G^l \frac{\partial P}{\partial v^l} \mathrm{d}\lambda_x(v).
\end{aligned}$$

The boundary terms vanish, and the derivative can be taken out of the integral, since $K(\mathscr{F}(x,v))$ and $K'(\mathscr{F}(x,v))$ decay exponentially in v by (A1), while G^l, its derivatives and P grow polynomially. Moreover, in a chart, λ_x is the Lebesgue measure of $\mathbb{R}^m$, which does not depend on x.

First identity. Taking $P=1$, we get $\int K(\mathscr{F})\partial_{v^l}G^l\mathrm{d}\lambda_x = -\frac{1}{2}\partial_j M_1^j$. Hence, by Equation (19),

$$s_0 = \sigma\int_{T_x\mathscr{M}} K(\mathscr{F})S\mathrm{d}\lambda_x(v) = -\frac{\sigma}{2}\partial_j M_1^j - \partial_j\sigma M_1^j = -\frac{1}{2\sigma}\partial_j\left(\sigma^2 M_1^j\right) = -\frac{1}{2}\operatorname{div}_{BH}(m_1).$$

Second identity. Taking $P=v^k$, we get $\int K(\mathscr{F})\partial_{v^l}G^l v^k\mathrm{d}\lambda_x = -\frac{1}{2}\partial_j M_2^{jk} - \int K(\mathscr{F})G^k\mathrm{d}\lambda_x$. Moreover, $\tilde{\Gamma}_{ij}^k(x,v)v^iv^j = 2G^k(x,v)$, so that $\Lambda^k = 2\sigma\int K(\mathscr{F})G^k\mathrm{d}\lambda_x$. Hence, by Equation (19),

$$\begin{aligned}\Lambda^k + 2s_1^k &= 2\sigma\int_{T_x\mathscr{M}} K(\mathscr{F})\left(G^k + \frac{\partial G^l}{\partial v^l}v^k\right)\mathrm{d}\lambda_x(v) - 2\partial_j\sigma M_2^{jk}\\ &= -\sigma\partial_j M_2^{jk} - 2\partial_j\sigma M_2^{jk} = -\frac{1}{\sigma}\partial_j\left(\sigma^2 M_2^{jk}\right) = -\frac{1}{\sigma}\partial_j\left(\sigma m_2^{jk}\right).\end{aligned}$$

The expressions of $\mathscr{A}_1$ and $\mathscr{A}_2$ follow by substitution in Definition 21. Finally, for smooth f and h, Stokes' theorem (Lee, 2012, Theorem 16.11) on the compact manifold without boundary $\mathscr{M}$ gives $\int_{\mathscr{M}}\operatorname{div}_{BH}(fhm_1)\mathrm{d}\mathfrak{m}_{\mathrm{BH}} = 0$, that is

$$\int_{\mathscr{M}}\mathscr{A}_1[f]h\mathrm{d}\mathfrak{m}_{\mathrm{BH}} = -\int_{\mathscr{M}} f\mathscr{A}_1[h]\mathrm{d}\mathfrak{m}_{\mathrm{BH}},$$

and similarly $\int_{\mathscr{M}}\operatorname{div}_{BH}(hm_2\nabla f - fm_2\nabla h)\mathrm{d}\mathfrak{m}_{\mathrm{BH}} = 0$ gives the self-adjointness of $\mathscr{A}_2$. □

Remark 34. *Lemma 33 is the infinitesimal counterpart of the adjointness relation $\mathscr{G}'_{\mathscr{F}} = \mathscr{G}^*_{\mathscr{F}}$ (Proposition 19). Indeed, plugging the expansions of Theorem 1 and Proposition 32 into $\int_{\mathscr{M}}\mathscr{G}_{\mathscr{F}}[f]h\mathrm{d}\mathfrak{m}_{\mathrm{BH}} = \int_{\mathscr{M}} f\mathscr{G}'_{\mathscr{F}}[h]\mathrm{d}\mathfrak{m}_{\mathrm{BH}}$, the terms of order ε and ε^2 show that $\mathscr{A}_1$ is skew-adjoint and $\mathscr{A}_2$ is self-adjoint on $L^2(\mathscr{M},\mathfrak{m}_{\mathrm{BH}})$, which gives back the two identities after an integration by parts.*

Dividing the second identity by the constant m_0, the second-order and first-order terms of Equation (18) combine into a divergence,

$$\frac{1}{2}\tilde{m}_2^{ij}\partial_i\partial_j f - \frac{1}{2}\left(\tilde{\Lambda}^k + 2\tilde{s}_1^k\right)\partial_k f = \frac{1}{2\sigma_{\mathrm{BH}}}\partial_j\left(\sigma_{\mathrm{BH}}\tilde{m}_2^{jk}\partial_k f\right) = \frac{1}{2}\operatorname{div}_{BH}\left(\tilde{m}_2\nabla f\right),$$

where $\tilde{m}_2\nabla f$ denotes the vector field $\tilde{m}_2^{jk}\partial_k f\partial_j$. Moreover, writing $\omega = \rho^{2(1-\theta)}$, the Leibniz rule gives $\frac{1}{2\omega}\operatorname{div}_{BH}(\omega X) = \frac{1}{2}\operatorname{div}_{BH}(X) + (1-\theta)X^j\partial_j\log\rho$ for any vector field X. Taking $X = \tilde{m}_2\nabla f$, and using $\tilde{m}_2^{jk} = cg_{\mathrm{BL}}^{jk}$ (Proposition 35), we get

$$\mathscr{L}^s f = \frac{1}{2\omega}\operatorname{div}_{BH}\left(\omega\tilde{m}_2\nabla f\right) = \frac{c_2}{\rho^{2(1-\theta)}}\operatorname{div}_{BH}\left(\rho^{2(1-\theta)}\nabla_{\mathrm{BL}} f\right).$$

Finally, since $\mathscr{M}$ is compact and without boundary (assumption (A2)), Stokes' theorem (Lee, 2012, Theorem 16.11) gives $\int_{\mathscr{M}}\operatorname{div}_{BH}(X)\mathrm{d}\mathfrak{m}_{\mathrm{BH}} = 0$ for any smooth vector field X. Applying it to $X = u\omega\nabla_{\mathrm{BL}} f$, for smooth $u, f : \mathscr{M}\to\mathbb{R}$, gives

$$\int_{\mathscr{M}} u\mathscr{L}^s[f]\omega\mathrm{d}\mathfrak{m}_{\mathrm{BH}} = -c_2\int_{\mathscr{M}}\langle\nabla u,\nabla f\rangle_{\mathrm{BL}}\omega\mathrm{d}\mathfrak{m}_{\mathrm{BH}},$$

which is symmetric in u and f: $\mathscr{L}^s$ is self-adjoint on $L^2(\mathscr{M},\omega\mathfrak{m}_{\mathrm{BH}})$. Its principal part is $c_2 g_{\mathrm{BL}}^{ij}\partial_i\partial_j$, with g_{BL}^{ij} positive definite, so that it is elliptic. This concludes the proof of Theorem 2.

D.4 Links between the moments of the kernel and the unit ball

We recall that the moments of the kernel are defined as (Definition 21):

$$m_n(x) = \sigma_{\mathrm{BH}}(x) \int_{T_x\mathscr{M}} K(\mathscr{F}(x,v)) v^{\otimes n} \mathrm{d}\lambda_x(v),$$

We denote the normalized moments of the kernel as $\tilde{m}_n = m_n/m_0$. We define the unit tangent ball and its indicatrix as:

$$\mathscr{B}_x = \{v \in T_x\mathscr{M} : \mathscr{F}(x,v) \leq 1\}, \quad \mathscr{I}_x = \{v \in T_x\mathscr{M} : \mathscr{F}(x,v) = 1\}.$$

We also introduce the volume, the centroid, and the raw second moment of the unit ball read

$$\lambda_x(\mathscr{B}_x) = \int_{\mathscr{B}_x} \mathrm{d}\lambda_x(v), \quad \mathbf{c}^i(x) = \frac{1}{\lambda_x(\mathscr{B}_x)} \int_{\mathscr{B}_x} v^i \mathrm{d}\lambda_x(v), \quad \text{and} \quad \mathbf{S}^{ij}(x) = \frac{1}{\lambda_x(\mathscr{B}_x)} \int_{\mathscr{B}_x} v^i v^j \mathrm{d}\lambda_x(v).$$

Proposition 35. *We can relate the moments of the kernel to the unit ball as follows:*

$$\begin{aligned} m_0(x) &= m\mu_0 \sigma_{\mathrm{BH}}(x) \lambda_x(\mathscr{B}_x) = m\mu_0 \mathrm{Vol}_{\mathrm{Euc}}(\mathbb{B}^m), \\ \tilde{m}_1^i(x) &= \frac{(m+1)\mu_1}{m\mu_0} \mathbf{c}^i(x), \\ \tilde{m}_2^{ij}(x) &= \frac{(m+2)\mu_2}{m\mu_0} \mathbf{S}^{ij}(x), \end{aligned}$$

where $\mu_n = \int_0^\infty K(r) r^{m+n-1} \mathrm{d}r$ are the raw moments of the radial kernel K.

Proof. Since $\mathscr{F}$ is positively homogeneous of degree 1 in the second variable, we use a polar change of variable $v = ru$ with $r \in \mathbb{R}_+$ and $u \in \mathscr{I}_x$. Under this change of variable, we have that $\mathrm{d}\lambda_x(v) = r^{m-1} \mathrm{d}r \mathrm{d}\omega_x$, where $\omega_x(A) = m\lambda_x(\{ru : u \in A, 0 \leq r \leq 1\})$ for $A \subset \mathscr{I}_x$. With this change of variable, the volume, centroids, and raw second moment of the unit ball as:

$$\lambda_x(\mathscr{B}_x) = \int_{\mathscr{I}_x} \int_0^1 r^{m-1} \mathrm{d}r \mathrm{d}\omega_x = \frac{1}{m} \int_{\mathscr{I}_x} \mathrm{d}\omega_x, \tag{21}$$

$$\mathbf{c}^i(x) = \frac{1}{\lambda_x(\mathscr{B}_x)} \int_{\mathscr{I}_x} \int_0^1 (ru)^i r^{m-1} \mathrm{d}r \mathrm{d}\omega_x = \frac{1}{(m+1)\lambda_x(\mathscr{B}_x)} \int_{\mathscr{I}_x} u^i \mathrm{d}\omega_x, \tag{22}$$

$$\mathbf{S}^{ij}(x) = \frac{1}{\lambda_x(\mathscr{B}_x)} \int_{\mathscr{I}_x} \int_0^1 (ru)^i (ru)^j r^{m-1} \mathrm{d}r \mathrm{d}\omega_x = \frac{1}{(m+2)\lambda_x(\mathscr{B}_x)} \int_{\mathscr{I}_x} u^i u^j \mathrm{d}\omega_x. \tag{23}$$

We now turn to the moments $m_n(x)$ of the kernel. Since $\sigma_{\mathrm{BH}}(x)$ does not depend on v, we compute them up to this factor. For the zeroth moment,

$$\begin{aligned} \frac{m_0(x)}{\sigma_{\mathrm{BH}}(x)} &= \int_{T_x\mathscr{M}} K(\mathscr{F}(x,v)) \mathrm{d}\lambda_x(v) \\ &= \int_{\mathscr{I}_x} \int_0^\infty K(r) r^{m-1} \mathrm{d}r \mathrm{d}\omega_x && \text{By the change of variable } v = ru \\ &= \mu_0 \int_{\mathscr{I}_x} \mathrm{d}\omega_x && \text{By Fubini's theorem} \\ &= m\mu_0 \lambda_x(\mathscr{B}_x) && \text{By Equation (21).} \end{aligned}$$

Similarly, the first moment is given by:

$$\begin{aligned}\frac{m_1^i(x)}{\sigma_{\mathrm{BH}}(x)} &= \int_{T_x\mathscr{M}} K(\mathscr{F}(x,v))v^i \mathrm{d}\lambda_x(v) \\ &= \int_{\mathscr{I}_x}\int_0^\infty K(r)(ru)^i r^{m-1}\mathrm{d}r\mathrm{d}\omega_x && \text{By the change of variable } v=ru \\ &= \mu_1 \int_{\mathscr{I}_x} u^i \mathrm{d}\omega_x && \text{By Fubini's theorem} \\ &= (m+1)\mu_1\lambda_x(\mathscr{B}_x)\mathbf{c}^i(x) && \text{By Equation (22).}\end{aligned}$$

The second moment is given by:

$$\begin{aligned}\frac{m_2^{ij}(x)}{\sigma_{\mathrm{BH}}(x)} &= \int_{T_x\mathscr{M}} K(\mathscr{F}(x,v))v^i v^j \mathrm{d}\lambda_x(v) \\ &= \int_{\mathscr{I}_x}\int_0^\infty K(r)(ru)^i(ru)^j r^{m-1}\mathrm{d}r\mathrm{d}\omega_x && \text{By the change of variable } v=ru \\ &= \mu_2 \int_{\mathscr{I}_x} u^i u^j \mathrm{d}\omega_x && \text{By Fubini's theorem} \\ &= (m+2)\mu_2\lambda_x(\mathscr{B}_x)\mathbf{S}^{ij}(x) && \text{By Equation (23).}\end{aligned}$$

Multiplying by $\sigma_{\mathrm{BH}}(x)$, Equation (11) gives $m_0(x) = m\mu_0\sigma_{\mathrm{BH}}(x)\lambda_x(\mathscr{B}_x) = m\mu_0\mathrm{Vol}_{\mathrm{Euc}}(\mathbb{B}^m)$. Dividing the expression of m_n by m_0, in which $\sigma_{\mathrm{BH}}(x)$ cancels, yields the desired result for the normalized moments $\tilde{m}_n = m_n/m_0$.

□

Implications for the limit operators

Zeroth moment. The zeroth moment $m_0(x)$ of the kernel appears in two ways: as the normalization factor in the definitions of $\tilde{m}_n$, $\tilde{\Lambda}^k$ and $\tilde{s}_1^k$, which is discussed at the end of this subsection, and as the leading coefficient of the expansion of Theorem 1. By Proposition 35, $m_0 = m\mu_0\mathrm{Vol}_{\mathrm{Euc}}(\mathbb{B}^m)$ does not depend on x: the Busemann–Hausdorff measure is precisely the one for which small forward balls have, at leading order, the volume of Euclidean balls. As a consequence, $q_\varepsilon = m_0\rho + O(\varepsilon^2)$ carries no geometric factor, and $\partial_i \log\phi^{(\theta)} = (1-\theta)\partial_i\log\rho + O(\varepsilon^2)$ (Section D.3): the θ-normalization acts exactly as in the Riemannian setting, and $\theta = 1$ removes the sampling density from the limit operator $\mathscr{L}^s$.

First moment. By Proposition 35, the normalized first moment $\tilde{m}_1^i(x)$ is proportional to the centroid $\mathbf{c}^i(x)$ of the unit ball $\mathscr{B}_x$. If the Finsler metric $\mathscr{F}$ is reversible, the unit ball $\mathscr{B}_x$ is symmetric with respect to the origin, so that $\mathbf{c}^i(x) = 0$. Hence, for reversible Finsler metrics, $\tilde{m}_1^i = 0$ and the anti-symmetric limit operator $\mathscr{L}^a$ vanishes[6]. When $\mathscr{F}$ is not reversible, the unit ball is in general not symmetric, and its centroid defines a vector field $\mathbf{c}$ on $\mathscr{M}$ which measures the asymmetry of $\mathscr{F}$ at first order. Note that $\mathbf{c}(x) = 0$ does not imply that $\mathscr{F}$ is reversible at x, since a non-symmetric unit ball can have its centroid at the origin. For Randers metrics, the centroid is given explicitly in terms of b by Proposition 43. This vector field is exactly what the anti-symmetric part of the graph Laplacian recovers: by Theorem 2,

$$\mathscr{L}^a f(x) = \tilde{m}_1^i(x)\partial_i f(x) = \frac{(m+1)\mu_1}{m\mu_0}\mathbf{c}^i(x)\partial_i f(x),$$

[6]One can note that in this context, the anti-symmetric operator $\mathscr{P}^{(\theta,a)}$ also vanishes: $\mathscr{P}^{(\theta,a)}[f] \equiv 0$ for any f. This is due to the fact that for a reversible Finsler metric, $\mathrm{dst}_{\mathscr{F}}(x,y) = \mathrm{dst}_{\mathscr{F}}(y,x)$ so that $W(x,y) = W(y,x)$.

so that $\mathscr{L}^a$ is the derivative along the centroid vector field, up to a constant that only depends on the kernel.

Second moment. By Proposition 35 and Equation (10), the normalized second moment is proportional to the raw second moment $\mathbf{S}^{ij}(x)$ of the unit ball, and hence to the Binet–Legendre metric g_{BL} (Definition 16):

$$\tilde{m}_2^{ij}(x) = \frac{(m+2)\mu_2}{m\mu_0}\mathbf{S}^{ij}(x) = \frac{\mu_2}{m\mu_0} g_{\mathrm{BL}}^{ij}(x).$$

Hence, up to the constant $2c_2 = \mu_2/(m\mu_0)$ of Theorem 2, the normalized second moment coincides with the Binet–Legendre metric, which is a Riemannian metric capturing the local geometry of $(\mathscr{M}, \mathscr{F})$. It depends on the kernel only through this constant, and does not depend on the bandwidth ε.

Normalization by m_0. Dividing by m_0 is what makes $\tilde{m}_1$ and $\tilde{m}_2$ the natural geometric objects governing the limit operators. Indeed, by Proposition 35, the raw moments m_1 and m_2 carry the factor $\sigma_{\mathrm{BH}}(x)\lambda_x(\mathscr{B}_x)$, as m_0 does, which cancels in the ratio. The normalized moments thus only involve the centroid $\mathbf{c}$ and the raw second moment $\mathbf{S}$, which are averages over the unit ball, and the kernel only enters through the constants μ_n/μ_0. This is also why the Binet–Legendre metric, which is itself normalized by $\lambda_x(\mathscr{B}_x)$ in Equation (10), appears in the limit.

D.5 Proof of Corollary 3

We start from the divergence form of Theorem 2. Let $\sqrt{\det g_{\mathrm{BL}}}$ denote the density, in a chart, of the Riemannian volume $\mathrm{vol}_{\mathrm{BL}}$ of g_{BL}, and let

$$w = \frac{\sigma_{\mathrm{BH}}}{\sqrt{\det g_{\mathrm{BL}}}} = \frac{\mathrm{Vol}_{\mathrm{Euc}}(\mathbb{B}^m)}{\mathrm{vol}_{\mathrm{BL}}(\mathscr{B}_x)}$$

be the density of $\mathfrak{m}_{\mathrm{BH}}$ with respect to $\mathrm{vol}_{\mathrm{BL}}$, where we used Equation (11), and $\mathrm{vol}_{\mathrm{BL}}(\mathscr{B}_x) = \sqrt{\det g_{\mathrm{BL}}(x)}\lambda_x(\mathscr{B}_x)$ is the volume of the unit ball of $\mathscr{F}$ for the Binet–Legendre metric. In particular, w does not depend on the chart. Since $\sigma_{\mathrm{BH}} = w\sqrt{\det g_{\mathrm{BL}}}$, we have $\mathrm{div}_{BH} X = \mathrm{div}_{\mathrm{BL}} X + X^i \partial_i \log w$ for any vector field X, where $\mathrm{div}_{\mathrm{BL}}$ is the Riemannian divergence of g_{BL}. Applying it in Theorem 2, and using $\mathrm{div}_{\mathrm{BL}}(\nabla_{\mathrm{BL}} f) = \Delta_{\mathrm{BL}} f$, we get, for any Finsler metric,

$$\mathscr{L}^s f = c_2\left(\Delta_{\mathrm{BL}} f + 2(1-\theta)\langle \nabla \log \rho, \nabla f\rangle_{\mathrm{BL}} + \langle \nabla \log w, \nabla f\rangle_{\mathrm{BL}}\right). \tag{24}$$

The discrepancy with a weighted Laplace–Beltrami operator of g_{BL} is thus a drift along the gradient of $\log w$, which measures the variations of the Binet–Legendre volume of the unit ball of $\mathscr{F}$ across $\mathscr{M}$.

Berwald metrics. When $\mathscr{F}$ is Berwald, the parallel transport along any curve from x to y is a linear map $P : T_x\mathscr{M} \to T_y\mathscr{M}$ which preserves $\mathscr{F}$ (Bao et al., 2012), so that $P(\mathscr{B}_x) = \mathscr{B}_y$. By Equation (10), the Binet–Legendre metric is built from the unit ball alone, and the change of variable $v = Pu$ gives $g_{\mathrm{BL}}^{-1}(y) = P g_{\mathrm{BL}}^{-1}(x) P^{\mathsf{T}}$, the Jacobian of P being cancelled by the normalization by the volume of the ball. Hence P is an isometry from $(T_x\mathscr{M}, g_{\mathrm{BL}}(x))$ to $(T_y\mathscr{M}, g_{\mathrm{BL}}(y))$ which maps $\mathscr{B}_x$ onto $\mathscr{B}_y$, so that $\mathrm{vol}_{\mathrm{BL}}(\mathscr{B}_x) = \mathrm{vol}_{\mathrm{BL}}(\mathscr{B}_y)$. Since $\mathscr{M}$ is connected, w is constant, and Equation (24) reduces to Equation (7).

Riemannian metrics. Let $\mathscr{F}(x,v) = \sqrt{v^{\mathsf{T}}\mathbf{A}(x)v}$. Its fundamental tensor is $g_{ij}(x,v) = \mathbf{A}_{ij}(x)$ for every v, hence does not depend on the direction, so $\mathscr{F}$ is Berwald. Its unit ball $\mathscr{B}_x = \{v : v^{\mathsf{T}}\mathbf{A}(x)v \le 1\}$ is the ellipsoid defined by $\mathbf{A}(x)$, whose second moment is given by the classical identity $\int_{\mathscr{B}_x} v^i v^j \mathrm{d}\lambda_x(v) = \frac{\lambda_x(\mathscr{B}_x)}{m+2}\mathbf{A}^{ij}(x)$. Substituting it into Equation (10) gives $g_{\mathrm{BL}}^{ij} = \mathbf{A}^{ij}$, i.e. $g_{\mathrm{BL}} = \mathbf{A}$. Equation (7) then reads

$$\mathscr{L}^s f = c_2\left(\Delta_{\mathbf{A}} f + 2(1-\theta)\langle \nabla \log \rho, \nabla f\rangle_{\mathbf{A}}\right),$$

which is the θ-normalized diffusion generator of Coifman & Lafon (2006). Note also that the S-curvature of a Riemannian metric vanishes, so that $s_0 = s_1 = 0$ and, in this setting, Υ reduces to

$$\Upsilon(x) = -\frac{1}{3}\sigma_{\mathrm{BH}}(x)\int_{T_x\mathscr{M}} K(\mathscr{F}(x,v))\mathrm{Ric}^{\mathscr{F}}(x,v)\mathrm{d}\lambda_x(v).$$

Finally, a Riemannian metric is reversible, so $v \mapsto K(\mathscr{F}(x,v))$ is an even function on $T_x\mathscr{M}$, and the change of variable $v \mapsto -v$ gives $m_1 = 0$. By Theorem 2, $\mathscr{L}^a f = \tilde{m}_1^i \partial_i f = 0$.

D.6 Proof of Theorem 4

This subsection is devoted to the proof of Theorem 4. We first give the matrix forms of the discrete operators, then we prove the convergence of the empirical operators to their population counterparts, and finally we prove the convergence of the population operators to their limit.

Matrix forms

We start by recalling the basic definitions of the discrete operators. The matrices of Section 4 are all built from the empirical kernel matrix $\mathbf{W}_N$, and the two members of the family $\mathbf{P}^{(\theta,\bullet)}$ take a slightly different shape. For $\bullet = s$, the matrix $\mathbf{D}_N^{(\theta,s)}$ is by construction the degree matrix of $\mathbf{W}_N^{(\theta,s)}$, so that the two terms merge into

$$\mathbf{P}^{(\theta,s)}[f] = \left(\left(\mathbf{D}_N^{(\theta,s)}\right)^{-1}\mathbf{W}_N^{(\theta,s)} - \mathbf{I}\right)\mathbf{f},$$

a normalized-graph-Laplacian-type matrix, whose eigenvectors provide the embedding coordinates (Section 5.3). For $\bullet = a$, the matrix $\mathbf{D}_N^{(\theta,a)}$ is not the degree matrix of $\mathbf{W}_N^{(\theta,s)}$, and the two terms remain separate:

$$\mathbf{P}^{(\theta,a)}[f] = \left(\mathbf{D}_N^{(\theta,s)}\right)^{-1}\mathbf{W}_N^{(\theta,a)}\mathbf{f} - \left(\mathbf{D}_N^{(\theta,s)}\right)^{-1}\mathbf{D}_N^{(\theta,a)}\mathbf{f}$$

is the estimator of the advection term, whose limit is $\varepsilon\tilde{m}_1^i\partial_i f$.

From finite samples to the limit operator: setting and proof sketch

We now let $N \to \infty$ and $\varepsilon = \varepsilon(N) \to 0$ jointly. Throughout, C denotes a constant depending on $\mathscr{M}$, $\mathscr{F}$, K, ρ, θ and f, but never on N or ε, and possibly changing from line to line.

It is convenient to extend the discrete objects to the whole manifold, so that they can be compared pointwise with their continuous counterparts. For $x, y \in \mathscr{M}$, we set

$$\hat{d}_N(x) = \frac{1}{N}\sum_{j=1}^N W(x,X_j), \qquad \hat{d}'_N(x) = \frac{1}{N}\sum_{j=1}^N W(X_j,x), \qquad \hat{q}_N = \frac{\hat{d}_N + \hat{d}'_N}{2},$$

$$\hat{W}_N^{(\theta)}(x,y) = \frac{W(x,y)}{\hat{q}_N(x)^\theta \hat{q}_N(y)^\theta}, \qquad \hat{W}_N^{(\theta,\bullet)}(x,y) = \frac{\hat{W}_N^{(\theta)}(x,y) \pm \hat{W}_N^{(\theta)}(y,x)}{2},$$

$$\hat{d}_N^{(\theta,\bullet)}(x) = \frac{1}{N}\sum_{j=1}^N \hat{W}_N^{(\theta,\bullet)}(x,X_j),$$

with the sign $+$ for $\bullet = s$ and $-$ for $\bullet = a$, and finally

$$\hat{\mathbf{P}}_N^{(\theta,\bullet)}[f](x) = \frac{1}{\hat{d}_N^{(\theta,s)}(x)}\left(\frac{1}{N}\sum_{j=1}^N \hat{W}_N^{(\theta,\bullet)}(x,X_j)f(X_j) - f(x)\hat{d}_N^{(\theta,\bullet)}(x)\right). \tag{25}$$

We also write $\hat{\mathbf{L}}_N^s = \varepsilon^{-2}\hat{\mathbf{P}}_N^{(\theta,s)}$ and $\hat{\mathbf{L}}_N^a = \varepsilon^{-1}\hat{\mathbf{P}}_N^{(\theta,a)}$ for the rescaled versions, extending $\mathbf{L}^s$ and $\mathbf{L}^a$ in the same way. Evaluated at the sample points $x = X_i$, these are exactly the matrices of Section 4, the factors $1/N$ cancelling in the ratio Equation (25), so that $\hat{\mathbf{P}}_N^{(\theta,\bullet)}[f](X_i) = \big(\mathbf{P}^{(\theta,\bullet)}\mathbf{f}\big)_i$ and $\hat{\mathbf{L}}_N^{\bullet}[f](X_i) = \big(\mathbf{L}^{\bullet}\mathbf{f}\big)_i$. Their continuous counterparts are d, d', q_ε, $W^{(\theta)}$, $d^{(\theta,s)} = \mathscr{G}_{\mathscr{F}}^{(\theta,s)}[\rho]$, $d^{(\theta,a)} = \mathscr{G}_{\mathscr{F}}^{(\theta,a)}[\rho]$ and $\mathscr{P}^{(\theta,\bullet)}$.

Two errors then separate the empirical operator from its limit, and we bound them one at a time. Writing the symmetric case for concreteness, and inserting the population operator $\mathscr{P}^{(\theta,s)}$ between the two, we have that:

$$\big|\hat{\mathbf{L}}_N^s[f] - \mathscr{L}^s f\big| \le \underbrace{\frac{1}{\varepsilon^2}\big|\hat{\mathbf{P}}_N^{(\theta,s)}[f] - \mathscr{P}^{(\theta,s)}[f]\big|}_{\text{fluctuation}} + \underbrace{\frac{1}{\varepsilon^2}\big|\mathscr{P}^{(\theta,s)}[f] - \varepsilon^2\mathscr{L}^s f\big|}_{\text{bias}}. \tag{26}$$

The *bias* is deterministic: it is the gap between the population operator and its limit, and it decreases with ε. The *fluctuation* is stochastic: it is the gap between the empirical operator and its population counterpart, and it grows as ε decreases, since the kernel then concentrates and fewer samples contribute to each evaluation. The next two subsections bound each of them, and the last one combines the two and reads off the admissible scalings of $\varepsilon(N)$.

Bounding the bias

The deterministic part of the error is the expansion already established in Appendix D.3, read quantitatively rather than as a limit.

Lemma 36 (Bias). *Let $f \in \mathscr{C}^3(\mathscr{M})$. There exist $C, \varepsilon_0 > 0$ such that for $\varepsilon \le \varepsilon_0$,*

$$\sup_{x\in\mathscr{M}}\big|\mathscr{P}^{(\theta,s)}[f](x) - \varepsilon^2\mathscr{L}^s f(x)\big| \le C\varepsilon^3, \qquad \sup_{x\in\mathscr{M}}\big|\mathscr{P}^{(\theta,a)}[f](x) - \varepsilon\mathscr{L}^a f(x)\big| \le C\varepsilon^2.$$

Proof. The proof of Theorem 2 in Appendix D.3 establishes these expansions with an explicit remainder, inherited from the $\mathcal{O}(\varepsilon^3)$ of Theorem 1. The remainder is uniform in x, since $\mathscr{M}$ is compact and the derivatives of f are bounded. □

Bounding the fluctuation

We now bound the stochastic part of the error, which is the deviation of the empirical operator from its population counterpart. The next proposition gives a uniform bound on this deviation, which is the main technical result of this section.

Proposition 37 (Fluctuation). *Assume $\varepsilon = \varepsilon(N) \to 0$ with $N\varepsilon^m/\log N \to \infty$, and let $f \in \mathscr{C}^1(\mathscr{M})$. Then there is a constant C such that, almost surely for N large enough,*

$$\sup_{x\in\mathscr{M}}\Big|\hat{\mathbf{P}}_N^{(\theta,\bullet)}[f](x) - \mathscr{P}^{(\theta,\bullet)}[f](x)\Big| \le C\varepsilon\zeta_N, \qquad \bullet \in \{s,a\},$$

where

$$\zeta_N = \sqrt{\frac{\log N}{N\varepsilon^m}}. \tag{27}$$

To prove this bound, we rely on standard concentration inequalities for sums of independent random variables, which give a bound on the deviation of the empirical average from its expectation. To do so, we use a δ-net argument[7]: we first control the deviation on a finite set of points, then extend it to the whole manifold by continuity, and conclude almost surely with the Borel–Cantelli lemma.

[7] A δ-net is a finite set of points in a metric space such that every point in the space is within distance δ of at least one point in the net.

Proof of Proposition 37. Defining the net. The intrinsic distance $\mathrm{dst}_{\mathscr{F}}$ is asymmetric, so that it yields two different notions of δ-net, depending on whether we use forward or backward balls. We therefore rely instead on the Euclidean distance $|\cdot|$ of $\mathbb{R}^D$ restricted to the tangent spaces of $\mathscr{M}$, which is symmetric and induces a natural Riemannian distance dist on $\mathscr{M}$. In particular, the distances $\mathrm{dst}_{\mathscr{F}}$ and dist are comparable: the map $(x,v)\mapsto \mathscr{F}(x,v)$ is continuous and positive on the unit sphere bundle $\{(x,v)\in T\mathscr{M} : |v|=1\}$, which is compact by (A2), so that it is bounded there between two constants. By positive homogeneity, we get that

$$\kappa_-|v| \le \mathscr{F}(x,v) \le \kappa_+|v|, \qquad (x,v)\in T\mathscr{M}, \qquad \text{for some } 0<\kappa_-\le\kappa_+<\infty.$$

Hence, for any $x,y\in\mathscr{M}$,

$$\kappa_- \,\mathrm{dist}(x,y) \le \mathrm{dst}_{\mathscr{F}}(x,y) \le \kappa_+ \,\mathrm{dist}(x,y). \tag{28}$$

Once this is established, we can define a δ-net according to this Riemannian distance. Counting the number of points in such a net is standard, and we give the result below for completeness. In this section, Vol denotes the Riemannian volume of $\mathscr{M}$ associated with the Riemannian metric that induces dist.

Lemma 38. *There are $C_{\mathscr{M}},\delta_0>0$ such that, for every $\delta\le\delta_0$, $\mathscr{M}$ admits a δ-net of cardinality $n_\delta \le C_{\mathscr{M}}\delta^{-m}$.*

Proof sketch. The input is the volume of small balls. On a compact Riemannian manifold, the volume density in normal coordinates around x is $1+\mathcal{O}(r^2)$, with a remainder that is uniform in x since the curvature is bounded and the injectivity radius is bounded below, hence there are $v_-,v_+,\delta_0>0$ such that $v_-r^m \le \mathrm{Vol}\big(B(x,r)\big)\le v_+r^m$ for every $x\in\mathscr{M}$ and every $r\le\delta_0$. Let $x_1,\dots,x_{n_\delta}$ be a maximal δ-separated subset of $\mathscr{M}$, which exists by compactness. It is a δ-net: any x at distance more than δ from all the x_k could be added to the family, contradicting maximality. Moreover the balls $B(x_k,\delta/2)$ are pairwise disjoint, by δ-separation, so that summing their volumes gives $n_\delta v_-(\delta/2)^m\le\mathrm{Vol}(\mathscr{M})$, which is the announced bound. The same packing argument, in the Euclidean setting, is detailed in Vershynin (2019, Section 4.2). □

Using this δ-net, the supremum over $\mathscr{M}$ of a function g reduces to a maximum over the net points: picking for each x a net point x_k with $\mathrm{dist}(x,x_k)\le\delta$, and denoting by L_g a Lipschitz constant of g,

$$\sup_{x\in\mathscr{M}}|g(x)| \le \max_{1\le k\le n_\delta}|g(x_k)| + L_g\delta.$$

Deterministic estimates on the kernel. Besides the net, we need three elementary properties of the kernel W: it is of size ε^{-m}, it is concentrated at scale ε around the diagonal, and it varies at scale ε. We write $W^\bullet(x,y)=\big(W(x,y)\pm W(y,x)\big)/2$ for its symmetric ($\bullet=s$) and antisymmetric ($\bullet=a$) parts, so that $|W^\bullet|\le W^s$.

Lemma 39 (Kernel at scale ε)**.** *There is a constant C such that, for every $\varepsilon\le1$, every $x,x',y\in\mathscr{M}$ and every $p\in\{0,1\}$,*

$$W^s(x,y)\le\frac{K(0)}{\varepsilon^m}, \qquad W^s(x,y)\,\mathrm{dist}(x,y)\le\frac{C}{\varepsilon^{m-1}}, \qquad \int_{\mathscr{M}} W^s(x,y)\,\mathrm{dist}(x,y)^p\mathrm{d}\mathbb{P}(y)\le C\varepsilon^p, \tag{29}$$

and, for $\bullet\in\{s,a\}$,

$$\big|W^\bullet(x,y)-W^\bullet(x',y)\big|\le\frac{C\,\mathrm{dist}(x,x')}{\varepsilon^{m+1}}. \tag{30}$$

Proof. Since K is decreasing and $\mathrm{dst}_{\mathscr{F}}\ge0$, both $W(x,y)$ and $W(y,x)$ are bounded by $K(0)\varepsilon^{-m}$, which gives the first bound. For the second one, the distance comparison inequality (Equation (28)) gives $\mathrm{dst}_{\mathscr{F}}(x,y)\ge\kappa_-\,\mathrm{dist}(x,y)$ and $\mathrm{dst}_{\mathscr{F}}(y,x)\ge\kappa_-\,\mathrm{dist}(x,y)$, so that $W^s(x,y)\le\varepsilon^{-m}K\big(\kappa_-\,\mathrm{dist}(x,y)/\varepsilon\big)$. Writing $r=\kappa_-\,\mathrm{dist}(x,y)/\varepsilon$, it holds that

$$W^s(x,y)\,\mathrm{dist}(x,y)\le\frac{\varepsilon^{1-m}}{\kappa_-}rK(r)\le\frac{\varepsilon^{1-m}}{\kappa_-}\sup_{r\ge0}rK(r),$$

which is finite since K is sub-exponential (A1). For the third one, we use that $\mathrm{Vol}\big(B(x,r)\big) \le v_+ r^m$ for every $r > 0$: this is shown in the proof of Lemma 38 for $r \le \delta_0$, and it extends to $r > \delta_0$ by enlarging v_+ by $\mathrm{Vol}(\mathscr{M})\delta_0^{-m}$. We split $\mathscr{M}$ into the annuli $A_k = \{y \in \mathscr{M} : k\varepsilon \le \mathrm{dist}(x,y) < (k+1)\varepsilon\}$, $k \ge 0$. On A_k, we have $W^s(x,y) \le \varepsilon^{-m} K(\kappa_- k) \le C_K \varepsilon^{-m} e^{-\nu_K \kappa_- k}$ and $\mathrm{dist}(x,y)^p \le \big((k+1)\varepsilon\big)^p$, and moreover $\mathrm{Vol}(A_k) \le v_+\big((k+1)\varepsilon\big)^m$. Since $\mathrm{d}\mathbb{P} = \rho \mathrm{d}\mathfrak{m}_{\mathrm{BH}}$, where ρ and the ratio of the densities of $\mathfrak{m}_{\mathrm{BH}}$ and Vol are bounded on the compact manifold $\mathscr{M}$, we have $\mathrm{d}\mathbb{P} \le C d\mathrm{Vol}$, and we get

$$\int_{\mathscr{M}} W^s(x,y)\,\mathrm{dist}(x,y)^p \mathrm{d}\mathbb{P}(y) \le \frac{C}{\varepsilon^m} \sum_{k\ge 0} e^{-\nu_K \kappa_- k} \big((k+1)\varepsilon\big)^{m+p} = C\varepsilon^p \sum_{k\ge 0} (k+1)^{m+p} e^{-\nu_K \kappa_- k},$$

and the series converges. Finally, for the Lipschitz bound, the triangle inequality for $\mathrm{dst}_{\mathscr{F}}$ gives $\mathrm{dst}_{\mathscr{F}}(x,y) \le \mathrm{dst}_{\mathscr{F}}(x,x') + \mathrm{dst}_{\mathscr{F}}(x',y)$ and $\mathrm{dst}_{\mathscr{F}}(x',y) \le \mathrm{dst}_{\mathscr{F}}(x',x) + \mathrm{dst}_{\mathscr{F}}(x,y)$, so that, by Equation (28),

$$\big|\mathrm{dst}_{\mathscr{F}}(x,y) - \mathrm{dst}_{\mathscr{F}}(x',y)\big| \le \max\{\mathrm{dst}_{\mathscr{F}}(x,x'), \mathrm{dst}_{\mathscr{F}}(x',x)\} \le \kappa_+ \,\mathrm{dist}(x,x'),$$

and the same holds in the second argument. Since K' is bounded by (A1), we get $|W(x,y) - W(x',y)| \le \kappa_+ \|K'\|_\infty \varepsilon^{-m-1} \,\mathrm{dist}(x,x')$, and similarly for $x \mapsto W(y,x)$. The bound Equation (30) follows, since $W^\bullet$ is the half-sum or the half-difference of these two functions. □

We also need the normalization q_ε to be bounded above and away from zero so that the normalized kernel is well-defined. On the one hand, the assumption on the degrees gives $q_\varepsilon = (d'+d)/2 > d_{\min}$. On the other hand, since $q_\varepsilon(x) = \int_{\mathscr{M}} W^s(x,y)\mathrm{d}\mathbb{P}(y)$, the case $p = 0$ of Equation (29) shows that q_ε is bounded. Hence, there are constants C_q, c_θ and C_θ, independent of ε, such that for every $x \in \mathscr{M}$,

$$d_{\min} \le q_\varepsilon(x) \le C_q \qquad \text{and} \qquad 0 < c_\theta \le q_\varepsilon(x)^{-\theta} \le C_\theta. \tag{31}$$

Reduction to numerator and denominator. To apply the δ-net argument, the natural choice is the function $g(x) = \hat{\mathbf{P}}_N^{(\theta,\bullet)}[f](x) - \mathscr{P}^{(\theta,\bullet)}[f](x)$. However, this function is not itself an empirical average, so we apply this argument to its numerator and its denominator separately. One can remark that Equation (25) can be rewritten as a ratio of two empirical averages:

$$\hat{\mathbf{P}}_N^{(\theta,\bullet)}[f](x) = \frac{\frac{1}{N}\sum_{j=1}^N \hat{W}_N^{(\theta,\bullet)}(x,X_j)(f(X_j) - f(x))}{\frac{1}{N}\sum_{j=1}^N \hat{W}_N^{(\theta,s)}(x,X_j)}.$$

Moreover, by using the definition of θ-normalization, it holds:

$$\hat{\mathbf{P}}_N^{(\theta,\bullet)}[f](x) = \frac{\frac{1}{N}\sum_{j=1}^N W^\bullet(x,X_j)\hat{q}_N(X_j)^{-\theta}(f(X_j) - f(x))}{\frac{1}{N}\sum_{j=1}^N W^s(x,X_j)\hat{q}_N(X_j)^{-\theta}}.$$

Importantly, the same rewriting holds for the population operator $\mathscr{P}^{(\theta,\bullet)}[f]$, with $\hat{q}_N$ replaced by q_ε and the empirical average replaced by an integral over $\mathscr{M}$. Finding bounds on the deviation of $\hat{\mathbf{P}}_N^{(\theta,\bullet)}[f]$ from $\mathscr{P}^{(\theta,\bullet)}[f]$ therefore reduces to bounding the deviation of the numerator and denominator separately. Note that due to its definition, $\hat{q}_N$ is itself an empirical average, so that the numerator and denominator are not sums of independent random variables.

Lemma 40 (Uniform deviation of an empirical average)**.** *Let $X_1, \dots, X_N$ be i.i.d. with law $\mathbb{P}$, and let $\psi : \mathscr{M} \times \mathscr{M} \to \mathbb{R}$ be measurable, possibly depending on N, and assume that for every $x, x', y \in \mathscr{M}$,*

$$|\psi(y,x)| \le B, \qquad \int_{\mathscr{M}} \psi(y,x)^2 \mathrm{d}\mathbb{P}(y) \le \sigma^2, \qquad |\psi(y,x) - \psi(y,x')| \le L\,\mathrm{dist}(x,x'),$$

where B, σ and L may depend on N. Then, with $A_0 = 2(5m+2)$, almost surely for N large enough,

$$\sup_{x\in\mathscr{M}} \left| \frac{1}{N}\sum_{j=1}^N \psi(X_j,x) - \int_{\mathscr{M}} \psi(y,x)\mathrm{d}\mathbb{P}(y) \right| \le A_0 \left(\sigma \sqrt{\frac{\log N}{N}} + \frac{B \log N}{N} \right) + \frac{2L}{N^5}. \tag{32}$$

Proof. Let $\delta = N^{-5}$ and let $x_1, \dots, x_{n_\delta}$ be a δ-net of $\mathscr{M}$, with $n_\delta \le C_{\mathscr{M}} N^{5m}$ by Lemma 38, which holds for N large so that $\delta \le \delta_0$. For any $x \in \mathscr{M}$, let x_k be a net point with $\operatorname{dist}(x, x_k) \le \delta$. By the triangle inequality,

$$\left| \frac{1}{N}\sum_{j=1}^{N} \psi(X_j, x) - \int_{\mathscr{M}} \psi(y,x)\mathrm{d}\mathbb{P}(y) \right| \le \underbrace{\left| \frac{1}{N}\sum_{j=1}^{N} \big(\psi(X_j,x) - \psi(X_j,x_k)\big) \right|}_{(A)} + \underbrace{\left| \frac{1}{N}\sum_{j=1}^{N} \psi(X_j,x_k) - \int_{\mathscr{M}} \psi(y,x_k)\mathrm{d}\mathbb{P}(y) \right|}_{(B)} + \underbrace{\left| \int_{\mathscr{M}} \big(\psi(y,x_k) - \psi(y,x)\big)\mathrm{d}\mathbb{P}(y) \right|}_{(C)}.$$

For (A) and (C), the Lipschitz assumption gives (A) $\le L\delta$ and (C) $\le L\delta$, for every realization of the sample.

For (B), fix a net point x_k. The net points do not depend on the sample, so $Y_j = \psi(X_j, x_k)$ are i.i.d. random variables with mean $\mu_k = \int_{\mathscr{M}} \psi(y, x_k)\mathrm{d}\mathbb{P}(y)$. By assumption, Y_j and $-Y_j$ are bounded above by B and have second moment at most σ^2. Bernstein's inequality, applied to both of them, gives for any $t > 0$

$$\mathbb{P}\left(\left| \frac{1}{N}\sum_{j=1}^{N} Y_j - \mu_k \right| > t \right) \le 2\exp\left(-\frac{Nt^2}{2(\sigma^2 + Bt/3)} \right). \tag{33}$$

We set $t = A_0\left(\sigma\sqrt{\frac{\log N}{N}} + \frac{B\log N}{N}\right)$ and check that the exponent is then at least $A_0 \log N/2$. Let $S = \sigma^2 + B^2 \log N/N$, which we can assume positive, since otherwise $\psi = 0$ and there is nothing to prove. For the numerator, dropping the nonnegative cross term in the square, we have

$$Nt^2 \ge N A_0^2 \left(\sigma^2 \frac{\log N}{N} + B^2 \frac{(\log N)^2}{N^2} \right) = A_0^2 S \log N.$$

For the denominator, the inequality $2ab \le a^2 + b^2$ gives $\sigma B\sqrt{\log N/N} \le S/2$, and we have $B^2 \log N/N \le S$, so that

$$\begin{aligned} 2\left(\sigma^2 + \frac{Bt}{3}\right) &= 2\sigma^2 + \frac{2A_0}{3}\left(\sigma B\sqrt{\frac{\log N}{N}} + \frac{B^2 \log N}{N}\right) \\ &\le 2S + \frac{2A_0}{3}\left(\frac{S}{2} + S\right) = (2 + A_0)S. \end{aligned}$$

Combining the two, the exponent in Equation (33) is at least $\frac{A_0^2}{2+A_0}\log N$, and

$$\frac{A_0^2}{2+A_0} \ge \frac{A_0}{2} \qquad \Longleftrightarrow \qquad 2A_0 \ge 2 + A_0 \qquad \Longleftrightarrow \qquad A_0 \ge 2,$$

which holds since $A_0 = 2(5m+2)$. Hence (B) exceeds t at the point x_k with probability at most $2N^{-A_0/2}$. Taking a union bound over the n_δ net points, the maximum of (B) over the net exceeds t with probability at most

$$2n_\delta N^{-A_0/2} = 2n_\delta N^{-(5m+2)} \le 2C_{\mathscr{M}} N^{-2},$$

which is summable in N. By the Borel–Cantelli lemma, (B) $\le t$ for every net point, almost surely for N large enough. Combining the three terms gives Equation (32). □

To apply this lemma to the quantity of interest, we first need to remove the dependence of the kernel on the sample, which is done by replacing $\hat{q}_N$ by its expectation q_ε.

We can apply Lemma 40 to $\hat{q}_N$ since it is an empirical average of i.i.d. random variables, $\hat{q}_N(x) = \frac{1}{N}\sum_{j=1}^N W^s(x, X_j)$, with mean $q_\varepsilon(x)$. By Lemma 39 and Equation (31), the kernel $\psi(y,x) = W^s(x,y)$ satisfies the assumptions of the lemma with $B \le K(0)\varepsilon^{-m}$, $\sigma^2 \le Bq_\varepsilon(x) \le C\varepsilon^{-m}$ and $L \le C\varepsilon^{-m-1}$. Recalling that $\zeta_N^2 = \log N/(N\varepsilon^m)$, the right-hand side of Equation (32) is thus at most

$$C\left(\sqrt{\frac{\log N}{N\varepsilon^m}} + \frac{\log N}{N\varepsilon^m}\right) + \frac{C}{\varepsilon^{m+1}N^5} = C\left(\zeta_N + \zeta_N^2\right) + \frac{C}{\varepsilon^{m+1}N^5}.$$

The assumption $N\varepsilon^m/\log N \to \infty$ gives $\zeta_N \to 0$, as well as $\varepsilon^{-m} \le N$ and $\varepsilon^{-1} \le N^{1/m} \le N$ for N large enough, so that the last term is at most CN^{-3}. Since moreover $\zeta_N \ge \sqrt{\log N/N} \ge N^{-1/2}$ and $\varepsilon \ge N^{-1}$, we have $N^{-3} \le \varepsilon\zeta_N$, so that this term is negligible here and in all the bounds below. Hence, almost surely for N large enough,

$$\sup_{x\in\mathcal{M}} |\hat{q}_N(x) - q_\varepsilon(x)| \le C\zeta_N. \tag{34}$$

Using this, we can rewrite the numerator of $\hat{\mathbf{P}}_N^{(\theta,\bullet)}[f]$ as

$$\begin{aligned}
&\frac{1}{N}\sum_{j=1}^N W^\bullet(x, X_j)\hat{q}_N(X_j)^{-\theta}(f(X_j) - f(x)) \\
&\quad= \underbrace{\frac{1}{N}\sum_{j=1}^N W^\bullet(x, X_j)q_\varepsilon(X_j)^{-\theta}(f(X_j) - f(x))}_{(\star)} \\
&\qquad+ \underbrace{\frac{1}{N}\sum_{j=1}^N W^\bullet(x, X_j)\left(\hat{q}_N(X_j)^{-\theta} - q_\varepsilon(X_j)^{-\theta}\right)(f(X_j) - f(x))}_{(\star\star)}.
\end{aligned}$$

We can bound the second term $(\star\star)$ by

$$\begin{aligned}
|(\star\star)| &\le \sup_{y\in\mathcal{M}} \left|\hat{q}_N(y)^{-\theta} - q_\varepsilon(y)^{-\theta}\right| \frac{1}{N}\sum_{j=1}^N W^s(x, X_j)|f(X_j) - f(x)| \\
&\le C \sup_{y\in\mathcal{M}} |\hat{q}_N(y) - q_\varepsilon(y)| \frac{1}{N}\sum_{j=1}^N W^s(x, X_j)|f(X_j) - f(x)| \\
&\le C\zeta_N \frac{1}{N}\sum_{j=1}^N W^s(x, X_j)|f(X_j) - f(x)|,
\end{aligned}$$

where we used that $|W^\bullet| \le W^s$, and that $t \mapsto t^{-\theta}$ is Lipschitz on $[d_{\min}/2, 2C_q]$, which contains all the values of $\hat{q}_N$ and q_ε for N large enough by Equations (31) and (34). Note that the supremum over $y \in \mathcal{M}$ controls in particular the values at the sample points X_j, even though they are random. The last factor is itself an empirical average of i.i.d. random variables, and we show below that it is at most $C\varepsilon$, so that $|(\star\star)| \le C\varepsilon\zeta_N$. The same reasoning applies to the denominator of $\hat{\mathbf{P}}_N^{(\theta,\bullet)}[f]$, without the factor $|f(X_j) - f(x)|$: its second term is at most $C\zeta_N\hat{q}_N(x) \le C\zeta_N$, since $\hat{q}_N \le 2C_q$ for N large. To conclude, we need to bound $(\star)$, which

is an empirical average of i.i.d. random variables. To do so, we must show that the lemma applies to the kernels $\psi_\bullet(y,x) = W^\bullet(x,y) q_\varepsilon(y)^{-\theta}(f(y)-f(x))$ and $\psi_s(y,x) = W^s(x,y) q_\varepsilon(y)^{-\theta}$.

By Lemma 39 and Equation (31), the kernel ψ_s satisfies the assumptions of Lemma 40 with $B \le C\varepsilon^{-m}$, $\sigma^2 \le C\varepsilon^{-m}$ and $L \le C\varepsilon^{-m-1}$, exactly as for $\hat{q}_N$. The kernel $\psi_\bullet$ carries the increment $f(y)-f(x)$, which is of size ε on the effective support of W. Since $f \in \mathscr{C}^1(\mathscr{M})$ is L_f-Lipschitz and $|W^\bullet| \le W^s$, Lemma 39 gives

$$B \le C_\theta L_f \sup_{y\in\mathscr{M}} W^s(x,y)\operatorname{dist}(x,y) \le C\varepsilon^{1-m},$$
$$\sigma^2 \le B\int_{\mathscr{M}} |\psi_\bullet(y,x)|\mathrm{d}\mathbb{P}(y) \le C\varepsilon^{1-m} C_\theta L_f \int_{\mathscr{M}} W^s(x,y)\operatorname{dist}(x,y)\mathrm{d}\mathbb{P}(y) \le C\varepsilon^{2-m},$$

and $L \le C\varepsilon^{-m-1}$, since

$$|\psi_\bullet(y,x) - \psi_\bullet(y,x')| \le 2C_\theta \|f\|_\infty |W^\bullet(x,y) - W^\bullet(x',y)| + C_\theta L_f W^s(x',y)\operatorname{dist}(x,x').$$

The same bounds hold for the kernel $W^s(x,y)|f(y)-f(x)|$ of the last factor in the bound on $(\star\star)$, using $\big||f(y)-f(x)| - |f(y)-f(x')|\big| \le L_f \operatorname{dist}(x,x')$. For these kernels, B and σ are smaller than for ψ_s by a factor ε, and so is the first term of the right-hand side of Equation (32), the last one being negligible as seen above. Hence, almost surely for N large enough,

$$\sup_{x\in\mathscr{M}} \left| \frac{1}{N}\sum_{j=1}^N W^\bullet(x,X_j) q_\varepsilon(X_j)^{-\theta}(f(X_j)-f(x)) - \int_{\mathscr{M}} W^\bullet(x,y) q_\varepsilon(y)^{-\theta}(f(y)-f(x))\mathrm{d}\mathbb{P}(y) \right| \le C\varepsilon\zeta_N,$$

and similarly $\frac{1}{N}\sum_j W^s(x,X_j)|f(X_j)-f(x)| \le \int_{\mathscr{M}} W^s(x,y)|f(y)-f(x)|\mathrm{d}\mathbb{P}(y) + C\varepsilon\zeta_N \le C\varepsilon$ for every $x \in \mathscr{M}$, which completes the bound on $(\star\star)$. The same reasoning applies to the denominator:

$$\sup_{x\in\mathscr{M}} \left| \frac{1}{N}\sum_{j=1}^N W^s(x,X_j) q_\varepsilon(X_j)^{-\theta} - \int_{\mathscr{M}} W^s(x,y) q_\varepsilon(y)^{-\theta}\mathrm{d}\mathbb{P}(y) \right| \le C\zeta_N.$$

Thus, it remains to show that the ratio of these two empirical averages is close to the ratio of their expectations.

Combining the bounds on $(\star)$ and $(\star\star)$, and their analogues for the denominator, we get that, almost surely for N large enough and for every $x \in \mathscr{M}$, the numerator $\mathscr{N}_N(x)$ and the denominator $\mathscr{D}_N(x)$ of $\hat{\mathbf{P}}_N^{(\theta,\bullet)}[f](x)$ satisfy

$$|\mathscr{N}_N(x) - \mathscr{N}(x)| \le C\varepsilon\zeta_N, \qquad |\mathscr{D}_N(x) - \mathscr{D}(x)| \le C\zeta_N,$$

where $\mathscr{N}(x) = \int_{\mathscr{M}} W^\bullet(x,y) q_\varepsilon(y)^{-\theta}(f(y)-f(x))\mathrm{d}\mathbb{P}(y)$ and $\mathscr{D}(x) = \int_{\mathscr{M}} W^s(x,y) q_\varepsilon(y)^{-\theta}\mathrm{d}\mathbb{P}(y)$ are the numerator and the denominator of $\mathscr{P}^{(\theta,\bullet)}[f](x)$, up to the common factor $q_\varepsilon(x)^{-\theta}$. Moreover, since $\int_{\mathscr{M}} W^s(x,y)\mathrm{d}\mathbb{P}(y) = q_\varepsilon(x)$, Equation (31) and Lemma 39 give $\mathscr{D}(x) \ge c_\theta q_\varepsilon(x) \ge c_\theta d_{\min}$ and $|\mathscr{N}(x)| \le C_\theta L_f \int_{\mathscr{M}} W^s(x,y)\operatorname{dist}(x,y)\mathrm{d}\mathbb{P}(y) \le C\varepsilon$. In particular $\mathscr{D}_N(x) \ge c_\theta d_{\min}/2$ for N large. Adding and subtracting $\mathscr{N}(x)/\mathscr{D}_N(x)$, we get

$$\left|\hat{\mathbf{P}}_N^{(\theta,\bullet)}[f](x) - \mathscr{P}^{(\theta,\bullet)}[f](x)\right| = \left| \frac{\mathscr{N}_N(x)-\mathscr{N}(x)}{\mathscr{D}_N(x)} + \frac{\mathscr{N}(x)\big(\mathscr{D}(x)-\mathscr{D}_N(x)\big)}{\mathscr{D}_N(x)\mathscr{D}(x)} \right|$$
$$\le \frac{2C\varepsilon\zeta_N}{c_\theta d_{\min}} + \frac{2C\varepsilon\cdot C\zeta_N}{(c_\theta d_{\min})^2} \le C\varepsilon\zeta_N.$$

The second term is small because the deviation of the denominator, of order ζ_N, is multiplied by $\mathscr{N}(x)$, which is of order ε. Since this holds for every $x \in \mathscr{M}$, almost surely for N large enough, this gives Proposition 37, which concludes the proof. □

Combining the bias and the fluctuation

Combining the fluctuation and the bias gives the quantitative form of Theorem 4.

Theorem 41. *Let $f \in \mathscr{C}^3(\mathscr{M})$ and let $\varepsilon = \varepsilon(N) \to 0$ with $N\varepsilon^m/\log N \to \infty$. Then, almost surely for N large enough,*

$$\sup_{x\in\mathscr{M}} \left|\hat{\mathbf{L}}_N^s[f](x) - \mathscr{L}^s f(x)\right| \le C\left(\varepsilon + \frac{\zeta_N}{\varepsilon}\right), \qquad \sup_{x\in\mathscr{M}} \left|\hat{\mathbf{L}}_N^a[f](x) - \mathscr{L}^a f(x)\right| \le C(\varepsilon + \zeta_N), \tag{35}$$

with ζ_N as in Equation (27). In particular, both right-hand sides vanish, and $\mathbf{L}^\bullet[f] \to \mathscr{L}^\bullet f$ uniformly and almost surely, as soon as

$$\varepsilon(N) \to 0 \qquad \text{and} \qquad \frac{N\varepsilon(N)^{m+2}}{\log N} \to \infty. \tag{36}$$

Proof. By definition $\hat{\mathbf{L}}_N^s = \varepsilon^{-2}\hat{\mathbf{P}}_N^{(\theta,s)}$, so that, splitting the error into its fluctuation and its bias, we have that:

$$\begin{aligned}
\left|\hat{\mathbf{L}}_N^s[f] - \mathscr{L}^s f\right| &\le \frac{1}{\varepsilon^2}\left|\hat{\mathbf{P}}_N^{(\theta,s)}[f] - \mathscr{P}^{(\theta,s)}[f]\right| + \frac{1}{\varepsilon^2}\left|\mathscr{P}^{(\theta,s)}[f] - \varepsilon^2\mathscr{L}^s f\right| \\
&\le \frac{C\varepsilon\zeta_N}{\varepsilon^2} + \frac{C\varepsilon^3}{\varepsilon^2} && \text{By Proposition 37 and Lemma 36} \\
&= C\left(\frac{\zeta_N}{\varepsilon} + \varepsilon\right).
\end{aligned}$$

The second bound is obtained in the same way, with ε^{-1} in place of ε^{-2}. Under Equation (36), we have that $\zeta_N/\varepsilon = \sqrt{\log N/(N\varepsilon^{m+2})} \to 0$, so that both bounds tend to 0. □

Remark 42 (Which scaling to choose)**.** *The two terms of Equation (35) vary in opposite directions with ε: a large bandwidth averages over many samples but blurs the geometry, while a small one is faithful but noisy. Balancing ε against ζ_N/ε gives the optimal choice*

$$\varepsilon(N) \asymp \left(\frac{\log N}{N}\right)^{\frac{1}{m+4}}, \qquad \text{for an error of order } \left(\frac{\log N}{N}\right)^{\frac{1}{m+4}},$$

which exhibits the usual curse of dimensionality in the intrinsic dimension m, as in the Riemannian case (Singer, 2006; Hein et al., 2007). We note that the antisymmetric part is the easier of the two: being rescaled by ε^{-1} instead of ε^{-2}, it only requires $N\varepsilon^m/\log N \to \infty$, whereas the symmetric part requires the stronger condition Equation (36). In that sense, estimating the direction of the data is statistically cheaper than estimating its geometry.

D.7 Computation of the moments in the Randers setting

If $\mathscr{F}$ is a Randers metric, the moments of the unit ball $\mathscr{B}_x$ can be computed in closed form. We recall that a Randers metric is defined as $\mathscr{F}(x,v) = \sqrt{v^\mathsf{T}\mathbf{A}(x)v} + b(x)^\mathsf{T}v$, where $\mathbf{A}(x)$ is a positive-definite matrix and $b(x)$ is a vector field such that $\|b(x)\|_{\mathbf{A}^{-1}} < 1$.

Proposition 43. *For a Randers metric, we have that:*

$$\lambda_x(\mathscr{B}_x) = \mathrm{Vol}_{\mathrm{Euc}}(\mathbb{B}^m)(1 - \|b(x)\|^2_{\mathbf{A}^{-1}})^{-\frac{m+1}{2}}\sqrt{\det \mathbf{A}^{-1}(x)}, \tag{37}$$

$$\mathbf{c}^i(x) = -\frac{b^i(x)}{1 - \|b(x)\|^2_{\mathbf{A}^{-1}}}, \tag{38}$$

$$\mathbf{S}^{ij}(x) = \frac{R_x^2}{m+2}\mathbf{E}^{ij}(x) + \mathbf{c}^i(x)\mathbf{c}^j(x), \tag{39}$$

where $R_x^2 = \frac{1}{1-\|b(x)\|^2_{\mathbf{A}^{-1}}}$, $b^i(x) = \mathbf{A}^{ij}(x)b_j(x)$, $\mathbf{E}_{ij}(x) = \mathbf{A}_{ij}(x) - b_i(x)b_j(x)$ *and* $\mathbf{E}^{ij} = (\mathbf{E}_{ij})^{-1} = \mathbf{A}^{ij} + \frac{b^i b^j}{1-\|b(x)\|^2_{\mathbf{A}^{-1}}}$.

Proof. Using (Ohta, 2021, Lemma 10.8), we have that:

$$\mathscr{B}_x = \{v \in T_x\mathscr{M} : \mathbf{E}_{ij}(x)(v-v_0)^i(v-v_0)^j \leq R_x^2\}$$

where $\mathbf{E}_{ij}(x) = \mathbf{A}_{ij}(x) - b_i(x)b_j(x)$, $v_0^i = -\frac{\mathbf{A}^{ij}(x)b_j(x)}{1-\|b(x)\|^2_{\mathbf{A}^{-1}}}$ and $R_x^2 = \frac{1}{1-\|b(x)\|^2_{\mathbf{A}^{-1}}}$. This shows that the unit tangent ball of a Randers metric is an ellipsoid, described by the matrix $\mathbf{E}(x)$, centered at v_0 and with a radius R_x.

We can compute the volume of the unit ball by using the change of variable $u = \mathbf{E}^{1/2}(x)(v-v_0)$, which yields that:

$$\lambda_x(\mathscr{B}_x) = \frac{R_x^m}{\sqrt{\det \mathbf{E}(x)}}\mathrm{Vol}_{\mathrm{Euc}}(\mathbb{B}^m) = \mathrm{Vol}_{\mathrm{Euc}}(\mathbb{B}^m)R_x^m\sqrt{\det \mathbf{E}^{-1}(x)}$$

We have the expression of R_x^m: $R_x^m = (1-\|b(x)\|^2_{\mathbf{A}^{-1}})^{-\frac{m}{2}}$. Moreover, the matrix determinant lemma yields $\det(\mathbf{E}) = \det(\mathbf{A} - bb^\mathsf{T}) = \det(\mathbf{A})(1-\|b(x)\|^2_{\mathbf{A}^{-1}})$, which yields that $\sqrt{\det \mathbf{E}^{-1}(x)} = \sqrt{\det \mathbf{A}^{-1}(x)}(1-\|b(x)\|^2_{\mathbf{A}^{-1}})^{-\frac{1}{2}}$. Hence the expression of the volume of the unit ball as:

$$\lambda_x(\mathscr{B}_x) = \mathrm{Vol}_{\mathrm{Euc}}(\mathbb{B}^m)(1-\|b(x)\|^2_{\mathbf{A}^{-1}})^{-\frac{m+1}{2}}\sqrt{\det \mathbf{A}^{-1}(x)}.$$

Now we can compute the first moments. Because of the symmetry around the center v_0, we have that $\int_{\mathscr{B}_x}(v-v_0)^i \mathrm{d}\lambda_x(v) = 0$. This yields that $\int_{\mathscr{B}_x} v^i \mathrm{d}\lambda_x(v) = \lambda_x(\mathscr{B}_x)v_0^i$ and thus:

$$\mathbf{c}^i(x) = \frac{\int_{\mathscr{B}_x} v^i \mathrm{d}\lambda_x(v)}{\lambda_x(\mathscr{B}_x)} = v_0^i = -\frac{\mathbf{A}^{ij}(x)b_j(x)}{1-\|b(x)\|^2_{\mathbf{A}^{-1}}} = -\frac{b^i(x)}{1-\|b(x)\|^2_{\mathbf{A}^{-1}}}.$$

And to compute the second moments, we observe that:

$$v^i v^j = (v-v_0)^i(v-v_0)^j + (v-v_0)^i v_0^j + v_0^i(v-v_0)^j + v_0^i v_0^j.$$

Here the cross terms vanish when integrated over $\mathscr{B}_x$ because of the symmetry around the center v_0. Using the change of variable $u = \mathbf{E}^{1/2}(x)(v-v_0)$, we can compute the first term as:

$$\begin{aligned}
\int_{\mathscr{B}_x}(v-v_0)^i(v-v_0)^j \mathrm{d}\lambda_x(v) &= \int_{\{u:\|u\|_2^2<R_x^2\}}(\mathbf{E}^{-1/2}(x)u)^i(\mathbf{E}^{-1/2}(x)u)^j \det(\mathbf{E}^{-1/2}(x))\mathrm{d}u \\
&= \det(\mathbf{E}^{-1/2}(x))(\mathbf{E}^{-1/2}(x))^i_k(\mathbf{E}^{-1/2}(x))^j_l \int_{\{u:\|u\|_2^2<R_x^2\}} u^k u^l \mathrm{d}u \\
&= \det(\mathbf{E}^{-1/2}(x))(\mathbf{E}^{-1/2}(x))^i_k(\mathbf{E}^{-1/2}(x))^j_l \frac{R_x^{m+2}}{m+2}\mathrm{Vol}_{\mathrm{Euc}}(\mathbb{B}^m)\delta^{kl} \\
&= \frac{R_x^{m+2}}{m+2}\mathrm{Vol}_{\mathrm{Euc}}(\mathbb{B}^m)\mathbf{E}^{ij}(x)\sqrt{\det \mathbf{E}^{-1}(x)} \\
&= \frac{R_x^2}{m+2}\lambda_x(\mathscr{B}_x)\mathbf{E}^{ij}(x).
\end{aligned}$$

where we have that $\int_{\{u:\|u\|_2^2<R_x^2\}} u^k u^l \mathrm{d}u = \frac{R_x^{m+2}}{m+2}\mathrm{Vol}_{\mathrm{Euc}}(\mathbb{B}^m)\delta^{kl}$ by rotation invariance of the Euclidean ball.

So we have that:

$$\mathbf{S}^{ij}(x) = \frac{\int_{\mathscr{B}_x} v^i v^j \mathrm{d}\lambda_x(v)}{\lambda_x(\mathscr{B}_x)} = \frac{\int_{\mathscr{B}_x} (v-v_0)^i (v-v_0)^j \mathrm{d}\lambda_x(v) + \lambda_x(\mathscr{B}_x) v_0^i v_0^j}{\lambda_x(\mathscr{B}_x)}$$
$$= \frac{R_x^2}{m+2}\mathbf{E}^{ij}(x) + v_0^i v_0^j.$$

□

Remark 44. *We can see that in the Randers setting, contrary to the Riemannian setting, the centroid of the unit ball is not zero but is shifted proportionally to the vector field b. Similarly, the second moment of the unit ball is not proportional to the inverse of the matrix $\mathbf{A}$ but has an additional term proportional to the outer product of the centroid with itself. When $b \to 0$, we recover the Riemannian setting with $\mathbf{c}^i(x) \to 0$ and $\mathbf{S}^{ij}(x) \to \frac{1}{m+2}\mathbf{A}^{ij}(x)$.*

Proposition 45. *In the Randers setting, the moments of the kernel m_0, $\tilde{m}_1$ and $\tilde{m}_2$ are given by:*

$$m_0(x) = m\mu_0 \mathrm{Vol}_{\mathrm{Euc}}(\mathbb{B}^m), \tag{40}$$

$$\tilde{m}_1^i(x) = -\frac{(m+1)\mu_1}{m\mu_0} \frac{b^i(x)}{1 - \|b(x)\|^2_{\mathbf{A}^{-1}}}, \tag{41}$$

$$\tilde{m}_2^{ij}(x) = \frac{\mu_2}{\mu_0}\frac{m+2}{m}\left(\frac{R_x^2}{m+2}\mathbf{E}^{ij}(x) + \mathbf{c}(x)^i \mathbf{c}(x)^j\right). \tag{42}$$

Proof. This is a direct application of Proposition 43 and the expressions of the moments of the kernel in Proposition 35. □

D.8 The moment-Randers approximation

This subsection proves the results of Section 5. We proceed in three steps: we first check that the moment-Randers metric of Definition 5 is well defined, identify its unit ball, and show that it matches the centroid and the second moment of $\mathscr{F}$; we then show how $\mathbf{c}$ and g_{BL} are recovered from $\mathscr{L}^s$ and $\mathscr{L}^a$ through an embedding; and we finally relate the resulting empirical metric $\hat{\mathscr{F}}_R$ to $\mathscr{F}_R$.

Throughout, we work in a chart, so that $\mathbf{c}(x) \in \mathbb{R}^m$, $\mathbf{S}(x) \in \mathbb{R}^{m\times m}$ and $\mathscr{B}_x \subset \mathbb{R}^m$. We recall from Equations (10), (22) and (23) that

$$\mathbf{c}(x) = \frac{1}{\lambda_x(\mathscr{B}_x)}\int_{\mathscr{B}_x} v \mathrm{d}\lambda_x(v), \qquad \mathbf{S}(x) = \frac{1}{\lambda_x(\mathscr{B}_x)}\int_{\mathscr{B}_x} vv^\mathsf{T} \mathrm{d}\lambda_x(v), \qquad g_{\mathrm{BL}}(x)^{-1} = (m+2)\mathbf{S}(x),$$

so that the matrix $\mathbf{H}$ of Definition 5 can be written in the two equivalent ways

$$\mathbf{H}(x)^{-1} = g_{\mathrm{BL}}(x)^{-1} - (m+2)\mathbf{c}(x)\mathbf{c}(x)^\mathsf{T} = (m+2)\big(\mathbf{S}(x) - \mathbf{c}(x)\mathbf{c}(x)^\mathsf{T}\big).$$

Well-definedness of the moment-Randers metric

We begin by proving Proposition 6, which gives a necessary and sufficient condition for $\mathbf{H}$ to be positive definite, and relates the norms of $\mathbf{c}$ in the two metrics. We restate it here for convenience:

Proposition 6. *Let $\mathscr{F}$ be a Finsler metric and let $x \in \mathscr{M}$ be such that $\|\mathbf{c}(x)\|^2_{g_{\mathrm{BL}}} < (m+2)^{-1}$. Then the matrix $\mathbf{H}(x)$ is well defined and positive definite, and*

$$\|\mathbf{c}(x)\|^2_{g_{\mathrm{BL}}} = \frac{\|\mathbf{c}(x)\|^2_{\mathbf{H}}}{1 + (m+2)\|\mathbf{c}(x)\|^2_{\mathbf{H}}}, \qquad \textit{hence} \qquad \|\mathbf{c}(x)\|_{\mathbf{H}} < 1 \iff \|\mathbf{c}(x)\|^2_{g_{\mathrm{BL}}} < (m+3)^{-1}.$$

Proof of Proposition 6. We first check that $\mathbf{H}(x)$ is positive definite, hence invertible. The matrix $\mathbf{H}^{-1}$ is a rank-one downdate of g_{BL}^{-1}, so its positivity only depends on the size of $\mathbf{c}$: for any $u \in T_x^*\mathscr{M}$, the Cauchy–Schwarz inequality gives $\langle \mathbf{c}, u\rangle^2 \le \|\mathbf{c}\|_{g_{\mathrm{BL}}}^2 \|u\|_{g_{\mathrm{BL}}^{-1}}^2$, so that

$$u^{\mathsf{T}}\mathbf{H}^{-1}u = \|u\|_{g_{\mathrm{BL}}^{-1}}^2 - (m+2)\langle \mathbf{c}, u\rangle^2 \ge \left(1-(m+2)\|\mathbf{c}\|_{g_{\mathrm{BL}}}^2\right)\|u\|_{g_{\mathrm{BL}}^{-1}}^2,$$

which is positive for every $u \neq 0$ under the hypothesis. The bound is attained at $u = g_{\mathrm{BL}}\mathbf{c}$, the equality case of Cauchy–Schwarz, where $u^{\mathsf{T}}\mathbf{H}^{-1}u = \|\mathbf{c}\|_{g_{\mathrm{BL}}}^2\left(1-(m+2)\|\mathbf{c}\|_{g_{\mathrm{BL}}}^2\right)$, so the hypothesis is also necessary when $\mathbf{c} \neq 0$.

We now turn to the identity. We recall that $\mathbf{H}^{-1} = (m+2)(\mathbf{S} - \mathbf{c}\mathbf{c}^{\mathsf{T}})$, so that $\mathbf{S} = \frac{1}{m+2}\mathbf{H}^{-1} + \mathbf{c}\mathbf{c}^{\mathsf{T}}$. Setting $A = \frac{1}{m+2}\mathbf{H}^{-1}$, so that $A^{-1} = (m+2)\mathbf{H}$, the Sherman–Morrison formula yields

$$\mathbf{S}^{-1} = (A + \mathbf{c}\mathbf{c}^{\mathsf{T}})^{-1} = A^{-1} - \frac{A^{-1}\mathbf{c}\mathbf{c}^{\mathsf{T}}A^{-1}}{1+\mathbf{c}^{\mathsf{T}}A^{-1}\mathbf{c}} = (m+2)\mathbf{H} - \frac{(m+2)^2\mathbf{H}\mathbf{c}\mathbf{c}^{\mathsf{T}}\mathbf{H}}{1+(m+2)\mathbf{c}^{\mathsf{T}}\mathbf{H}\mathbf{c}}.$$

Hence

$$\mathbf{c}^{\mathsf{T}}\mathbf{S}^{-1}\mathbf{c} = (m+2)\|\mathbf{c}\|_{\mathbf{H}}^2 - \frac{(m+2)^2\|\mathbf{c}\|_{\mathbf{H}}^4}{1+(m+2)\|\mathbf{c}\|_{\mathbf{H}}^2} = \frac{(m+2)\|\mathbf{c}\|_{\mathbf{H}}^2}{1+(m+2)\|\mathbf{c}\|_{\mathbf{H}}^2},$$

and, because $g_{\mathrm{BL}}^{-1} = (m+2)\mathbf{S}$,

$$\|\mathbf{c}\|_{g_{\mathrm{BL}}}^2 = \mathbf{c}^{\mathsf{T}}g_{\mathrm{BL}}\mathbf{c} = \frac{1}{m+2}\mathbf{c}^{\mathsf{T}}\mathbf{S}^{-1}\mathbf{c} = \frac{\|\mathbf{c}\|_{\mathbf{H}}^2}{1+(m+2)\|\mathbf{c}\|_{\mathbf{H}}^2}.$$

The right-hand side is an increasing function of $\|\mathbf{c}\|_{\mathbf{H}}^2$, and

$$\|\mathbf{c}\|_{g_{\mathrm{BL}}}^2 < \frac{1}{m+3} \iff (m+3)\|\mathbf{c}\|_{\mathbf{H}}^2 < 1+(m+2)\|\mathbf{c}\|_{\mathbf{H}}^2 \iff \|\mathbf{c}\|_{\mathbf{H}}^2 < 1.$$

This concludes the proof of Proposition 6. □

The metric of Definition 5 is written in the $\alpha+\beta$ form Equation (1). The following remark shows that it can equivalently be written in the *navigation form*.

Remark 46 (Equivalent navigation form)**.** *With the $\alpha+\beta$ form of Definition 5, we have $\mathbf{A}_R = \lambda^{-1}\mathbf{H} + \lambda^{-2}(\mathbf{H}\mathbf{c})(\mathbf{H}\mathbf{c})^{\mathsf{T}}$ and $b_R = -\lambda^{-1}\mathbf{H}\mathbf{c}$, where $\lambda = 1-\|\mathbf{c}\|_{\mathbf{H}}^2$. Hence*

$$v^{\mathsf{T}}\mathbf{A}_R v = \frac{\lambda\|v\|_{\mathbf{H}}^2 + \langle \mathbf{c}, v\rangle_{\mathbf{H}}^2}{\lambda^2}, \qquad b_R^{\mathsf{T}}v = -\frac{\langle \mathbf{c}, v\rangle_{\mathbf{H}}}{\lambda},$$

so that $\mathscr{F}_R$ can equivalently be written as

$$\mathscr{F}_R(x,v) = \frac{1}{\lambda(x)}\left(\sqrt{\lambda(x)\|v\|_{\mathbf{H}}^2 + \langle \mathbf{c}(x), v\rangle_{\mathbf{H}}^2} - \langle \mathbf{c}(x), v\rangle_{\mathbf{H}}\right), \qquad \lambda(x) = 1-\|\mathbf{c}(x)\|_{\mathbf{H}}^2. \tag{43}$$

Importantly, the admissibility condition $\|b_R(x)\|_{\mathbf{A}_R^{-1}} < 1$ is equivalent to $\|\mathbf{c}(x)\|_{\mathbf{H}} < 1$, which is exactly the hypothesis of Definition 5. Indeed,

$$\|b_R\|_{\mathbf{A}_R^{-1}}^2 = b_R^{\mathsf{T}}\mathbf{A}_R^{-1}b_R = \lambda^2\mathbf{c}^{\mathsf{T}}\mathbf{H}^{-1}\left(\lambda(\mathbf{H}^{-1} - \mathbf{c}\mathbf{c}^{\mathsf{T}})\right)\mathbf{H}^{-1}\mathbf{c} = \lambda^2(\|\mathbf{c}\|_{\mathbf{H}}^2 - \|\mathbf{c}\|_{\mathbf{H}}^4)/\lambda^2 = \|\mathbf{c}\|_{\mathbf{H}}^2.$$

We can now justify the name: $\mathscr{F}_R$ is the unique Randers metric whose unit ball has the same centroid and the same second moment as $\mathscr{B}_x$, and it therefore leaves $\tilde{m}_1$ and $\tilde{m}_2$ unchanged.

Proposition 47 (Moment matching)**.** *Let $\mathscr{F}$ be a Finsler metric with $\|\mathbf{c}(x)\|_{\mathbf{H}} < 1$ for all $x \in \mathscr{M}$, and let $\mathscr{F}_R$ be its moment-Randers metric (Definition 5). Then $\mathscr{F}_R$ is a Randers metric and:*

1. *its unit tangent ball is the ellipsoid $E_x = \{v \in T_x\mathscr{M} : \|v - \mathbf{c}(x)\|_{\mathbf{H}} \leq 1\}$;*
2. *$\mathscr{F}$ and $\mathscr{F}_R$ have the same centroid and the same raw second moment, $\mathbf{c}_R = \mathbf{c}$ and $\mathbf{S}_R = \mathbf{S}$; consequently they have the same Binet–Legendre metric and the same normalized kernel moments $\tilde{m}_1$ and $\tilde{m}_2$;*
3. *if $\mathscr{F}$ is itself a Randers metric, then $\mathscr{F}_R = \mathscr{F}$.*

Proof. Unit tangent ball. We start by showing that $\mathscr{F}_R$ is a valid Randers metric. If $\|\mathbf{c}(x)\|_{\mathbf{H}} < 1$, then $\lambda(x) > 0$. Consequently we have that $\mathscr{F}_R(x, v) \geq 0$ for all $v \in T_x\mathscr{M}$. Now if we assume that $\mathscr{F}_R(x, v) = 0$ for some $v \in T_x\mathscr{M}$, then we have that $\lambda(x)\|v\|_{\mathbf{H}}^2 + \langle \mathbf{c}(x), v\rangle_{\mathbf{H}}^2 = \langle \mathbf{c}(x), v\rangle_{\mathbf{H}}^2$, which implies that $\lambda(x)\|v\|_{\mathbf{H}}^2 = 0$. Since $\lambda(x) > 0$, we have that $\|v\|_{\mathbf{H}}^2 = 0$, which implies that $v = 0$. Therefore, $\mathscr{F}_R(x, v) = 0$ if and only if $v = 0$. Positive homogeneity and the triangle inequality are trivial. Hence, $\mathscr{F}_R$ is a valid Randers metric.

Now let $v \in T_x\mathscr{M}$. We have the following chain of equivalences:

$$
\begin{aligned}
\mathscr{F}_R(x, v) \leq 1 &\iff \sqrt{\lambda(x)\|v\|_{\mathbf{H}}^2 + \langle \mathbf{c}(x), v\rangle_{\mathbf{H}}^2} - \langle \mathbf{c}(x), v\rangle_{\mathbf{H}} \leq \lambda(x) \\
&\iff \sqrt{\lambda(x)\|v\|_{\mathbf{H}}^2 + \langle \mathbf{c}(x), v\rangle_{\mathbf{H}}^2} \leq \lambda(x) + \langle \mathbf{c}(x), v\rangle_{\mathbf{H}} \\
&\iff \lambda(x)\|v\|_{\mathbf{H}}^2 + \langle \mathbf{c}(x), v\rangle_{\mathbf{H}}^2 \leq (\lambda(x) + \langle \mathbf{c}(x), v\rangle_{\mathbf{H}})^2 \\
&\iff \lambda(x)\|v\|_{\mathbf{H}}^2 + \langle \mathbf{c}(x), v\rangle_{\mathbf{H}}^2 \leq \lambda(x)^2 + 2\lambda(x)\langle \mathbf{c}(x), v\rangle_{\mathbf{H}} + \langle \mathbf{c}(x), v\rangle_{\mathbf{H}}^2 \\
&\iff \lambda(x)\|v\|_{\mathbf{H}}^2 \leq \lambda(x)^2 + 2\lambda(x)\langle \mathbf{c}(x), v\rangle_{\mathbf{H}} \\
&\iff \|v\|_{\mathbf{H}}^2 \leq \lambda(x) + 2\langle \mathbf{c}(x), v\rangle_{\mathbf{H}} \\
&\iff \|v - \mathbf{c}(x) + \mathbf{c}(x)\|_{\mathbf{H}}^2 \leq \lambda(x) + 2\langle \mathbf{c}(x), v\rangle_{\mathbf{H}} \\
&\iff \|v - \mathbf{c}(x)\|_{\mathbf{H}}^2 + 2\langle v - \mathbf{c}(x), \mathbf{c}(x)\rangle_{\mathbf{H}} + \|\mathbf{c}(x)\|_{\mathbf{H}}^2 \leq 1 - \|\mathbf{c}(x)\|_{\mathbf{H}}^2 + 2\langle \mathbf{c}(x), v\rangle_{\mathbf{H}} \\
&\iff \|v - \mathbf{c}(x)\|_{\mathbf{H}}^2 + 2\langle v, \mathbf{c}(x)\rangle_{\mathbf{H}} - 2\|\mathbf{c}(x)\|_{\mathbf{H}}^2 + \|\mathbf{c}(x)\|_{\mathbf{H}}^2 \leq 1 - \|\mathbf{c}(x)\|_{\mathbf{H}}^2 + 2\langle \mathbf{c}(x), v\rangle_{\mathbf{H}} \\
&\iff \|v - \mathbf{c}(x)\|_{\mathbf{H}}^2 \leq 1
\end{aligned}
$$

Moment matching. By the first item above, the unit tangent ball of $\mathscr{F}_R$ is $E_x = \{v \in T_x\mathscr{M} : \|v - \mathbf{c}(x)\|_{\mathbf{H}} \leq 1\}$. Hence its centroid is $\mathbf{c}(x)$.

We now compute the second moment of the unit tangent ball of the moment-Randers metric:

$$
\begin{aligned}
\mathbf{S}_R(x) &= \frac{1}{\lambda_x(E_x)} \int_{E_x} v v^{\mathsf{T}} \mathrm{d}\lambda_x(v) \\
&= \frac{1}{\lambda_x(E_x)} \int_{E_x} (v - \mathbf{c}(x) + \mathbf{c}(x))(v - \mathbf{c}(x) + \mathbf{c}(x))^{\mathsf{T}} \mathrm{d}\lambda_x(v) \\
&= \frac{1}{\lambda_x(E_x)} \int_{E_x} (v - \mathbf{c}(x))(v - \mathbf{c}(x))^{\mathsf{T}} \mathrm{d}\lambda_x(v) + \mathbf{c}(x)\mathbf{c}(x)^{\mathsf{T}},
\end{aligned}
$$

where we used that $\int_{E_x} (v - \mathbf{c}(x)) \mathrm{d}\lambda_x(v) = 0$. Now, as E_x is an ellipsoid, we have that $\int_{E_x} (v - \mathbf{c}(x))(v - \mathbf{c}(x))^{\mathsf{T}} \mathrm{d}\lambda_x(v) = \frac{1}{m+2}\mathbf{H}(x)^{-1}$, and thus

$$
\begin{aligned}
\mathbf{S}_R(x) &= \frac{1}{m+2}\mathbf{H}(x)^{-1} + \mathbf{c}(x)\mathbf{c}(x)^{\mathsf{T}} \\
&= \mathbf{S}(x)
\end{aligned}
$$

where we used that $\mathbf{H}(x)^{-1} = (m+2)(\mathbf{S}(x) - \mathbf{c}(x)\mathbf{c}(x)^\mathsf{T})$.

Uniqueness. Since $\tilde{\mathscr{F}}(x,\cdot)$ is positively homogeneous of degree 1 for any Finsler metric $\tilde{\mathscr{F}}$, we have $\tilde{\mathscr{F}}(x,v/t) = \tilde{\mathscr{F}}(x,v)/t$ for $t>0$, so that $v/t \in \{\tilde{\mathscr{F}}(x,\cdot) \le 1\} \iff \tilde{\mathscr{F}}(x,v) \le t$, and thus

$$\tilde{\mathscr{F}}(x,v) = \inf\{t>0 : v/t \in \{\tilde{\mathscr{F}}(x,\cdot) \le 1\}\}, \qquad \forall v \in T_x\mathscr{M}.$$

In particular, a Finsler metric is determined by its unit tangent balls: two Finsler metrics with the same unit tangent ball at every $x \in \mathscr{M}$ are equal. The goal of this paragraph is to show that, for a Randers metric, the unit tangent ball is uniquely determined by its centroid and its raw second moment.

By definition, the centroid $\mathbf{c}$ and the raw second moment $\mathbf{S}$ of a Randers metric are given by

$$\mathbf{c} = \frac{1}{\lambda_x(\mathscr{B}_x)} \int_{\mathscr{B}_x} v \mathrm{d}\lambda_x(v), \qquad \mathbf{S} = \frac{1}{\lambda_x(\mathscr{B}_x)} \int_{\mathscr{B}_x} vv^\mathsf{T} \mathrm{d}\lambda_x(v),$$

where $\mathscr{B}_x = \{v \in T_x\mathscr{M} : \mathscr{F}(x,v) \le 1\}$ is the unit tangent ball of the Randers metric. By Proposition 43, we have that $\mathscr{B}_x$ is an ellipsoid, and by Proposition 45, we have that $\mathbf{c}$ and $\mathbf{S}$ are given by explicit formulas in terms of the parameters $\mathbf{A}$ and b of the Randers metric. These formulas are invertible, so that $\mathbf{A}$ and b can be recovered from $\mathbf{c}$ and $\mathbf{S}$. Hence the unit tangent ball $\mathscr{B}_x$ is uniquely determined by $\mathbf{c}$ and $\mathbf{S}$, and so is the Randers metric $\mathscr{F}$. □

Recovering the centroid and the Binet–Legendre metric from an embedding

Let $\Psi = (\psi_1, \dots, \psi_\ell) : \mathscr{M} \to \mathbb{R}^\ell$ be a smooth embedding, and denote by $J(x) \in \mathbb{R}^{\ell \times m}$ its Jacobian at x, whose k-th row is $\nabla\psi_k(x)^\mathsf{T}$. We assume throughout that $J(x)$ has full row rank m, so that $\ell \ge m$ and $J(x)^+ = (J(x)^\mathsf{T}J(x))^{-1}J(x)^\mathsf{T}$ is the pseudo-inverse of $J(x)$, i.e. $J(x)^+J(x) = I_m$. Operators are applied to Ψ coordinate-wise, so that $\mathscr{L}^a[\Psi] = (\mathscr{L}^a[\psi_1], \dots, \mathscr{L}^a[\psi_\ell])^\mathsf{T}$. Recall from Proposition 35 and Theorem 2 that the anti-symmetric operator $\mathscr{L}^a$ is given by:

$$\mathscr{L}^a[f](x) = \langle \tilde{m}_1(x), \nabla f(x) \rangle = c_1 \langle \mathbf{c}(x), \nabla f(x) \rangle,$$

where $\mathbf{c}(x)$ is the centroid of the unit ball $\mathscr{B}_x$ and $c_1 = ((m+1)\mu_1)/(m\mu_0)$ is a constant that depends on the kernel. Applying $\mathscr{L}^a$ to the embedding Ψ yields:

$$V(x) := \mathscr{L}^a[\Psi](x) = c_1 \begin{pmatrix} \langle \mathbf{c}(x), \nabla\psi_1(x) \rangle \\ \vdots \\ \langle \mathbf{c}(x), \nabla\psi_\ell(x) \rangle \end{pmatrix} = c_1 J(x)\mathbf{c}(x).$$

Consequently, we can recover the centroid $\mathbf{c}(x)$ from $\mathscr{L}^a\Psi(x)$ by

$$\mathbf{c}(x) = \frac{1}{c_1} J(x)^+ \mathscr{L}^a\Psi(x).$$

The Binet–Legendre metric is recovered in the same way, from the symmetric part, through the carré du champ of $\mathscr{L}^s$. The point is that this bilinear form only retains the principal part of $\mathscr{L}^s$: the sampling density ρ and the normalization θ, which enter $\mathscr{L}^s$ only through first-order terms, disappear. This is the content of Proposition 7, which we now prove.

Proof of Proposition 7. By Theorem 2, in a chart $\mathscr{L}^s$ is a second-order differential operator with principal part $c_2 g_{\mathrm{BL}}^{ij} \partial_i\partial_j$ and no zeroth-order term. We denote by ξ the coefficients collecting all first-order contributions, so that $\mathscr{L}^s = c_2 g_{\mathrm{BL}}^{ij} \partial_i\partial_j + \xi^i \partial_i$. For smooth f and h, the Leibniz rule gives $\partial_i(fh) = f\partial_i h + h\partial_i f$ and

$$\partial_i\partial_j(fh) = f\partial_i\partial_j h + h\partial_i\partial_j f + \partial_i f\, \partial_j h + \partial_j f\, \partial_i h,$$

so that all terms involving ξ and all second derivatives of f and h cancel in the combination $\mathscr{L}^s[fh] - f\mathscr{L}^s[h] - h\mathscr{L}^s[f]$. Using the symmetry of g_{BL}^{ij}, we are left with

$$\frac{1}{2}\Big(\mathscr{L}^s[fh] - f\mathscr{L}^s[h] - h\mathscr{L}^s[f]\Big) = c_2\, g_{\mathrm{BL}}^{ij}\, \partial_i f\, \partial_j h.$$

Applying this with $f = \psi_k$ and $h = \psi_l$ gives $\Gamma^{kl} = c_2\, g_{\mathrm{BL}}^{ij} \partial_i \psi_k \partial_j \psi_l$, which is the (k,l) entry of $c_2 J g_{\mathrm{BL}}^{-1} J^{\mathsf{T}}$, with $c_2 = \frac{\mu_2}{2m\mu_0}$. □

Proposition 7 gives the following corollary, which allows us to compute the norm of the drift in the Binet–Legendre metric from the embedding coordinates $\Psi(x)$ and the operators $\mathscr{L}^s$ and $\mathscr{L}^a$.

Corollary 48. *If $J(x)$ has full column rank, then*

$$\Gamma(x)^+ = \frac{1}{c_2} J(x)^{+\mathsf{T}} g_{\mathrm{BL}}(x) J(x)^+, \qquad \textit{and} \qquad \|\mathbf{c}(x)\|_{g_{\mathrm{BL}}}^2 = \frac{c_2}{c_1^2}\|V(x)\|_{\Gamma^+}^2.$$

Consequently, by Proposition 6, the moment-Randers metric of $\mathscr{F}$ is well defined at x if and only if $\frac{c_2}{c_1^2}\|V(x)\|_{\Gamma^+}^2 < (m+3)^{-1}$, which is computable from $\mathscr{L}^s$ and $\mathscr{L}^a$ alone.

Proof. Fix $x \in \mathscr{M}$, write $M(x) = c_2^{-1} J(x)^{+\mathsf{T}} g_{\mathrm{BL}}(x) J(x)^+$, and let $P(x) = J(x)J(x)^+$ be the orthogonal projector onto the range of $J(x)$[8]; both $\Gamma(x)$ and $M(x)$ are symmetric. Using Proposition 7 and $J(x)^+ J(x) = I_m$,

$$\Gamma(x)M(x) = J(x) g_{\mathrm{BL}}(x)^{-1} \big(J(x)^+ J(x)\big)^{\mathsf{T}} g_{\mathrm{BL}}(x) J(x)^+ = J(x)J(x)^+ = P(x).$$

Since $P(x)$ is symmetric, we also have $M(x)\Gamma(x) = P(x)$. Applying this, it holds that:

$$\Gamma(x)M(x)\Gamma(x) = \Gamma(x)P(x) = \Gamma(x), \qquad M(x)\Gamma(x)M(x) = M(x)P(x) = M(x).$$

Together with the symmetry of $\Gamma(x)M(x)$ and $M(x)\Gamma(x)$, these are exactly the Penrose conditions characterizing the Moore–Penrose pseudoinverse, so $M(x) = \Gamma(x)^+$. The second identity then follows from $V(x) = c_1 J(x)\mathbf{c}(x)$ and the definition of the pseudo-inverse:

$$\|\mathbf{c}(x)\|_{g_{\mathrm{BL}}}^2 = \frac{1}{c_1^2} V(x)^{\mathsf{T}} J(x)^{+\mathsf{T}} g_{\mathrm{BL}}(x) J(x)^+ V(x) = \frac{c_2}{c_1^2} V(x)^{\mathsf{T}} \Gamma(x)^+ V(x) = \frac{c_2}{c_1^2}\|V(x)\|_{\Gamma^+}^2. \qquad \square$$

To sum up, we can recover both the direction and the magnitude, in the Binet–Legendre metric, of the vector field $\mathbf{c}(x)$ from the vector field $V(x)$ in the embedding space. Note that neither the sampling density ρ nor the normalization exponent θ appears in $\|\mathbf{c}(x)\|_{g_{\mathrm{BL}}}^2$.

The empirical moment-Randers metric

Building on the above, we can now define a metric from the observed quantities $\Gamma(x)$ and $V(x)$. Following Section 5, set

$$\hat{g}_{BL}(x)^{-1} = \Gamma(x), \qquad \hat{\mathbf{c}}(x) = \frac{\sqrt{c_2}}{c_1} V(x), \tag{44}$$

$$\hat{\mathbf{H}}(x)^{-1} = \Gamma(x) - (m+2)\frac{c_2}{c_1^2} V(x)V(x)^{\mathsf{T}}, \qquad \hat{\lambda}(x) = 1 - \|\hat{\mathbf{c}}(x)\|_{\hat{\mathbf{H}}}^2,$$

[8] Since $J(x)$ has full column rank, $J(x)^+ = \big(J(x)^{\mathsf{T}}J(x)\big)^{-1}J(x)^{\mathsf{T}}$, so $P(x) = J(x)\big(J(x)^{\mathsf{T}}J(x)\big)^{-1}J(x)^{\mathsf{T}}$ is symmetric.

with $\hat{\mathbf{H}} := (\hat{\mathbf{H}}(x)^{-1})^{+}$. Let $\hat{\mathcal{F}}_R$ be defined by Definition 5 with $(\hat{\mathbf{c}}, \hat{\mathbf{H}}, \hat{\lambda})$, so that $\hat{\mathcal{F}}_R$ is defined as:

$$\hat{\mathcal{F}}_R(x,v) = \frac{1}{\hat{\lambda}(x)} \left(\sqrt{\hat{\lambda}(x)\|v\|^2_{\hat{\mathbf{H}}} + \langle \hat{\mathbf{c}}(x), v\rangle^2_{\hat{\mathbf{H}}}} - \langle \hat{\mathbf{c}}(x), v\rangle_{\hat{\mathbf{H}}} \right).$$

Lemma 49. *If $J(x)$ has full column rank, then $\hat{\mathbf{H}}^{-1}$ and $\hat{\mathbf{c}}$ are the pushforwards of $\mathbf{H}^{-1}$ and $\mathbf{c}$ by the Jacobian $J(x)$, i.e.*

$$\hat{\mathbf{H}}^{-1}(x) = c_2 J(x)\mathbf{H}^{-1}(x)J(x)^{\mathsf{T}}, \qquad \hat{\mathbf{c}}(x) = \sqrt{c_2} J(x)\mathbf{c}(x), \qquad \hat{\mathbf{H}}(x) = \frac{1}{c_2} J(x)^{+\mathsf{T}}\mathbf{H}(x)J(x)^{+},$$

and $\hat{\lambda}(x) = \lambda(x)$.

In particular, this lemma implies that $\|\hat{\mathbf{c}}(x)\|_{\hat{\mathbf{H}}} = \|\mathbf{c}(x)\|_{\mathbf{H}}$, so that the admissibility condition $\|\hat{\mathbf{c}}(x)\|_{\hat{\mathbf{H}}} < 1$ is equivalent to $\|\mathbf{c}(x)\|_{\mathbf{H}} < 1$, which is exactly the hypothesis of Definition 5. Hence, if $J(x)$ has full column rank, then $\hat{\mathcal{F}}_R$ is a valid Randers metric on the embedded tangent space $J(T_x\mathcal{M})$.

Proof. By Proposition 7, $\Gamma(x) = c_2 J(x) g_{\mathrm{BL}}(x)^{-1} J(x)^{\mathsf{T}}$, and $V(x)V(x)^{\mathsf{T}} = c_1^2 J(x)\mathbf{c}(x)\mathbf{c}(x)^{\mathsf{T}}J(x)^{\mathsf{T}}$ since $V(x) = c_1 J(x)\mathbf{c}(x)$. Substituting into the definition of $\hat{\mathbf{H}}^{-1}(x)$ and using $\mathbf{H}(x)^{-1} = g_{\mathrm{BL}}(x)^{-1} - (m+2)\mathbf{c}(x)\mathbf{c}(x)^{\mathsf{T}}$ (Definition 5),

$$\begin{aligned}
\hat{\mathbf{H}}^{-1}(x) &= c_2 J(x) g_{\mathrm{BL}}(x)^{-1} J(x)^{\mathsf{T}} - (m+2) c_2 J(x)\mathbf{c}(x)\mathbf{c}(x)^{\mathsf{T}} J(x)^{\mathsf{T}} \\
&= c_2 J(x) \left(g_{\mathrm{BL}}(x)^{-1} - (m+2)\mathbf{c}(x)\mathbf{c}(x)^{\mathsf{T}} \right) J(x)^{\mathsf{T}} \\
&= c_2 J(x)\mathbf{H}(x)^{-1} J(x)^{\mathsf{T}},
\end{aligned}$$

and, similarly, $\hat{\mathbf{c}}(x) = \frac{\sqrt{c_2}}{c_1} V(x) = \frac{\sqrt{c_2}}{c_1} c_1 J(x)\mathbf{c}(x) = \sqrt{c_2} J(x)\mathbf{c}(x)$.

For $\hat{\mathbf{H}}(x)$: the identity $\hat{\mathbf{H}}^{-1}(x) = c_2 J(x)\mathbf{H}(x)^{-1}J(x)^{\mathsf{T}}$ just established has exactly the same form as $\Gamma(x) = c_2 J(x) g_{\mathrm{BL}}(x)^{-1} J(x)^{\mathsf{T}}$, with $g_{\mathrm{BL}}(x)^{-1}$ replaced by $\mathbf{H}(x)^{-1}$. Since $\mathbf{H}(x)^{-1} \succ 0$ (Proposition 6) and $J(x)$ has full column rank, the proof of Corollary 48 applies verbatim under this substitution, giving $\hat{\mathbf{H}}(x) = \left(\hat{\mathbf{H}}^{-1}(x)\right)^{+} = c_2^{-1} J(x)^{+\mathsf{T}}\mathbf{H}(x)J(x)^{+}$.

It remains to compare the two norms. Using $J(x)^{+}J(x) = I_m$,

$$\|\hat{\mathbf{c}}(x)\|^2_{\hat{\mathbf{H}}} = c_2\, \mathbf{c}(x)^{\mathsf{T}} J(x)^{\mathsf{T}} \cdot \frac{1}{c_2} J(x)^{+\mathsf{T}}\mathbf{H}(x)J(x)^{+} \cdot J(x)\mathbf{c}(x) = \mathbf{c}(x)^{\mathsf{T}}\mathbf{H}(x)\mathbf{c}(x) = \|\mathbf{c}(x)\|^2_{\mathbf{H}},$$

so that $\hat{\lambda}(x) = 1 - \|\hat{\mathbf{c}}(x)\|^2_{\hat{\mathbf{H}}} = 1 - \|\mathbf{c}(x)\|^2_{\mathbf{H}} = \lambda(x)$. □

Linking the two Randers metrics $\mathcal{F}_R$ and $\hat{\mathcal{F}}_R$ together results in Proposition 9, which we restate here for convenience:

Proposition 9. *If $\|\hat{\mathbf{c}}(x)\|_{\hat{g}_{BL}} < 1/\sqrt{m+3}$ and J is of full column rank, then $\hat{\mathcal{F}}_R$ is related to the original moment-Randers metric $\mathcal{F}_R$ by*

$$\hat{\mathcal{F}}_R(x, J(x)v) = \frac{1}{\sqrt{c_2}} \mathcal{F}_R(x,v), \quad \forall v \in T_x\mathcal{M}.$$

In particular, $\hat{\mathcal{F}}_R$ is a valid Randers metric on $J(T_x\mathcal{M})$.

Proof of Proposition 9. By Corollary 48 and Proposition 6 the assumption $\|\hat{\mathbf{c}}(x)\|_{\hat{g}_{BL}} < (m+3)^{-1/2}$ is equivalent to $\|\mathbf{c}(x)\|_{\mathbf{H}} < 1$, so that $\mathscr{F}_R$ is a valid Randers metric at x. By Lemma 49 and $J^+J = I_m$, for every $v \in T_x\mathscr{M}$,

$$\|Jv\|^2_{\hat{\mathbf{H}}} = \frac{1}{c_2}\|v\|^2_{\mathbf{H}}, \qquad \langle \hat{\mathbf{c}}, Jv\rangle_{\hat{\mathbf{H}}} = \sqrt{c_2}\,\mathbf{c}^\mathsf{T} J^\mathsf{T} \frac{1}{c_2} J^{+\mathsf{T}} \mathbf{H} J^+ Jv = \frac{1}{\sqrt{c_2}}\langle \mathbf{c}, v\rangle_{\mathbf{H}}.$$

Substituting these and $\hat{\lambda} = \lambda$ into Equation (43),

$$\hat{\mathscr{F}}_R(x, J(x)v) = \frac{1}{\lambda}\left(\sqrt{\frac{\lambda}{c_2}\|v\|^2_{\mathbf{H}} + \frac{1}{c_2}\langle \mathbf{c}, v\rangle^2_{\mathbf{H}}} - \frac{1}{\sqrt{c_2}}\langle \mathbf{c}, v\rangle_{\mathbf{H}}\right) = \frac{1}{\sqrt{c_2}}\mathscr{F}_R(x, v). \qquad \square$$

In other words, $\hat{\mathscr{F}}_R = c_2^{-1/2}\Psi_* \mathscr{F}_R$ on the embedded tangent space $J(x)(T_x\mathscr{M})$: the moment-Randers construction commutes with the embedding, which is what makes it computable. Note that $\hat{\mathscr{F}}_R$ is a Randers metric on $J(x)(T_x\mathscr{M})$ only, and not on the whole of $\mathbb{R}^\ell$, since $\hat{\mathbf{H}}^{-1}$ has rank m.

Discrete setting. All the quantities above are evaluated from the discrete operators of Equation (8) by replacing $\mathscr{L}^s$ and $\mathscr{L}^a$ with $\mathbf{L}^s$ and $\mathbf{L}^a$, and Ψ with the embedded samples $Y_i = \Psi(X_i)$. Writing $\psi_k \in \mathbb{R}^N$ for the k-th embedding coordinate evaluated at the samples and $\odot$ for the entrywise product,

$$V_i = \mathbf{L}^a[\Psi]_i, \qquad \Gamma_i^{kl} = \frac{1}{2}\Big[\big(\mathbf{L}^s(\psi_k \odot \psi_l)\big)_i - \psi_k(X_i)\big(\mathbf{L}^s\psi_l\big)_i - \psi_l(X_i)\big(\mathbf{L}^s\psi_k\big)_i\Big],$$

and $\|\mathbf{c}(X_i)\|^2_{g_{\mathrm{BL}}} = \frac{c_2}{c_1^2} V_i^\mathsf{T} \Gamma_i^+ V_i$ by Corollary 48. Whenever this quantity is below $(m+3)^{-1}$, Equation (44) defines $\hat{\mathbf{c}}_i$, $\hat{\mathbf{H}}_i$ and $\hat{\lambda}_i$, and hence the metric $\hat{\mathscr{F}}_R(X_i, \cdot)$ at the sample X_i. By Theorem 4, $\mathbf{L}^s$ and $\mathbf{L}^a$ converge uniformly and almost surely to $\mathscr{L}^s$ and $\mathscr{L}^a$, so these empirical quantities converge to their continuous counterparts at every sample.

D.9 Computation of the kernel moments

The values of the kernel moments μ_n can be computed analytically for the usual kernel functions K.

Exponential kernel

If K is the exponential kernel $K(x) = \exp(-x)$, then we can compute, for $m \in \mathbb{N}$:

$$\mu_n = \int_0^{+\infty} K(r) r^{n+m-1} \mathrm{d}r = \int_0^{+\infty} e^{-r} r^{n+m-1} \mathrm{d}r = \Gamma(n+m) = (n+m-1)!$$

Gaussian kernel

If K is the Gaussian kernel $K(x) = \exp(-x^2)$, then we can compute, for $m \in \mathbb{N}$, using the change of variable $u = r^2$, so that $\mathrm{d}r = \mathrm{d}u/(2\sqrt{u})$:

$$\begin{aligned}
\mu_n = \int_0^{+\infty} K(r) r^{n+m-1} \mathrm{d}r &= \int_0^{+\infty} e^{-r^2} r^{n+m-1} \mathrm{d}r \\
&= \frac{1}{2}\int_0^{+\infty} e^{-u} u^{(n+m-1)/2 - 1/2} \mathrm{d}u \\
&= \frac{1}{2}\int_0^{+\infty} e^{-u} u^{(n+m)/2 - 1} \mathrm{d}u \\
&= \frac{1}{2}\Gamma\left(\frac{n+m}{2}\right)
\end{aligned}$$