WHAT A POINT SET KNOWS ABOUT A RIEMANNIAN MINIMIZER: A SCALING LAW AT THE CUT LOCUS

Anonymous authors
Paper under double-blind review

ABSTRACT

Quasi-Monte Carlo (QMC) point sets come with integration certificates: a discrepancy bounds the error of every integrand in a function class. Riemannian optimization asks where the minimizer of the sampled loss lies. One object connects the two. The point set minus the target law is a signed measure, the discrepancy charge; through the loss it generates a potential whose gradient, the field, is the error of the sampled gradient. The minimizer error is the field at the minimizer times an inverse Hessian, and reusing a point set, refreshing it, or solving once reads the field in three norms, which gives their sample complexities. What a certificate says about the field passes through one integrand, the gradient of the loss, and on a manifold this integrand is singular at the cut locus. For positive rules, dilation predicts that a singularity of order m on a set of codimension c leaves the exponent $(m+c)/(c+\beta)$ of the discrepancy, where β is the dilation exponent of the certificate norm; we prove this scaling law, and that it cannot be improved, for point singularities on spheres in Sobolev norms and for point and polygonal singularities in the square in Hardy–Krause variation. On spheres the Fréchet problem is electrostatics: the cap discrepancy is the energy of the charge, the value error its potential at the antipode, the gradient error its field there. This gives the sharp exponent $2d/(2d+1)$ for the cap discrepancy on $\mathbb{S}^d$; a floor of order $1/N$ on the worst-case gradient error, which spherical designs meet up to N^{ϵ}; and a chart that removes the singularity when it blows up the cut point. Experiments on positive definite matrices, Grassmannians and spheres test the main predictions.

> *"It is upon the exactness with which we follow phenomena into the infinitely small that our knowledge of their causal relations essentially depends."* (Appendix I explains how this relates to our paper)
> — Bernhard Riemann, *On the Hypotheses which lie at the Bases of Geometry* (1854; trans. Clifford, 1873)

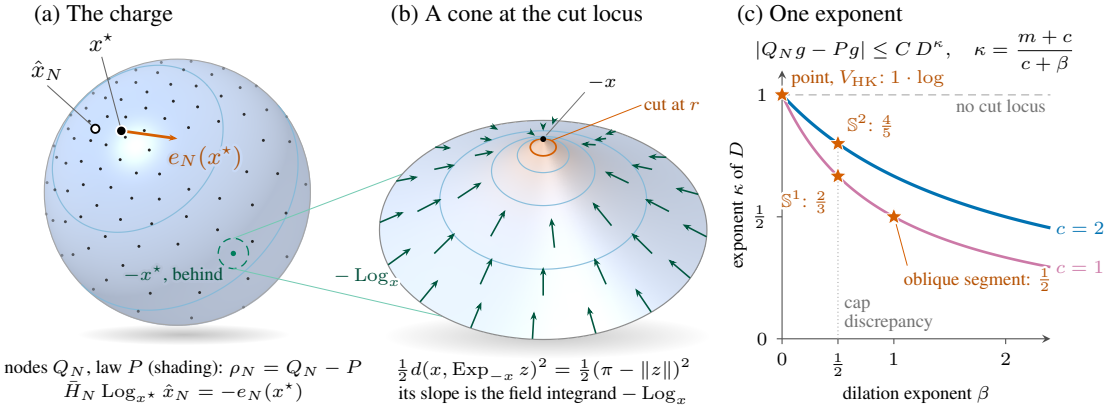

Figure 1: **What a point set knows about a Riemannian minimizer.** (a) Quasi-Monte Carlo nodes for a von Mises–Fisher law on $\mathbb{S}^2$, minus the law (shading), form the discrepancy charge; its field e_N at the minimizer $x^\star$ moves the sample minimizer $\hat{x}_N$ (Proposition 2.2). (b) The loss $\frac{1}{2}d(x,\cdot)^2$ has a cone at the antipode $-x$, the cut locus of x; its slope is the field integrand $-\operatorname{Log}_x$, whose direction jumps at the tip. (c) The exponent of the scaling law (Theorem 3.3) for $m = 0$; stars mark the cases proved sharp (Theorem 3.4).

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.

1 INTRODUCTION

Many learning problems have their parameters on a curved space: a principal subspace on a Grassmannian, a covariance or a metric on the positive definite cone, a mean direction or a normalized embedding on a sphere, a Gaussian posterior on the Bures–Wasserstein manifold (Boumal, 2023; Luu et al., 2025). Their losses are expectations, over data or over injected noise, and are optimized through sample averages. Replacing independent samples by quasi-Monte Carlo (QMC) point sets makes such averages converge faster for smooth integrands, and has sped up Euclidean stochastic optimization and variational inference (Buchholz et al., 2018; Liu & Owen, 2021); whether the advantage survives curvature, and what it buys an optimizer, is our question. The problem is

$$\min_{x\in\mathcal{M}} f(x), \qquad f(x) = \int_\Omega F(x,\omega)\,\mathrm{d}P(\omega). \tag{1}$$

In practice the integral becomes an average over N points, $f_N = Q_N F$ with $Q_N = \frac{1}{N}\sum_i \delta_{\omega_i}$, and QMC places the points so that a certificate holds: $|Q_N g - Pg| \le \|g\|_H \; D_H(Q_N)$ for every g in a function class H, with a discrepancy D_H that decays faster than the $N^{-1/2}$ of independent samples (Dick et al., 2013). We ask a naive question: *what does a point set know about the minimizer of f_N?*

The answer passes through one object. Subtract the target law from the rule and a signed measure of total mass zero remains, the *discrepancy charge* $\rho_N = Q_N - P$. Paired with the loss it defines a *potential* on the decision manifold, $\Phi_N(x) = \rho_N F(x,\cdot) = f_N(x) - f(x)$, and the gradient of the potential, its *field* $e_N = \operatorname{grad}\Phi_N$, is the error of the sampled gradient. Section 2 shows that the minimizer error is the field at the minimizer multiplied by an inverse Hessian; that reusing one point set along an optimization path, refreshing it at every step, and solving once read the field in three norms, which is where their sample complexities come from; and that whatever a certificate says about the field, it says through one tangent-vector-valued integrand, the gradient $G_x = \operatorname{grad}_x F(x,\cdot)$ of the loss in the decision variable.

On a manifold this integrand is singular. For squared-distance losses $G_x = -\operatorname{Log}_x$, whose direction jumps when the data point crosses the cut locus of x; on $\mathbb{S}^2$ its components are not even in H^1, let alone in the Sobolev space $H^{3/2}$ that the cap discrepancy controls there. A certificate does not apply to such an integrand as stated, and the question becomes how much of it survives. Our main result is a scaling law (Section 3). Quadrature rules have positive weights, so a singular integrand can be sandwiched between two members of H and only the gap has to be paid for. Balancing the certificate norm of a sandwich against its gap is a dilation computation, and it leaves one exponent,

$$\frac{m+c}{c+\beta}, \tag{2}$$

for a singularity of order m (its degree of homogeneity: 0 for a gradient that jumps in direction, 1 for a value with a cone) on a set of codimension c, when the certificate norm of a bump of width r grows like $r^{-\beta}$. We prove the law for point singularities on spheres in Sobolev norms and for point singularities and polygonal jumps in the square in Hardy–Krause variation; for cut loci of other codimensions on other manifolds it is a prediction. Measures built on the dual side, multipoles on spheres and shaped discrepancy functions in the square, show that no bound in terms of the discrepancy alone does better, and the field of a single node sets a floor of order $1/N$ on the worst-case gradient error that no point set avoids. Table 1 lists the exponents proved here.

Two geometric observations make the law concrete (Section 4). On the sphere the Fréchet problem is electrostatics: half the squared geodesic distance equals $-\pi$ times the chordal distance to the antipode, up to a term too smooth to matter, so the Fréchet value error is the potential of the discrepancy charge at the antipode, the gradient error is its field there, and Stolarsky's invariance principle says that the cap discrepancy is its energy (Stolarsky, 1973; Brauchart & Dick, 2013). And a chart can remove the singularity: a polar chart is the real blow-up of its poles, on which the jumping direction becomes the smooth angle coordinate, so a chart resolves the cut locus of the points whose antipodes it blows up and leaves a point singularity at the others.

The experiments test these predictions (Section 5). On positive definite matrices the Karcher-mean error of every replicate equals the inverse Hessian applied to the field at the minimizer, to within a relative $3\cdot 10^{-6}$ for scrambled nets at the largest N; on a Grassmannian the sampler and the transport set the error, and reuse of samples costs a constant; rules exact to high degree and nodes optimized

for the energy have maximal gradient errors at the $1/N$ floor; and when the data law, the sampler and the minimizer are fixed and only the chart changes, the fitted slope of the Fréchet-mean error of scrambled nets changes from -1.01 to -1.55. The last two go beyond what is proved: the proved bounds for the two charts differ by one logarithm, and those for optimized nodes stop at $N^{-3/5}$.

Relation to prior work. QMC sample averages are classical in Euclidean stochastic programming (Koivu, 2005; Homem-de Mello, 2008), and Luu et al. (2025) used scrambled minibatches on the Bures–Wasserstein manifold. We use Riemannian stochastic optimization (Bonnabel, 2013; Zhang & Sra, 2016; Boumal et al., 2019), QMC designs on spheres (Brauchart et al., 2014), Koksma–Hlawka inequalities for piecewise smooth integrands (Brandolini et al., 2013) and one-sided approximation (Vaaler, 1985), which on the sphere gives Erdős–Turán bounds for the cap discrepancy (Grabner, 1991; Li & Vaaler, 1999; Grabner & Tichy, 1993). The readings of the field are known arguments organized around one object. New is the analysis at the cut locus: the scaling law in the settings above, with the exponent $2d/(2d+1)$ on spheres, its sharpness and the $1/N$ floor; the electrostatic form of the Fréchet problem on spheres; and the chart as a blow-up, which explains when transported nets keep their rate and prescribes a chart that does. A classical result reappears as a special case: on Hadamard manifolds the contraction of barycentres (Sturm, 2003) is the one-point reading with the Wasserstein certificate. Appendix H discusses further related work.

2 The field of a point set

Setting. $\mathcal{M}$ is a complete Riemannian manifold with distance d, exponential map Exp and logarithm Log. The integration space (Ω, P) is a probability space: the cube $[0,1]^s$ with Lebesgue measure, a sphere $\mathbb{S}^d$ with its normalized surface measure σ, or $\mathcal{M}$ itself. The loss $F\colon \mathcal{M}\times\Omega\to\mathbb{R}$ is C^1 in x with F and $\operatorname{grad}_x F$ jointly continuous (Assumption C.1), so that $\operatorname{grad} f = P\operatorname{grad}_x F$. Rules $Q_N = \frac{1}{N}\sum_i \delta_{\omega_i}$ have equal weights unless stated otherwise; density weights $h(\omega_i)$ are absorbed into the loss. A *certificate* is a class H of functions on Ω with a seminorm that vanishes on constants, and the *discrepancy* of a rule is $D_H(Q) = \sup\{|Qg - Pg| : \|g\|_H \le 1\}$. The Koksma–Hlawka inequality is the case of Hardy–Krause variation V_{HK}, with the star discrepancy D_N^* (Koksma, 1942; Hlawka, 1961); on $\mathbb{S}^d$, the L_2 cap discrepancy D_{L_2} is the case of $H^{(d+1)/2}(\mathbb{S}^d)$ with Stolarsky's norm (Brauchart & Dick, 2013). Appendix B collects these facts. Cost is counted in evaluations of $\operatorname{grad}_x F$.

Definition 2.1 (charge, potential, field)**.** The *discrepancy charge* of Q_N is the signed measure $\rho_N = Q_N - P$. Its *potential* and *field* on $\mathcal{M}$ are

$$\Phi_N(x) = \rho_N F(x,\cdot) = f_N(x) - f(x), \qquad e_N(x) = \operatorname{grad}\Phi_N(x) = \rho_N G_x.$$

We call the map $G_x = \operatorname{grad}_x F(x,\cdot)$ from Ω to $T_x\mathcal{M}$ the *field integrand*, and we put $\|G_x\|_H = \sup_{\|v\|=1}\|\langle G_x, v\rangle\|_H$.

The field at x is one vector of one inner-product space, $T_x\mathcal{M}$, so a certificate applies to it in its own direction, with no frame, no parallel transport between tangent spaces, and no union over the iterates of an optimizer. The norm $\|G_x\|_H$ is the operator norm of the differential of the loss embedding $x \mapsto F(x,\cdot) \in H$.

Proposition 2.2 (three readings of the field; Appendix C)**.** *(a)* Uniform. *For every $S \subseteq \mathcal{M}$, $\sup_{x\in S}\|e_N(x)\| \le V_1(S)\, D_H(Q_N)$, where $V_1(S) = \sup_{x\in S}\|G_x\|_H$.*

(b) Mean square. *If Q_N is random with $\mathbb{E}\, Q_N g = Pg$, then $\mathbb{E}\,\|e_N(x)\|^2 = \sum_a \operatorname{Var}(Q_N\langle G_x, e_a\rangle)$ for every orthonormal basis (e_a) of $T_x\mathcal{M}$.*

(c) One point. *Let $\operatorname{grad} f(x^\star) = 0$ and let $\hat{x}_N$ minimize f_N over a closed geodesically convex set $C \ni x^\star$ on which f_N is μ_0-strongly geodesically convex. Then $\mu_0\, d(\hat{x}_N, x^\star) \le \|e_N(x^\star)\|$. If moreover $\operatorname{grad} f_N(\hat{x}_N) = 0$ and f_N is C^2 near the geodesic γ from $x^\star$ to $\hat{x}_N$, then, with $\bar{H}_N$ the average over γ of $\operatorname{Hess} f_N$ parallel-transported to $T_{x^\star}\mathcal{M}$,*

$$\bar{H}_N \operatorname{Log}_{x^\star}\hat{x}_N = -e_N(x^\star), \qquad \textit{hence} \qquad \frac{\|e_N(x^\star)\|}{L_0} \le d(\hat{x}_N, x^\star) \le \frac{\|e_N(x^\star)\|}{\mu_0} \tag{3}$$

whenever $\mu_0 \le \operatorname{Hess} f_N \le L_0$ along γ.

Each part is one line: scalarize in the direction of $e_N(x)$; expand in a frame; integrate the derivative of the transported gradient of f_N along γ. Their content is in what they need. Part (a) needs a uniform bound on the field integrand, part (b) its variance at one point at a time, part (c) only its quadrature error at $x^\star$, and to first order (c) is an identity, $d(\hat{x}_N, x^\star) = \left\|\mathrm{Hess}\, f(x^\star)^{-1} e_N(x^\star)\right\| (1 + o(1))$ as $\bar{H}_N \to \mathrm{Hess}\, f(x^\star)$: the finite-sample form of the influence expansion of M-estimators on manifolds (Bhattacharya & Patrangenaru, 2003; 2005), with quadrature error in place of sampling error.

The readings are the optimizers. An optimizer that *reuses* one point set produces iterates that depend on the points, and only (a) survives that dependence. One that *refreshes* a scrambled net at every step sees an unbiased gradient whose variance is (b). One that *solves* the sampled problem once needs (c). The standard descent and stochastic-approximation arguments then give the budgets.

Corollary 2.3 (budgets; Theorems C.9, C.13 and C.15)**.** *Let f be L-smooth and bounded below.* Reuse: *if every $F(\cdot, \omega)$ is L-smooth and the variations of F and G_x are bounded, uniformly in N, on a set containing the sublevel sets $\{f_N \le f_N(x_0)\}$, Riemannian gradient descent on f_N with the first N points of a Sobol' or Halton sequence reaches $\|\mathrm{grad}\, f\| \le \varepsilon$ with $\widetilde{O}(\varepsilon^{-3})$ evaluations of $\mathrm{grad}_x F$.* Refresh: *Riemannian SGD with a freshly scrambled net per step reaches $\mathbb{E}\,\|\mathrm{grad}\, f(x_\iota)\|^2 \le \varepsilon^2$ with $\widetilde{O}(\varepsilon^{-2-2/a})$ evaluations if $\mathbb{E}\,\|e_N\|^2 = \widetilde{O}(N^{-a})$; $a = 3$ for field integrands with continuous mixed derivatives.* Solve: *under (c), $d(\hat{x}_N, x^\star)$ is of the order of $\|e_N(x^\star)\|$: $\widetilde{O}(N^{-1})$ for low-discrepancy sets if $G_{x^\star}$ has bounded variation, and $\widetilde{O}(N^{-3/2})$ in root mean square for one scrambled net if $G_{x^\star}$ has continuous mixed derivatives.*

Independent minibatches need ε^{-4}, the optimal order for unbiased oracles of bounded variance (Arjevani et al., 2023). Under the mean-squared smoothness that reuse assumes, variance reduction reaches ε^{-3} (Zhou et al., 2019), which reuse matches up to logarithms; refresh is below both because it also uses regularity in ω.

Sturm's contraction is a reading of the field. On a Hadamard manifold every Fréchet functional $f_\nu(x) = \frac{1}{2}\int d(x,y)^2 \,\mathrm{d}\nu(y)$ is 1-strongly convex and $y \mapsto \langle \mathrm{Log}_x\, y, v\rangle$ is 1-Lipschitz for unit v. Reading (c) with H the Lipschitz functions, whose discrepancy is W_1, gives $d(b(\nu), b(\nu')) \le W_1(\nu, \nu')$ for barycentres of measures with finite second moments, the contraction Sturm (2003) proves on all spaces of nonpositive curvature; with H the Hardy–Krause variation of a transported integrand it gives the QMC rate of Karcher means (Theorem E.3). One inequality, two certificates.

3 A SCALING LAW FOR SINGULAR FIELD INTEGRANDS

The three readings need $\|G_x\|_H < \infty$. For $F = \frac{1}{2}d^2$ the field integrand is $G_x = -\mathrm{Log}_x$, and where the data can reach the cut locus of x this norm is infinite. On $\mathbb{S}^d$, in normal coordinates z centred at the antipode $-x$,

$$\mathrm{Log}_x\left(\mathrm{Exp}_{-x} z\right) = \pi \frac{z}{\|z\|} - z, \tag{4}$$

a direction that jumps at $z = 0$; on $\mathbb{S}^2$ its components lie in H^τ only for $\tau < 1$ (Proposition E.8). What survives is decided by positivity.

Lemma 3.1 (sandwich; Appendix D)**.** *Let Q be a rule with nonnegative weights summing to one, g a bounded function and $g_\pm \in H + \mathbb{R}$ with $g_- \le g \le g_+$. Then*

$$-\|g_-\|_H\, D_H(Q) - P(g - g_-) \;\le\; Qg - Pg \;\le\; \|g_+\|_H\, D_H(Q) + P(g_+ - g).$$

The proof is $Qg \le Qg_+ = Pg_+ + (Q - P)g_+$; one-sided approximation of this kind is classical (Vaaler, 1985). Let $\mathcal{S}_g(t)$ be the best such bound at discrepancy t, the larger of the two one-sided infima over sandwiches. It is the best bound a positive rule admits in terms of its discrepancy:

Proposition 3.2 (the sandwich is exact; Proposition D.2)**.** *If Ω is a compact metric space, H a reproducing kernel Hilbert space with continuous kernel, and g bounded and upper semicontinuous, then for every $t > 0$*

$$\max\{\nu g - Pg :\ \nu \in \mathcal{P}(\Omega),\ D_H(\nu) \le t\} = \inf\{t\,\|a\|_H + P(a + c - g) :\ a \in H,\ c \in \mathbb{R},\ a + c \ge g\},$$

and equal-weight rules approach the maximum.

This is Sion's minimax theorem for the pairing of measures and functions (Sion, 1958): the best sandwich and the worst measure coincide, so upper bounds come from sandwiches and lower bounds from measures. Both scale by dilation.

Theorem 3.3 (scaling law; Theorems D.6 and D.10)**.** *(i)* Sphere. *Let $\tau > d/2$ and $0 \le m < \tau - d/2$. Let g be bounded on $\mathbb{S}^d$ and equal, in normal coordinates z centred at a point p, to $b(z)\,\|z\|^m\,\omega(z/\|z\|)$ plus a function in $H^\tau(\mathbb{S}^d)$, with b and ω smooth and b supported in $\{\|z\| < \pi/4\}$. Then every positive rule Q with $D = \mathrm{wce}_{H^\tau}(Q) \le 1$ satisfies*

$$|Qg - Pg| \le C\,D^{(m+d)/(\tau+d/2)},$$

where C depends on finitely many derivatives of b and ω and on the norm of the remainder, and not on p; at $m = \tau - d/2$ the same holds with an extra factor $(1 + \log(1/D))^{1/2}$.

(ii) Square. *Let $g = a + b\,\omega(Z/\|Z\|)$ with $V_{\mathrm{HK}}(a) < \infty$, $\omega \in C^2(\mathbb{S}^1)$, $b \in C^2_c$ supported near an interior point u_0 of $[0,1]^2$, and Z a C^2 diffeomorphism there with $Z(u_0) = 0$, and $|(g-a)(u_0)| \le \|b\|_\infty\,\|\omega\|_\infty$. Then every point set satisfies $|Qg - Pg| \le C\,D_N^*(1+\log(1/D_N^*))$. If instead $g = a + \sum_k b_k \mathbf{1}_{\Pi_k}$ with $b_k \in C^2$ and finitely many convex polygons Π_k, whose edges may be oblique, then $|Qg - Pg| \le C\,(D_N^*)^{1/2}$, the classical order of the isotropic discrepancy (Kuipers & Niederreiter, 1974).*

Both parts are instances of (2). A bump of width r has norm of order $r^{-\beta}$ with $\beta = \tau - d/2$ in $H^\tau(\mathbb{S}^d)$ at a point ($c = d$), $\beta = 0$ in Hardy–Krause variation at a point ($c = 2$; the logarithm appears because $m = \beta$), and $\beta = 1$ along an oblique segment ($c = 1$).

Why this exponent. Cut g off at distance r from its singular set. The cut function costs a sum over dyadic shells $2^j r \le \|z\| < 2^{j+1} r$; each shell is a dilation of one fixed profile and costs $(2^j r)^{m-\beta}$, and the sum is geometric, so the cut costs $r^{m-\beta}$ in H. At $m = \beta$ every shell costs the same; the shells add in quadrature in a Hilbert norm and linearly in Hardy–Krause variation, whence the factors $(\log(1/D))^{1/2}$ in (i) and $\log(1/D)$ in (ii). The discarded part is at most r^m times the node mass within distance $2r$, and because the weights are positive the certificate bounds that mass through a bump: $r^c + r^{-\beta} D$. Balancing $r^{m-\beta} D$ against r^{m+c} gives $r = D^{1/(c+\beta)}$ and (2). The same dilation, run backwards on a multipole of width r whose moments vanish below an order above β, so that its Fourier transform is small where the certificate weight is large, shows that on spheres the exponent cannot be improved when the singularity is genuine ($b(0) \neq 0$ and $\|z\|^m\,\omega(z/\|z\|)$ not a polynomial). One limit of what is proved belongs here: for singular sets of codimension one on curved manifolds, such as the cut loci of projective spaces, the argument predicts $1/(\tau + \frac{1}{2})$ in H^τ but the proof needs the dilation lemma in a tubular neighbourhood.

Theorem 3.4 (sharpness and floor; Theorems D.11 and D.12 and Proposition D.13)**.** *Let $F = \frac{1}{2}d^2$ on $\mathbb{S}^d$, weighted by a continuous density $h \ge 0$.*

(a) For $h \equiv 1$, $\tau > d/2$ and $\gamma > d/(\tau + d/2)$ there are node sets Y_N with $D = \mathrm{wce}_{H^\tau}(Y_N) \to 0$ and a point x_0 with $\|e_N(x_0)\|/D^\gamma \to \infty$; if $\tau > d/2 + 1$, the same holds for $|\Phi_N(x_0)|$ and $\gamma > (d+1)/(\tau + d/2)$, and if $\tau = d/2 + 1$, $|\Phi_N(x_0)| \ge c\,D(\log(1/D))^{1/2}$.

(b) In Theorem 3.3(ii), if $b(u_0) \neq 0$ and ω is not constant, there are point sets with $D_N^* \to 0$ and $|Qg - Pg| \ge c\,D_N^* \log(1/D_N^*)$; a jump of nonzero size across an oblique segment gives $c\,(D_N^*)^{1/2}$.

(c) Every positive rule with distinct nodes y_i and weights w_i satisfies $\sup_{x \notin -Y_N} \|e_N(x)\| \ge \pi \max_i w_i h(y_i)$, that is $\pi \max_i h(y_i)/N$ for equal weights.

In (a) such a multipole next to $-x_0$ moves the field at x_0 by order r^d, the value by r^{d+1}, and the discrepancy only by $r^{\tau+d/2}$, so with Theorem 3.3(i) these exponents are exact among bounds in terms of wce_{H^τ} alone, as is the exponent of every genuine singularity (Theorem D.11); for the field and the cap discrepancy it is $2d/(2d+1)$. In (b) a discrepancy function, the dual of Hardy–Krause variation, of height ϵ and shaped like the singularity reaches down to the scale $\epsilon^{1/2}$ that positivity allows, each dyadic shell contributing ϵ: the balance $r = (D_N^*)^{1/2}$ of the upper bound, read backwards. Part (c) is the field of a single node: $\mathrm{Log}_x(y_i)$ reverses its direction as x crosses $-y_i$, and nothing the other nodes do cancels a jump. The floor bounds the worst case over decision points; at a fixed x the field can be far smaller (Section 5). It is met, up to N^ϵ, by rules equidistributed to every order.

Table 1: **One exponent, proved row by row** (Appendices D and E). Exponent κ in $|Q_N g - Pg| \le C\, D^\kappa$, uniform in x (locally in row 5), for the field ($m = 0$) and value ($m = 1$) integrands of $F = \frac{1}{2}d^2$. In every row $\kappa = (m + c)/(c + \beta)$ if $m < \beta$ and $\kappa = 1$ if $m > \beta$; at $m = \beta$, $\kappa = 1$ with a factor $(\log(1/D))^{1/2}$ in Sobolev norms and $\log(1/D)$ in Hardy–Krause variation, marked † (in row 3 at $\tau = \frac{d}{2} + 1$), except for the torus value. Sharp over all node sets in terms of D alone: rows 2 and 3 with their dagger, the logarithm of row 5 and the square root of row 6.

space, certificate	singular set of G_x	c	β	field	value
Hadamard manifold or small ball, D_N^*	none	–	–	1	1
$\mathbb{S}^d$, cap discrepancy D_{L_2}	the antipode	d	$\frac{1}{2}$	$\frac{2d}{2d+1}$, sharp	1
$\mathbb{S}^d$, Sobolev discrepancy wce_{H^τ}	the antipode	d	$\tau - \frac{d}{2}$	$\frac{d}{\tau+d/2}$, sharp	$\min\{1, \frac{d+1}{\tau+d/2}\}^\dagger$
polar chart, x a pole, D_N^*	an edge	–	–	1	1
polar chart, $-x$ interior, D_N^*	an interior point	2	0	$1^\dagger$	1
flat torus, oblique lattice, D_N^*	segments	1	1	$\frac{1}{2}$	1

Corollary 3.5 (designs meet the floor up to N^ϵ; Corollary D.14)**.** *Let h be smooth and Q_N positive rules with N nodes. If $\mathrm{wce}_{H^\tau}(Q_N) = O(N^{-\tau/d})$, as for QMC designs of order τ (Brauchart et al., 2014), then $\sup_{x \notin -\operatorname{supp} Q_N} \|e_N(x)\| = O(N^{-\tau/(\tau+d/2)})$. For $d \ge 2$, positive rules exact to degree t with $N \asymp t^d$ nodes satisfy this for every $\tau > d/2$ (Brauchart & Hesse, 2007), so their field error is $O(N^{-1+\epsilon})$ for every $\epsilon > 0$; they include spherical t-designs (Delsarte et al., 1977), which exist with $N \asymp t^d$ (Bondarenko et al., 2013).*

Node sets that maximize the sum of distances are QMC designs for $\tau \le (d+1)/2$ (Brauchart et al., 2014, Thms. 4, 16); for them the corollary gives $N^{-(d+1)/(2d+1)}$, that is $N^{-3/5}$ on $\mathbb{S}^2$, and the energy that defines them sees only $\tau = 3/2$. Equidistribution of higher order suffices to close the gap to the floor N^{-1}; whether it is necessary is open.

4 Geometry decides the singularity

The law turns every geometric question about certificates into two: where is the field integrand singular, and what are the order and codimension of the singularity as the integration space sees it?

No cut locus, no price. On a Hadamard manifold Log_x is smooth and every sample Fréchet functional is 1-strongly convex. A transport $T\colon [0,1]^s \to \mathcal{M}$ with $T_\# \lambda = \nu$ that is C^s up to the boundary, or whose mixed derivatives have integrable singularities (Lemma E.1), gives a field integrand of bounded variation, and the one-point reading gives $d(\hat{x}_N, x^\star) \le V_{\mathrm{HK}}(\mathrm{Log}_{x^\star} \circ T)\, D_N^*$ (Theorem E.3). With positive curvature the same holds when the data lie in a ball small enough to keep them away from the cut locus and the Fréchet functional convex (Theorem E.6).

The sphere is electrostatics. With full support the cut locus cannot be avoided. On $\mathbb{S}^d$, since $\|x + y\| = 2\cos\frac{d(x,y)}{2}$,

$$\tfrac{1}{2}d(x,y)^2 = \tfrac{\pi^2}{2} - \pi\,\|x+y\| + r\big(d(x,y)\big), \qquad r(\delta) = \tfrac{1}{2}\delta^2 + 2\pi\cos\tfrac{\delta}{2} - \tfrac{\pi^2}{2}. \tag{5}$$

The point is what r lacks. At distance $s = \pi - \delta$ from the antipode, $r = \frac{1}{2}s^2 - \frac{\pi}{24}s^3 + O(s^5)$ has no cone, and its gradient in x is $z + O(\|z\|^2)$, the smooth part of $-\mathrm{Log}_x$ in (4). The whole singularity of the Fréchet loss is the chordal distance to the antipode.

Proposition 4.1 (the sphere is electrostatics; Proposition E.9)**.** *Let $\rho_N = Q_N - \sigma$ and let $U_N(z) = \int \|z - y\|\, \mathrm{d}\rho_N(y)$ be the chordal potential of the charge. Then*

$$\Phi_N(x) = -\pi U_N(-x) + \rho_N R_x, \quad e_N(x) = \pi \nabla U_N(-x) + \rho_N S_x, \quad C_d D_{L_2}^2 = -\rho_N U_N,$$

the second for $x \notin -\operatorname{supp} Q_N$. Here $R_x = r(d(x,\cdot))$ and $S_x = \mathrm{grad}_x R_x$ lie in $H^\tau(\mathbb{S}^d)$ for $\tau < d/2 + 2$ with norms independent of x, $\|U_N\|_\infty \le C_d\, D_{L_2}$, and $C_d = d\sqrt{\pi}\,\Gamma(\frac{d}{2})/\Gamma(\frac{d+1}{2})$.

The last identity is Stolarsky's invariance principle: the cap discrepancy is the energy of the charge in the chordal-distance kernel. The first two say that the Fréchet value error is its potential and the

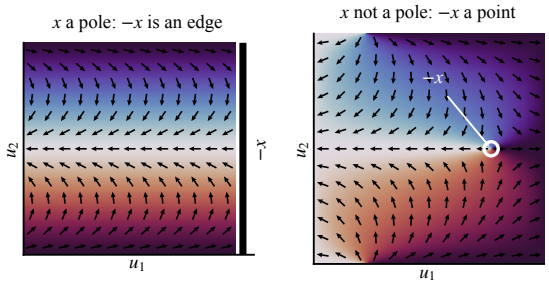

Figure 2: **A polar chart is a blow-up.** Direction of the field integrand $u \mapsto \mathrm{Log}_x(T(u))$ in the Lambert chart (color and arrows). Left: $x = e_3$; the chart replaces the cut point $-x$ by the edge $u_1 = 1$, along which the direction is the smooth angle $2\pi u_2$. Right: x at polar angle 1; $-x$ is an interior point around which every direction occurs.

gradient error its field, both read at the antipode, up to remainders controlled by discrepancies of higher order. The value certificate is therefore linear, with a constant the decomposition computes in closed form: on $\mathbb{S}^2$, $\|\Phi_N\|_\infty \le 6.062\, D_{L_2}$, and a single node attains $6.006\, D_{L_2}$. The field $\nabla U_N(z) = \frac{1}{N}\sum_i P_z(z - y_i)/\|z - y_i\|$, with P_z the projection onto $T_z\mathbb{S}^d$, averages vectors of length at most 1 that point away from the nodes and jumps by $2/N$ at each node, and how large it can be at a given energy is the content of Theorems 3.3 and 3.4.

Charts are blow-ups. Transported nets see the field integrand through a chart, and a chart can resolve the singularity. A cylindrical chart $T(u) = (\sqrt{1-z^2}\cos 2\pi u_2, \sqrt{1-z^2}\sin 2\pi u_2, z)$ with $z = z(u_1)$ is the real blow-up of the two poles: each pole is replaced by an edge of the square whose points are the directions at that pole (for smooth integrands and the equal-area chart, bounded variation is due to Basu & Owen, 2016). In these coordinates the jumping direction $z/\|z\|$ of (4) at a pole becomes the smooth function $u_2 \mapsto (\cos 2\pi u_2, \sin 2\pi u_2)$ (Figure 2).

Proposition 4.2 (a chart resolves the cut points it blows up; Proposition E.10)**.** *Let T be a cylindrical chart with a C^2 latitude profile, $z' > 0$, $h \in C^2(\mathbb{S}^2)$, and e_N the field of the weighted rule $\frac{1}{N}\sum_i h(T(u_i))\,\frac{1}{2}d(\cdot, T(u_i))^2$ with nodes u_i in the square, none mapped to $-x$. (a) If x is a pole of T, the field integrand $u \mapsto h(T(u))\,\mathrm{Log}_x T(u)$ has bounded Hardy–Krause variation, so $\|e_N(x)\| \le C\, D_N^*$. (b) If $-x = T(u_0)$ for an interior point u_0, the field integrand has a point singularity at u_0, of order 0 when $h(-x) \neq 0$, and $\|e_N(x)\| \le C\, D_N^*(1 + \log(1/D_N^*))$, uniformly on compact sets of such x.*

Where the sampled objective is smooth and convex near the minimizer, the one-point reading makes the minimizer inherit the rate of the field at $x^\star$; on the sphere this holds for typical draws, while a node within $O(1/N)$ of $-x^\star$ can move the minimizer by $O(1/N)$, a case the uniform bound of Corollary D.8 covers (Appendix E.5). The prescription is therefore to *blow up the antipode of the minimizer*: center the polar chart at an estimate of $x^\star$ and carry the data law by importance weights. The von Mises–Fisher chart of a law centred at its mean does this, since its pole is the mean.

Designing the charge. Riemannian optimization also produces the nodes. By Stolarsky's principle, ascent on the sum of distances $\sum_{i<j}\|y_i - y_j\|$ over $(\mathbb{S}^2)^N$ lowers the energy of the charge and evaluates the attained D_{L_2} exactly at every iterate (Proposition F.1); energy methods of this kind are classical (Hardin & Saff, 2004; Gräf & Potts, 2011; Damelin & Grabner, 2003). By the law, what the energy buys at optimal order is a certificate of order $N^{-3/4}$ for values and $N^{-3/5}$ for fields; equidistribution of higher order than the energy measures suffices to reach the floor.

5 Experiments: the predictions, tested

Each experiment tests one statement of Sections 2–4; Table 5 in the appendix lists them. MC denotes independent samples, Halton the deterministic sequence with its origin dropped, and SSobol SciPy's Sobol' nets with linear matrix scrambling, a digital shift and 30 random digits (Matoušek, 1998; Joe & Kuo, 2008); these have the variance of nested scrambling (Lemma B.4), up to a finite-precision bias that changes no reported Karcher or Grassmann error by more than a relative $5 \cdot 10^{-5}$ (Appendix G.9). Slopes are least-squares fits on the six largest sizes unless stated; bands are 10th–90th percentiles, not confidence intervals. Protocols, seeds and all replicate values are in Appendix G and the code.

One point decides the minimizer. We compute Karcher means on $\mathcal{P}_3$ with the affine-invariant metric. A bounded transport whose matrix terms do not commute makes every sample Karcher

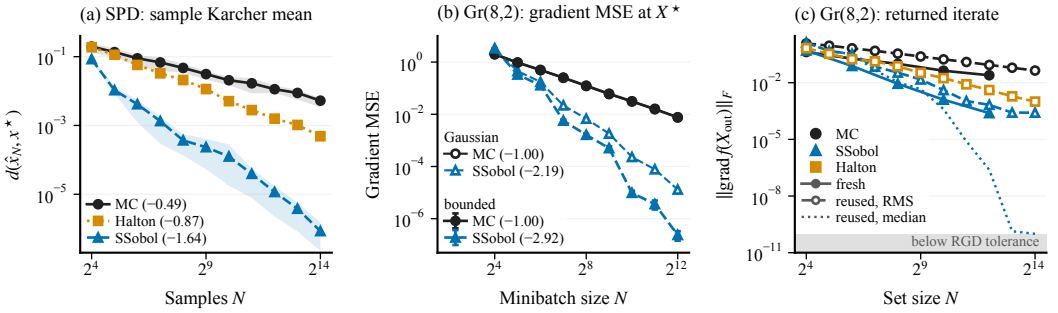

Figure 3: **One point decides the minimizer (a); the transport (b) and the sampler (c) set the rate, the reading a constant.** (a) Bounded non-commutative model on $\mathcal{P}_3$: Karcher-mean error, RMS over 20 replicates (Halton: one deterministic rule). (b) $\mathrm{Gr}(8,2)$: mean-square field at the optimum, 200 nets per size; Gaussian transport hollow, bounded filled. (c) Bounded transport: population gradient norm of the returned iterate; fresh (solid): RSGD with minibatches of size N, last iterate after 2^{20} samples, RMS over 10 runs; reused (dashed): gradient descent on one set of N points to $\|\mathrm{grad}\, f_N\|_F < 10^{-10}$, RMS over 64 sets; dotted: median over the SSobol sets.

functional 1-strongly convex and smooth, so Proposition 2.2(c) applies to every draw (Appendix G.3). At $N = 2^{14}$ the errors are $5.3 \cdot 10^{-3}$ (MC), $4.9 \cdot 10^{-4}$ (Halton) and $8.5 \cdot 10^{-7}$ (SSobol), with slopes -0.49, -0.87 and -1.64 (Figure 3a). The one-point formula predicts more than the rate. With $\mathrm{Hess}\, f(x^\star)$ computed once from a 2^{18}-point rule (eigenvalues 1.000 to 1.025), the prediction $\left\|\mathrm{Hess}\, f(x^\star)^{-1} e_N(x^\star)\right\|$ matches the error of every replicate, with relative deviation at most $2.8 \cdot 10^{-4}$ for MC and $2.4 \cdot 10^{-6}$ for SSobol at $N = 2^{14}$, shrinking with the error as a second-order remainder must (Table 10). The ratio of error to field, which the two-sided bound (3) only places between $1/L_0$ and 1 and which ranges over $[0.974, 0.999]$ across the draws of Table 10, is thereby resolved draw by draw. On the sphere, where the Hessian at the mean is a multiple λ of the identity, the median ratio is within 0.5% of $1/\lambda$ at the largest size in every chart of the experiment below.

The sampler and the transport set the rate; reuse costs a constant. On $\mathrm{Gr}(8,2)$ we estimate the leading eigenspace of a covariance with eigengap 2 (Appendices G.1 and G.2). The field integrand $-(I - XX^\top)a(u)a(u)^\top X$ is quadratic in the random vector $a(u)$, so its regularity in u is that of the transport. A Gaussian transport $a(u) = W\Phi^{-1}(u)$ is unbounded; a bounded one, $a(u) = \sqrt{3}\, W(2u - \mathbf{1})$, has the same covariance and continuous mixed derivatives. The mean-square reading predicts the scrambled-net rate N^{-3} for the bounded transport only, and the measured mean-square field has slope -2.92 ($[-3.04, -2.84]$) there, against -2.19 for the Gaussian transport and -1.00 for MC, fitted over all nine sizes (Figure 3b). Panel (c) reads the same field in two ways: at $N = 4096$ the population gradient norms of the returned iterates are $2.5 \cdot 10^{-2}$ (fresh MC), $8.7 \cdot 10^{-2}$ (reused MC), $2.5 \cdot 10^{-4}$ (fresh SSobol) and $6.7 \cdot 10^{-4}$ (reused SSobol). The sampler moves the error by two orders of magnitude and the reading by a factor of about three that depends on the step size and the fresh budget; the readings predict the shared rate, not the factor. The dotted median collapses because most scramblings integrate the quadratic integrand almost exactly, a digital-net cancellation absent from a smooth non-polynomial control (Appendix G.2).

The chart decides. By Proposition 4.2, a chart that blows up $-x$ gives the field integrand at x bounded variation and one that does not leaves it unbounded, at the price of one logarithm in the proved bounds. Scrambled nets gain their faster rates from regularity, so we expect a larger separation for them, and test this twice (Appendix G.6). For the uniform law on $\mathbb{S}^2$ the field has target 0 at every x; with one Lambert equal-area chart, the RMS field at the pole decays with slope -1.76, and at three points off the pole with slopes -0.91, -0.99 and -1.02, against about -0.5 for MC (Figure 4a). Next we fix a von Mises–Fisher law with concentration $\varkappa = 0.5$ and mean $x^\star$ at polar angle 1, whose Fréchet mean is $x^\star$, and change only the chart: the Lambert chart about e_3 leaves $-x^\star$ interior, while the same chart rotated to $x^\star$ and the von Mises–Fisher chart about $x^\star$ both blow it up. The Fréchet-mean error follows the field, as the one-point reading predicts for typical draws: its median decays with slope -1.01 in the fixed chart and -1.55 and -1.54 in the two resolving charts, and at $N = 2^{17}$ it is more than 60 times smaller there (Figure 4b), while the median ratio of error to field stays within 0.5% of $1/\lambda$ in every chart. For $\varkappa = 2$ the singular part of the field, which carries

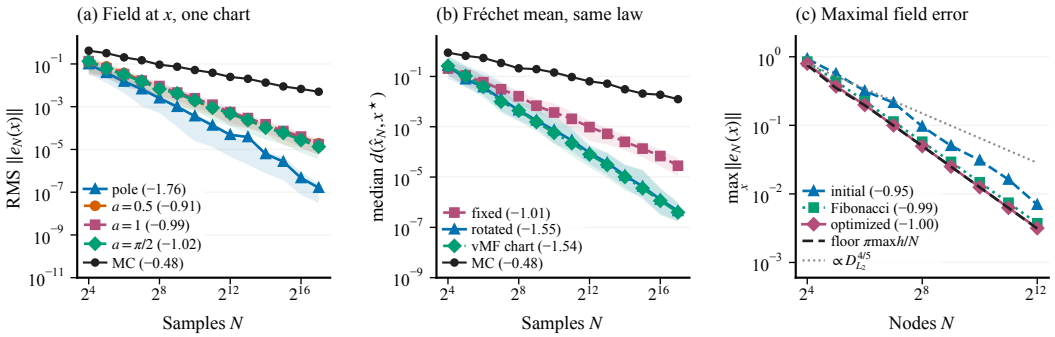

Figure 4: **The chart decides (a, b), and optimized nodes sit at the floor (c).** (a) Uniform law on $\mathbb{S}^2$, one Lambert chart about e_3, scrambled nets: RMS field at the pole, whose antipode the chart blows up, and at three points at polar angle a, antipodes interior. (b) von Mises–Fisher law, $\varkappa = 0.5$, mean at polar angle 1: median Fréchet-mean error in the fixed Lambert chart about e_3 (antipode interior), the same chart rotated to the mean and the von Mises–Fisher chart (both resolved); 64 scramblings per size in (a, b). (c) Nodes weighted by a von Mises–Fisher density ($\varkappa = 2$): maximal field error, the floor of Theorem 3.4(c), and the slope of the ceiling $D_{L_2}^{4/5}$. Legends give fitted slopes.

the data density $h(-x^\star) = 0.07$ at the antipode (0.58 for $\varkappa = 0.5$), overtakes the $N^{-3/2}$ term only beyond the tested range: the fixed chart shows the transitional slope -1.11 and a median error 3.4 and 3.6 times those of the rotated and the von Mises–Fisher chart at $N = 2^{17}$ (Appendix G.6). The separation is observed, not proved: no variance lower bound is known in the unresolved chart, and in a resolved chart the mixed derivatives near the blown-up edge are integrable but not square integrable.

Intrinsic nodes sit at the floor. Riemannian ascent on the sum of distances over $(\mathbb{S}^2)^N$, $N = 2^4, \dots, 2^{12}$, from Lambert images of scrambled nets lowers the deficit to 54–70% of its initial value, within 3.4% of the Fibonacci spiral (Appendix G.4). We weight the nodes by a von Mises–Fisher density ($\varkappa = 2$) and search the maximal field error over x on refined grids and next to every antipode. For these nodes Theorems 3.3 and 3.4 prove only the bracket between the floor $\pi \max_i h(y_i)/N$ and a ceiling of order $D_{L_2}^{4/5}$, yet the optimized sets sit at the floor: their maxima decay with slope -1.00 and lie within 2% of it for $N \geq 64$, while the Fibonacci maxima are 1.15–1.20 times the floor (slope -0.99) and the initial sets 1.2–2.6 times (slope -0.95; Figure 4c). Their value maxima decay with slope -1.48 and -1.47, faster than the proved $N^{-3/4}$. Rules exact to degree t, rotated Gauss–Legendre $\times$ trapezoid rules with $N \approx t^2/2$ nodes and positive weights w_i, test Corollary 3.5 directly: it allows their maximal field error anywhere between the floor $\pi \max_i w_i h(y_i)$ and $N^{-1+\epsilon}$, and it equals the floor to within 5% at $N = 32$ and 0.2% for $N \geq 1152$; their value maxima decay with slope -1.46 over the five largest sizes (Appendix G.7).

6 Discussion

A point set knows about a Riemannian minimizer what the field of its discrepancy charge knows at that point, and, where the sampled objective is smooth near the minimizer, to first order nothing else. Where the loss is regular in the data the certificate reaches the field; at the cut locus it keeps the exponent $(m + c)/(c + \beta)$, which positivity attains and dual measures show cannot be improved, a single node puts a floor under, and a chart that blows up the cut point makes unnecessary.

Three rules follow for practice: refresh a scrambled net at every step when the loss gradient is smooth in the data (Corollary 2.3); on a sphere, put the pole of the chart at an estimate of the minimizer (Proposition 4.2); and to approach the floor, use rules equidistributed to every order (Corollary 3.5).

The conditions of the present statements say what comes next. The budgets assume a chosen-point oracle and regularity in ω, with constants that grow with the dimension, for which weighted function spaces are natural (Dick et al., 2013). The chart separation needs variance bounds for scrambled nets at point singularities and for integrable but not square-integrable mixed derivatives, and codimension-one cut loci on curved manifolds need the dilation lemma in a tubular neighbourhood.

Reproducibility Statement

The complete statements, assumptions and proofs are in Appendices B–F, each proof following its result. Appendix G contains every experimental comparison and Appendix G.10 the implementation details, random seeds, solver tolerances and reference computations. The supplementary code contains the experiment scripts, their configurations, replicate-level JSON results, the stored spherical point sets, the checks of the one-point law and of the dipole scalings, and scripts that redraw the main figures from the stored results without rerunning experiments. All problems are synthetic; no external or private data are used.

The randomization in the implementation is SciPy's linear matrix scrambling with a digital shift and 30 randomized bits. Lemma B.4 shows that its ideal version has the variance of the nested scrambling assumed by the variance bounds; Lemma B.5 bounds the effect of the 30-bit truncation and Appendix G.9 measures it. Fitted slopes and resampling intervals are reported separately from asymptotic statements, and the software environments of all runs are recorded with their results.

Ethics Statement

This work concerns numerical integration and optimization on manifolds. The experiments use synthetic distributions and involve no human participants, personal data or sensitive attributes. We do not identify a direct ethical risk. The methods may be used inside larger learning systems, where questions of data consent, privacy, fairness and safety require separate assessment.

AI Use Statement

The authors developed the original research question and wrote the initial manuscript. AI tools were used to check mathematical arguments, fill in proof steps, draw TikZ figures, write and run code and numerical diagnostics, and polish the language. The lower-bound constructions of Theorem D.11 and Appendix G.8 came from AI-assisted review. The authors take full responsibility for the content and correctness of the manuscript.

References

P.-A. Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008. ISBN 978-0-691-13298-3. doi: 10.1515/9781400830244.

Bijan Afsari. Riemannian L^p center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673, 2011. doi: 10.1090/S0002-9939-2010-10541-5.

Christoph Aistleitner, Johann S. Brauchart, and Josef Dick. Point sets on the sphere $\mathbb{S}^2$ with small spherical cap discrepancy. *Discrete & Computational Geometry*, 48(4):990–1024, 2012. doi: 10.1007/s00454-012-9451-3.

Ralph Alexander. On the sum of distances between n points on a sphere. *Acta Mathematica Academiae Scientiarum Hungaricae*, 23(3–4):443–448, 1972. doi: 10.1007/BF01896964.

Yossi Arjevani, Yair Carmon, John C. Duchi, Dylan J. Foster, Nathan Srebro, and Blake Woodworth. Lower bounds for non-convex stochastic optimization. *Mathematical Programming*, 199(1–2):165–214, 2023. doi: 10.1007/s10107-022-01822-7.

Miroslav Bačák. *Convex Analysis and Optimization in Hadamard Spaces*, volume 22 of *De Gruyter Series in Nonlinear Analysis and Applications*. De Gruyter, Berlin, 2014. ISBN 978-3-11-036103-2. doi: 10.1515/9783110361629.

Kinjal Basu and Art B. Owen. Transformations and Hardy–Krause variation. *SIAM Journal on Numerical Analysis*, 54(3):1946–1966, 2016. doi: 10.1137/15M1052184.

Kinjal Basu and Art B. Owen. Scrambled geometric net integration over general product spaces. *Foundations of Computational Mathematics*, 17(2):467–496, 2017. doi: 10.1007/s10208-015-9293-5.

József Beck. Sums of distances between points on a sphere—an application of the theory of irregularities of distribution to discrete Geometry. *Mathematika*, 31(1):33–41, 1984. doi: 10.1112/S0025579300010639.

Glaydston C. Bento, Orizon P. Ferreira, and Jefferson G. Melo. Iteration-complexity of gradient, subgradient and proximal point methods on Riemannian manifolds. *Journal of Optimization Theory and Applications*, 173(2):548–562, 2017. doi: 10.1007/s10957-017-1093-4.

Rajendra Bhatia. *Positive Definite Matrices*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2007. ISBN 978-0-691-12918-1.

Rabi Bhattacharya and Vic Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *The Annals of Statistics*, 31(1):1–29, 2003. doi: 10.1214/aos/1046294456.

Rabi Bhattacharya and Vic Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *The Annals of Statistics*, 33(3):1225–1259, 2005. doi: 10.1214/009053605000000093.

Dmitriy Bilyk, Feng Dai, and Ryan Matzke. The Stolarsky principle and energy optimization on the sphere. *Constructive Approximation*, 48(1):31–60, 2018. doi: 10.1007/s00365-017-9412-4.

Andriy Bondarenko, Danylo Radchenko, and Maryna Viazovska. Optimal asymptotic bounds for spherical designs. *Annals of Mathematics*, 178(2):443–452, 2013. doi: 10.4007/annals.2013.178.2.2.

Silvère Bonnabel. Stochastic gradient descent on Riemannian manifolds. *IEEE Transactions on Automatic Control*, 58(9):2217–2229, 2013. doi: 10.1109/TAC.2013.2254619.

Sergiy V. Borodachov, Douglas P. Hardin, and Edward B. Saff. *Discrete Energy on Rectifiable Sets*. Springer Monographs in Mathematics. Springer, New York, 2019. doi: 10.1007/978-0-387-84808-2.

Nicolas Boumal. *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, Cambridge, 2023. ISBN 978-1-009-16617-1. doi: 10.1017/9781009166164.

Nicolas Boumal, P.-A. Absil, and Coralia Cartis. Global rates of convergence for nonconvex optimization on manifolds. *IMA Journal of Numerical Analysis*, 39(1):1–33, 2019. doi: 10.1093/imanum/drx080.

Luca Brandolini, Leonardo Colzani, Giacomo Gigante, and Giancarlo Travaglini. On the Koksma–Hlawka inequality. *Journal of Complexity*, 29(2):158–172, 2013. doi: 10.1016/j.jco.2012.10.003.

Luca Brandolini, William W. L. Chen, Leonardo Colzani, Giacomo Gigante, and Giancarlo Travaglini. Discrepancy and numerical integration on metric measure spaces. *The Journal of Geometric Analysis*, 29(1):328–369, 2019. doi: 10.1007/s12220-018-9993-6.

Johann S. Brauchart and Josef Dick. Quasi-Monte Carlo rules for numerical integration over the unit sphere $\mathbb{S}^2$. *Numerische Mathematik*, 121(3):473–502, 2012. doi: 10.1007/s00211-011-0444-6.

Johann S. Brauchart and Josef Dick. A simple proof of Stolarsky's invariance principle. *Proceedings of the American Mathematical Society*, 141(6):2085–2096, 2013. doi: 10.1090/S0002-9939-2013-11490-5.

Johann S. Brauchart and Peter J. Grabner. Distributing many points on spheres: Minimal energy and designs. *Journal of Complexity*, 31(3):293–326, 2015. doi: 10.1016/j.jco.2015.02.003.

Johann S. Brauchart and Kerstin Hesse. Numerical integration over spheres of arbitrary dimension. *Constructive Approximation*, 25(1):41–71, 2007. doi: 10.1007/s00365-006-0629-4.

Johann S. Brauchart, Edward B. Saff, Ian H. Sloan, and Robert S. Womersley. QMC designs: Optimal order quasi Monte Carlo integration schemes on the sphere. *Mathematics of Computation*, 83(290):2821–2851, 2014. doi: 10.1090/S0025-5718-2014-02839-1.

Anna Breger, Martin Ehler, and Manuel Gräf. Quasi Monte Carlo integration and kernel-based function approximation on Grassmannians. In *Frames and Other Bases in Abstract and Function Spaces*, Applied and Numerical Harmonic Analysis, pp. 333–353. Birkhäuser, Cham, 2017. doi: 10.1007/978-3-319-55550-8_14.

Alexander Buchholz, Florian Wenzel, and Stephan Mandt. Quasi-Monte Carlo variational inference. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 668–677. PMLR, 2018. URL https://proceedings.mlr.press/v80/buchholz18a.html.

Isaac Chavel. *Riemannian Geometry: A Modern Introduction*, volume 98 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2 edition, 2006. ISBN 9780521853682. doi: 10.1017/CBO9780511616822. Theorem IX.6.1 (p. 404): if $0 < r < \frac{1}{2}\min\{\mathrm{inj}(M), \pi/\sqrt{\Delta}\}$ with Δ an upper bound on the sectional curvatures, then the geodesic ball $B(p, r)$ is (strongly) convex.

Steven B. Damelin and Peter J. Grabner. Energy functionals, numerical integration and asymptotic equidistribution on the sphere. *Journal of Complexity*, 19(3):231–246, 2003. Corrigendum: Journal of Complexity 20 (2004), 883–884.

Philippe Delsarte, Jean-Marie Goethals, and Johan Jacob Seidel. Spherical codes and designs. *Geometriae Dedicata*, 6(3):363–388, 1977. doi: 10.1007/BF03187604.

Josef Dick. Walsh spaces containing smooth functions and quasi-Monte Carlo rules of arbitrary high order. *SIAM Journal on Numerical Analysis*, 46(3):1519–1553, 2008. doi: 10.1137/060666639.

Josef Dick. Higher order scrambled digital nets achieve the optimal rate of the root mean square error for smooth integrands. *The Annals of Statistics*, 39(3):1372–1398, 2011. doi: 10.1214/11-AOS880.

Josef Dick. Erratum to "higher order scrambled digital nets achieve the optimal rate of the root mean square error for smooth integrands", 2026. URL https://arxiv.org/abs/1007.0842v5. arXiv:1007.0842v5.

Josef Dick and Friedrich Pillichshammer. *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge University Press, Cambridge, 2010. ISBN 978-0-521-19159-3. doi: 10.1017/CBO9780511761188.

Josef Dick, Frances Y. Kuo, and Ian H. Sloan. High-dimensional integration: The quasi-Monte Carlo way. *Acta Numerica*, 22:133–288, 2013. doi: 10.1017/S0962492913000044.

Manfredo Perdigão do Carmo. *Riemannian Geometry*. Mathematics: Theory & Applications. Birkhäuser, Boston, MA, 1992. ISBN 0-8176-3490-8. Translated from the Portuguese by Francis Flaherty.

Martin Ehler, Manuel Gräf, and Chris J. Oates. Optimal Monte Carlo integration on closed manifolds. *Statistics and Computing*, 29(6):1203–1214, 2019. doi: 10.1007/s11222-019-09894-w.

Benjamin Eltzner and Stephan F. Huckemann. A smeary central limit theorem for manifolds with application to high-dimensional spheres. *The Annals of Statistics*, 47(6):3360–3381, 2019. doi: 10.1214/18-AOS1781.

Mathieu Gerber. On integration methods based on scrambled nets of arbitrary size. *Journal of Complexity*, 31(6):798–816, 2015. doi: 10.1016/j.jco.2015.06.001.

Saeed Ghadimi and Guanghui Lan. Stochastic first- and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013. doi: 10.1137/120880811.

Takashi Goda and Kosuke Suzuki. The median trick does not help for fully nested scrambling. *Archiv der Mathematik*, 126(2):221–230, 2026. doi: 10.1007/s00013-025-02205-y.

Peter J. Grabner. Erdős–Turán type discrepancy bounds. *Monatshefte für Mathematik*, 111(2): 127–136, 1991.

Peter J. Grabner and Robert F. Tichy. Spherical designs, discrepancy and numerical integration. *Mathematics of Computation*, 60(201):327–336, 1993.

Manuel Gräf and Daniel Potts. On the computation of spherical designs by a new optimization approach based on fast spherical Fourier transforms. *Numerische Mathematik*, 119(4):699–724, 2011. doi: 10.1007/s00211-011-0399-7.

Loukas Grafakos. *Classical Fourier Analysis*, volume 249 of *Graduate Texts in Mathematics*. Springer, New York, 3 edition, 2014. doi: 10.1007/978-1-4939-1194-3.

J. H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2(1):84–90, 1960. doi: 10.1007/BF01386213.

Douglas P. Hardin and Edward B. Saff. Discretizing manifolds via minimum energy points. *Notices of the American Mathematical Society*, 51(10):1186–1194, 2004.

Zhijian He and Xiaoqun Wang. On the convergence rate of randomized quasi-Monte Carlo for discontinuous functions. *SIAM Journal on Numerical Analysis*, 53(5):2488–2503, 2015. doi: 10.1137/15M1007963.

Edmund Hlawka. Funktionen von beschränkter Variation in der Theorie der Gleichverteilung. *Annali di Matematica Pura ed Applicata*, 54(1):325–333, 1961. doi: 10.1007/BF02415361.

Tito Homem-de Mello. On rates of convergence for stochastic optimization problems under non-independent and identically distributed sampling. *SIAM Journal on Optimization*, 19(2):524–551, 2008. doi: 10.1137/060657418.

Reshad Hosseini and Suvrit Sra. An alternative to EM for Gaussian mixture models: batch and stochastic Riemannian optimization. *Mathematical Programming*, 181(1):187–223, 2020. doi: 10.1007/s10107-019-01381-4.

Thomas Hotz and Stephan Huckemann. Intrinsic means on the circle: uniqueness, locus and asymptotics. *Annals of the Institute of Statistical Mathematics*, 67(1):177–193, 2015. doi: 10.1007/s10463-013-0444-7.

Stephen Joe and Frances Y. Kuo. Constructing Sobol sequences with better two-dimensional projections. *SIAM Journal on Scientific Computing*, 30(5):2635–2654, 2008. doi: 10.1137/070709359.

Hermann Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541, 1977. doi: 10.1002/cpa.3160300502.

Matti Koivu. Variance reduction in sample approximations of stochastic programs. *Mathematical Programming*, 103(3):463–485, 2005. doi: 10.1007/s10107-004-0557-0.

J. F. Koksma. Een algemeene stelling uit de theorie der gelijkmatige verdeeling modulo 1. *Mathematica B (Zutphen)*, 11:7–11, 1942.

L. Kuipers and Harald Niederreiter. *Uniform Distribution of Sequences*. Pure and Applied Mathematics. Wiley-Interscience [John Wiley & Sons], New York-London-Sydney, 1974.

John M. Lee. *Introduction to Riemannian Manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer, Cham, 2nd edition, 2018. ISBN 978-3-319-91754-2. doi: 10.1007/978-3-319-91755-9.

H. Leövey and W. Römisch. Quasi-Monte Carlo methods for two-stage stochastic mixed-integer programs. *Mathematical Programming*, 190(1–2):361–392, 2021. doi: 10.1007/s10107-020-01538-6.

Mario Lezcano-Casado. Curvature-dependant global convergence rates for optimization on manifolds of bounded geometry, 2020.

Xian-Jin Li and Jeffrey D. Vaaler. Some trigonometric extremal functions and the Erdős–Turán type inequalities. *Indiana University Mathematics Journal*, 48(1):183–236, 1999.

Sifan Liu. Langevin quasi-Monte Carlo. In *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*, 2023. URL https://papers.nips.cc/paper_files/paper/2023/hash/ee56aa4fe26a189782f507d843fd5272-Abstract-Conference.html.

Sifan Liu and Art B. Owen. Quasi-Monte Carlo quasi-Newton in variational Bayes. *Journal of Machine Learning Research*, 22(243):1–23, 2021. URL https://www.jmlr.org/papers/v22/21-0498.html.

Yucheng Lu, Si Yi Meng, and Christopher De Sa. A general analysis of example-selection for stochastic gradient descent. In *International Conference on Learning Representations (ICLR)*, 2022. URL https://openreview.net/forum?id=7gWSJrP3opB.

Hoang Phuc Hau Luu, Hanlin Yu, Bernardo Williams, Marcelo Hartmann, and Arto Klami. Stochastic variance-reduced Gaussian variational inference on the Bures–Wasserstein manifold. In *International Conference on Learning Representations (ICLR)*, 2025. URL https://openreview.net/forum?id=iMJpmcYucq.

Kanti V. Mardia and Peter E. Jupp. *Directional Statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester, 2000. ISBN 0-471-95333-4.

Ricardo Marques, Christian Bouville, Mickaël Ribardière, Luís Paulo Santos, and Kadi Bouatouch. Spherical Fibonacci point sets for illumination integrals. *Computer Graphics Forum*, 32(8): 134–143, 2013. doi: 10.1111/cgf.12190.

Jiří Matoušek. On the L_2-discrepancy for anchored boxes. *Journal of Complexity*, 14(4):527–556, 1998. doi: 10.1006/jcom.1998.0489.

Olena Melnikov and Johannes Milz. Randomized quasi-Monte Carlo methods for risk-averse stochastic optimization. *Journal of Optimization Theory and Applications*, 206(1):14, 2025. doi: 10.1007/s10957-025-02693-6.

Hrushikesh N. Mhaskar, Francis J. Narcowich, and Joseph D. Ward. Spherical Marcinkiewicz–Zygmund inequalities and positive quadrature. *Mathematics of Computation*, 70(235):1113–1130, 2001.

Harald Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*, volume 63 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1992. ISBN 0-89871-295-5. doi: 10.1137/1.9781611970081.

Art B. Owen. Randomly permuted (t, m, s)-nets and (t, s)-sequences. In Harald Niederreiter and Peter Jau-Shyong Shiue (eds.), *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, volume 106 of *Lecture Notes in Statistics*, pp. 299–317. Springer, New York, 1995. doi: 10.1007/978-1-4612-2552-2_19.

Art B. Owen. Scrambled net variance for integrals of smooth functions. *The Annals of Statistics*, 25 (4):1541–1562, 1997a. doi: 10.1214/aos/1031594731.

Art B. Owen. Monte Carlo variance of scrambled net quadrature. *SIAM Journal on Numerical Analysis*, 34(5):1884–1910, 1997b. doi: 10.1137/S0036142994277468.

Art B. Owen. Scrambling Sobol' and Niederreiter–Xing points. *Journal of Complexity*, 14(4): 466–489, 1998. doi: 10.1006/jcom.1998.0487.

Art B. Owen. Variance with alternative scramblings of digital nets. *ACM Transactions on Modeling and Computer Simulation*, 13(4):363–378, 2003. doi: 10.1145/945511.945518.

Art B. Owen. Multidimensional variation for quasi-Monte Carlo. In *Contemporary Multivariate Analysis and Design of Experiments*, volume 2 of *Series in Biostatistics*, pp. 49–74. World Scientific, Hackensack, NJ, 2005. doi: 10.1142/9789812567765_0004.

Art B. Owen. Halton sequences avoid the origin. *SIAM Review*, 48(3):487–503, 2006a. doi: 10.1137/S0036144504441573.

Art B. Owen. Quasi-Monte Carlo for integrands with point singularities at unknown locations. In Harald Niederreiter and Denis Talay (eds.), *Monte Carlo and Quasi-Monte Carlo Methods 2004*, pp. 403–417. Springer, Berlin, 2006b. doi: 10.1007/3-540-31186-6_24.

Art B. Owen. Local antithetic sampling with scrambled nets. *The Annals of Statistics*, 36(5): 2319–2343, 2008. doi: 10.1214/07-AOS548.

Art B. Owen and Daniel Rudolf. A strong law of large numbers for scrambled net integration. *SIAM Review*, 63(2):360–372, 2021. doi: 10.1137/20M1320535.

Teemu Pennanen and Matti Koivu. Epi-convergent discretizations of stochastic programs via integration quadratures. *Numerische Mathematik*, 100(1):141–163, 2005. doi: 10.1007/s00211-004-0571-4.

Xavier Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006. doi: 10.1007/s10851-006-6228-4.

Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006. doi: 10.1007/s11263-005-3222-z.

Bernhard Riemann. On the hypotheses which lie at the bases of geometry. *Nature*, 8:14–17, 36–37, 1873. Habilitation lecture, Göttingen, 1854; translated by W. K. Clifford.

K. F. Roth. On irregularities of distribution. *Mathematika*, 1(2):73–79, 1954. doi: 10.1112/S0025579300000541.

Hiroyuki Sato, Hiroyuki Kasai, and Bamdev Mishra. Riemannian stochastic variance reduced gradient algorithm with retraction and vector transport. *SIAM Journal on Optimization*, 29(2):1444–1472, 2019. doi: 10.1137/17M1116787.

Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*, volume 9 of *MPS/SIAM Series on Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2009. ISBN 978-0-89871-687-0. doi: 10.1137/1.9780898718751.

Maurice Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958. doi: 10.2140/pjm.1958.8.171.

I. M. Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4):86–112, 1967. doi: 10.1016/0041-5553(67)90144-9.

Kenneth B. Stolarsky. Sums of distances between points on a sphere. II. *Proceedings of the American Mathematical Society*, 41(2):575–582, 1973. doi: 10.1090/S0002-9939-1973-0333995-9.

Karl-Theodor Sturm. Probability measures on metric spaces of nonpositive curvature. In *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces*, volume 338 of *Contemporary Mathematics*, pp. 357–390. American Mathematical Society, Providence, RI, 2003. ISBN 0-8218-3383-9. doi: 10.1090/conm/338/06080.

Yann Thanwerdas and Xavier Pennec. Theoretically and computationally convenient geometries on full-rank correlation matrices. *SIAM Journal on Matrix Analysis and Applications*, 43(4): 1851–1872, 2022. doi: 10.1137/22M1471729.

Hans Triebel. *Theory of Function Spaces II*, volume 84 of *Monographs in Mathematics*. Birkhäuser, Basel, 1992. doi: 10.1007/978-3-0346-0419-2. Chapter 7: function spaces on Riemannian manifolds, defined through local charts and partitions of unity.

Nilesh Tripuraneni, Nicolas Flammarion, Francis Bach, and Michael I. Jordan. Averaging stochastic gradient descent on Riemannian manifolds. In *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pp. 650–687. PMLR, 2018. URL https://proceedings.mlr.press/v75/tripuraneni18a.html.

Jeffrey D. Vaaler. Some extremal functions in Fourier analysis. *Bulletin of the American Mathematical Society (N.S.)*, 12(2):183–216, 1985. doi: 10.1090/S0273-0979-1985-15349-2.

Anna Yershova, Swati Jain, Steven M. LaValle, and Julie C. Mitchell. Generating uniform incremental grids on $SO(3)$ using the Hopf fibration. *The International Journal of Robotics Research*, 29(7): 801–812, 2010. doi: 10.1177/0278364909352700.

Hongyi Zhang and Suvrit Sra. First-order methods for geodesically convex optimization. In *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pp. 1617–1638. PMLR, 2016. URL https://proceedings.mlr.press/v49/zhang16b.html.

Hongyi Zhang, Sashank J. Reddi, and Suvrit Sra. Riemannian SVRG: Fast stochastic optimization on Riemannian manifolds. In *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, 2016. URL https://papers.nips.cc/paper_files/paper/2016/hash/98e6f17209029f4ae6dc9d88ec8eac2c-Abstract.html.

Pan Zhou, Xiao-Tong Yuan, and Jiashi Feng. Faster first-order methods for stochastic non-convex optimization on Riemannian manifolds. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 89 of *Proceedings of Machine Learning Research*, pp. 138–147. PMLR, 2019. URL https://proceedings.mlr.press/v89/zhou19a.html. Journal version: IEEE Trans. Pattern Anal. Mach. Intell. 43(2):459–472, 2021.

APPENDIX CONTENTS

How the appendix is organized **18**

A Notation and the dependence of results **20**
A.1 Notation . . . 20
A.2 Dependence of results . . . 21

B Preliminaries **23**
B.1 Riemannian optimization . . . 23
B.2 Quasi-Monte Carlo integration . . . 24

C The field of a point set: readings, budgets and proofs **28**
C.1 Setting, potential and field . . . 28
C.2 The three readings . . . 31
C.3 Reuse: one point set along the optimization path . . . 31
C.4 Solve: minimizers of the sampled objective . . . 33
C.5 Refresh: Riemannian SGD with scrambled minibatches . . . 35
C.6 Sturm's contraction as a reading of the field . . . 40

D The scaling law: proofs **41**
D.1 The sandwich and its exactness . . . 41
D.2 Dilation . . . 42
D.3 The law on spheres . . . 43
D.4 The law on the square . . . 46
D.5 Sharpness . . . 47
D.6 The floor and the designs . . . 51

E Geometry: where the field integrand is singular **52**
E.1 Transported integrands . . . 52
E.2 No cut locus: Hadamard manifolds . . . 55
E.3 No cut locus: the Karcher ball . . . 56
E.4 The sphere is electrostatics . . . 57
E.5 Charts are blow-ups . . . 60

F Designing the charge: Riemannian optimization of node sets **63**

G Numerical experiments **64**
G.1 Principal subspaces on the Grassmannian . . . 64
G.2 The Grassmannian with a bounded transport: same objective, regular integrand . . . 67
G.3 Karcher means on the affine-invariant positive definite cone . . . 71

G.4 The sphere: Stolarsky nodes, transported nets and the cut locus 73
G.5 The optimized nodes: orientation, maximal errors and cost 75
G.6 The chart decides: resolved and unresolved cut points 79
G.7 Rules exact to high degree sit at the floor 80
G.8 Checks of the one-point law, the dipole scalings and the value constant 80
G.9 Finite-precision scrambling and endpoints 82
G.10 Implementation details 83

H Related work and open directions **85**
H.1 Additional related work 85
H.2 Open directions 86

I The epigraph and the method **87**

How the appendix is organized

The appendix follows the architecture of the main text: one object, the discrepancy charge $\rho_N = Q_N - P$ with its potential and its field (Definition 2.1); one law, the exponent $(m+c)/(c+\beta)$ of (2); and the settings in which the law is used. Every statement of the main text is proved here, each proof follows its statement, and every cross-reference is a link. Appendix A collects the notation in one table and draws the dependency graph of the results (Figure 5), which is the quickest way to see what a given proof uses. The paragraphs below describe the sections in order; the last two suggest three ways of reading them and list where each statement of the main text is proved.

Preliminaries (Appendix B). The Riemannian half recalls geodesic convexity and smoothness (Definitions B.1 and B.2), the two consequences (7) of strong convexity that the minimizer bounds use, the comparison inequality of Lemma B.3 and the Hessian comparison (10). The quadrature half recalls the star discrepancy, Hardy–Krause variation and the Koksma–Hlawka inequality (12); the four properties (S1)–(S4) of scrambled nets; the variance equivalence of the linear matrix scrambling implemented in SciPy and the nested scrambling of the theory (Lemma B.4) and the effect of scrambling finitely many digits (Lemma B.5); and QMC on spheres, where Stolarsky's kernel (Lemma B.6) serves twice: it makes the cap discrepancy a worst-case error, and it makes the Fréchet problem on the sphere a problem of electrostatics (Appendix E.4).

The field of a point set (Appendix C). This section treats the object of the main text in full generality. The field integrand takes values in a tangent space, and the frame-free Hardy–Krause variation of Definition C.1 carries the Koksma–Hlawka inequality to every tangent space at once (Lemma C.2 and Theorem C.3). Proposition C.7 proves the three readings of Proposition 2.2. Each optimization result is then a standard Riemannian argument applied to the sampled objective, followed by one reading: reusing one point set along gradient descent needs the uniform reading (Theorem C.9 and Corollary C.11); solving the sampled problem once needs only the one-point reading (Theorems C.12 and C.13); refreshing a scrambled net at every step of Riemannian SGD needs the mean-square reading, which Lemma C.14 reduces to scalar variance bounds (Theorems C.15 and C.18, with the budget ladder of Corollary C.17 and Table 4). The section ends with Sturm's contraction of barycentres, the one-point reading with the Wasserstein certificate (Proposition C.21).

The scaling law (Appendix D). Appendix D.1 proves the sandwich (Lemma D.1) and its exactness by Sion's minimax theorem (Proposition D.2): upper bounds come from sandwiches, lower bounds from measures, and at the optimum they meet. Appendix D.2 contains the two computations behind every exponent: a dilation changes a Sobolev norm by a power of the scale (Lemma D.3), and a homogeneous singularity cut at scale r is a sum of dilated dyadic shells (Lemma D.4). Appendices D.3 and D.4 prove the law on spheres in Sobolev norms and in the square in Hardy–Krause variation (Theorems D.6 and D.10) and derive the certificates of the Fréchet problem on spheres (Corollaries D.7 and D.8); Figure 6 draws the proof as a balance between a cut and a bump. Appendix D.5 shows that

no bound in terms of the discrepancy alone does better, with multipoles on spheres (Theorem D.11 and Figure 7) and discrepancy functions in the square (Theorem D.12 and Figure 8). Appendix D.6 proves the floor of order $1/N$ that a single node sets (Proposition D.13) and shows that rules equidistributed to every order meet it up to N^{ϵ} (Corollary D.14).

Geometry (Appendix E). This section locates the singularities of the field integrand, and so decides which row of Table 1 applies. Transports with integrable coordinate singularities cost nothing (Lemma E.1 and Proposition E.2). On Hadamard manifolds and inside the Karcher radius there is no cut locus to meet, and Fréchet means of transported low-discrepancy sets converge at the rate of the star discrepancy (Theorems E.3 and E.6). On spheres with full support the whole singularity of the Fréchet loss is the chordal distance to the antipode (Propositions E.8 and E.9 and Figure 9), a chart resolves the cut points it blows up (Proposition E.10), and flat tori with oblique lattices realize the row of codimension one (Proposition E.11).

Designing the charge (Appendix F). On a sphere, node sets are computed rather than enumerated. By Stolarsky's principle, Riemannian ascent on the sum of distances lowers the energy of the discrepancy charge and evaluates the attained cap discrepancy exactly at every iterate (Proposition F.1); the section ends with what that energy certifies about fields.

Experiments (Appendix G). Every experiment tests a statement, as listed in Table 5: the one-point reading on the positive definite cone (Appendix G.3), the readings and the role of the transport on the Grassmannian (Appendices G.1 and G.2), the charts and the floor on the sphere (Appendices G.4–G.6), rules exact to high degree (Appendix G.7), exact checks of the one-point law and of the constructions behind the sharpness theorems (Appendix G.8), and finite-precision scrambling (Appendix G.9). Appendix G.10 gives the implementation details, seeds, tolerances and reference computations.

Related work and open directions (Appendix H). The last section places the paper among work on QMC in stochastic programming and variational inference, Riemannian stochastic optimization, QMC on manifolds and energy-minimizing point sets, and it states the open directions that the conditions of the present theorems leave.

The epigraph (Appendix I). The last section explains how Riemann's sentence below the abstract describes the method of the paper: the rate is decided in the infinitely small neighbourhood of the singular set, the lower bounds show that this analysis cannot be avoided, and a blow-up follows the singularity into that neighbourhood and resolves it.

Three ways to read the appendix. For the scaling law alone: Appendices D.1–D.3, then Appendices D.5, D.6 and E.4, with Figures 6, 7 and 9 beside the proofs. For the optimization guarantees: Appendices C.1 and C.2, then Appendices C.3–C.5, and Appendices E.2 and E.3 for Fréchet means. For the experiments: Table 5 and Appendix G, whose subsections each name the statement they test, and the checks of Appendix G.8, which evaluate the lower-bound constructions exactly.

Table 2: **Where the statements of the main text are proved.**

main text	content	appendix
Definition 2.1	charge, potential, field	Appendix C.1, (17)
Proposition 2.2	three readings of the field	Proposition C.7
Corollary 2.3	budgets of reuse, refresh, solve	Theorems C.9, C.13 and C.15 and Corollaries C.11 and C.17
Lemma 3.1	sandwich	Lemma D.1
Proposition 3.2	the sandwich is exact	Proposition D.2
Theorem 3.3	scaling law	Theorems D.6 and D.10
Theorem 3.4	sharpness and floor	Theorems D.11 and D.12 and Proposition D.13
Corollary 3.5	designs meet the floor	Corollary D.14
Proposition 4.1	the sphere is electrostatics	Proposition E.9
Proposition 4.2	charts are blow-ups	Proposition E.10
Table 1, row 1	no cut locus	Theorems E.3 and E.6
Table 1, rows 2–3	the antipode on $\mathbb{S}^d$	Corollary D.7, Theorem D.11, and Proposition E.9
Table 1, rows 4–5	polar charts	Proposition E.10 and Theorems D.10 and D.12
Table 1, row 6	flat tori	Proposition E.11 and Theorems D.10 and D.12

A NOTATION AND THE DEPENDENCE OF RESULTS

A.1 NOTATION

Table 3 lists the symbols used in more than one place, grouped by the part of the paper that introduces them; the last column links to the definition. A few letters serve twice, as in the literature they come from: d is the geodesic distance and the dimension of $\mathbb{S}^d$, s the dimension of the cube and the singular part of an integrand, σ the surface measure and a standard deviation, and κ a curvature bound and, in Table 1 only, an exponent. The context always decides.

Conventions. C and c denote positive constants that may change from line to line and depend only on the quantities named in each statement; $\widetilde{O}(\cdot)$ is $O(\cdot)$ up to powers of logarithms, and $a \asymp b$ means that a/b is bounded above and below by positive constants. A positive rule has nonnegative weights that sum to one. Equations are numbered consecutively with those of the main text, and results are numbered within each appendix.

Table 3: **Notation.**

symbol	meaning	defined in
The decision manifold		
$\mathcal{M}$, n	complete connected Riemannian manifold of decisions; its dimension	Appendix B.1
$d(x, y)$	geodesic distance	Appendix B.1
Exp_x, Log_x	exponential map; its inverse wherever the minimizing geodesic is unique, so that $\|\mathrm{Log}_x\, y\| = d(x, y)$	Appendix B.1
$\mathrm{Cut}(x)$, $\mathrm{inj}(\mathcal{M})$	cut locus of x, which is $\{-x\}$ on $\mathbb{S}^d$; injectivity radius	Appendix B.1
grad, Hess, $P_{x\to y}$	Riemannian gradient and Hessian; parallel transport along the minimizing geodesic	Appendix B.1
μ, μ_0; L, L_0	moduli of strong geodesic convexity of f and of f_N; smoothness constants	Definitions B.1 and B.2
$-\kappa$, Δ	lower and upper bounds on the sectional curvature	(10)
$\mathrm{Gr}(n, p)$, $\mathcal{P}_n$	Grassmannian of p-planes in $\mathbb{R}^n$; cone of positive definite matrices with the affine-invariant metric	Examples E.4 and C.6
Integration, rules and certificates		
(Ω, P)	integration space: the cube $[0,1]^s$ with Lebesgue measure λ, the sphere $\mathbb{S}^d$ with its normalized surface measure σ, or $\mathcal{M}$ itself	Section 2
s, u	dimension of the cube; integration variable $u \in [0,1]^s$	Appendix C.1
Q_N, P_N, w_i	rule $\sum_i w_i \delta_{\omega_i}$, with $w_i = 1/N$ unless stated; its point set; its weights. A *positive rule* has $w_i \geq 0$ and $\sum_i w_i = 1$	Section 2, Appendix D
H, $\|\cdot\|_H$	certificate: a class of functions on Ω with a seminorm that vanishes on constants	Section 2
$D_H(Q)$	discrepancy of Q in H, $\sup\{\|Qg - Pg\| : \|g\|_H \leq 1\}$	Section 2
D_N^*, V_{HK}	star discrepancy; Hardy–Krause variation, frame-free for integrands with values in a tangent space	Appendix B.2, Definition C.1
$H^\tau(\mathbb{S}^d)$, wce_{H^τ}	Sobolev space of order τ on the sphere; worst-case error of a rule on its unit ball	Appendix B.2
D_{L_2}, $\mathcal{D}(\varrho)$	L_2 spherical cap discrepancy of a node set; the same for a signed measure ϱ	Appendix B.2
K_d, C_d	Stolarsky's kernel $1 - \|y - z\|$ /C_d; its constant, $C_2 = 4$	Lemma B.6
QMC design	node sets with $\mathrm{wce}_{H^\tau} \leq c_\tau N^{-\tau/d}$, the optimal order	Appendix B.2
(S1)–(S4), $\Gamma_{b,t,s}$	properties of scrambled (t, m, s)-nets in base b; the constant of their variance bound	Appendix B.2
$\widehat{I}_N(g)$, $\widehat{g}_N(x)$	equal-weight estimate of $\int g$; scrambled-net estimate of $\mathrm{grad}\, f(x)$	Appendix B.2, Appendix C.5
$\sigma_F^2(x)$, $\sigma_N^2(x)$	variance of the gradient integrand at x; mean squared error of its estimate	Lemma C.14, Theorem C.13
B, ϵ_B	number of randomized digits of an implemented scrambling; the deterministic error it adds	Lemma B.5

continued on the next page

Table 3, continued

symbol	meaning	defined in
The charge, its potential and its field		
$F(x,\omega)$, h	loss; a density weight absorbed into it, as in $F(x,y)=h(y)\frac{1}{2}d(x,y)^2$	Section 2, Corollary D.7
f, f_N	population objective PF; sampled objective Q_NF	(1), (15)
$x^\star$, $\hat{x}_N$	population minimizer; minimizer or stationary point of f_N	Proposition 2.2
ρ_N	discrepancy charge $Q_N - P$	Definition 2.1
Φ_N	potential $\rho_N F(x,\cdot) = f_N - f$, the value error	Definition 2.1
e_N	field $\operatorname{grad}\Phi_N = \rho_N G_x$, the gradient error	Definition 2.1, (17)
G_x, $\|G_x\|_H$	field integrand $\operatorname{grad}_x F(x,\cdot)\colon \Omega \to T_x\mathcal{M}$; its certificate norm $\sup_{\|v\|=1}\|\langle G_x, v\rangle\|_H$	Definition 2.1
$V_0(S)$, $V_1(S)$	suprema over $x \in S$ of the certificate norms of $F(x,\cdot)$ and of G_x	Assumption C.2, Proposition C.7
$\bar{H}_N$	Hessian of f_N averaged along the geodesic from $x^\star$ to $\hat{x}_N$ ($\bar{H}$ in Proposition C.7)	Proposition 2.2
U_ϱ, R_x, S_x	chordal potential $\int \|z-y\|\,\mathrm{d}\varrho(y)$ of a signed measure; smooth remainders of the value and field integrands on the sphere	Proposition E.9
Δ_0, ε, K, η	initial gap $f(x_0) - \inf f$; target accuracy; number of iterations; step size	Theorem C.9, Appendix C.5
The scaling law		
m	order of a singularity, its degree of homogeneity: 0 for a direction that jumps, 1 for a cone	(2)
c	codimension of the singular set	(2)
β	dilation exponent: a bump of width r has certificate norm of order $r^{-\beta}$; $\tau - d/2$ in $H^\tau(\mathbb{S}^d)$, 0 in V_{HK} at a point, 1 along a segment	(2), Section 3
p, z	singular point; normal coordinates centred at p, with $p=-x$ for the Fréchet loss	Theorem D.6, (30)
b, ω	smooth amplitude and angular profile of a singularity $b(z)\,\|z\|^m\,\omega(z/\|z\|)$	Theorem D.6, Theorem D.10
$g_\pm$, $\mathcal{S}_g$	minorant and majorant of g in $H+\mathbb{R}$; the best one-sided bound at a given discrepancy	Lemma D.1
r, s^{out}, s^{in}, ϑ_r^p	cut scale; cut part and discarded part of a singular function; the bump that dominates the discarded part	Theorem D.6
α, $\ell_\alpha(r)$	$\alpha = m + d/2 - \tau$; cost of the dyadic shells: r^α, $1+\log(1/r)$ or 1 as $\alpha<0$, $=0$, >0	Lemma D.4
q, ℓ, $\Lambda(q)$	multipole profile whose moments of order below ℓ vanish; the moment that pairs it with the singularity	Theorem D.11
Δ_ν	discrepancy function $\nu([0,a)) - a_1a_2$ of a measure on the square	Appendix D.5
Y_N	node set $\{y_1,\dots,y_N\}$ on the sphere	Proposition D.13

A.2 Dependence of results

Figure 5 shows which results each proof uses. The optimization results of Appendix C rest on one inequality, Koksma–Hlawka in the form Lemma C.2 for tangent vectors, or on its randomized counterpart Lemma C.14; each of them is a standard Riemannian argument followed by one reading of the field. The scaling law of Appendix D rests on two computations: every upper bound is a sandwich (Lemma D.1), and on spheres the cost of the sandwich is the sum over dyadic shells of Lemma D.4, a consequence of the dilation lemma; the lower bounds are the dual half of Proposition D.2. The geometry of Appendix E draws on both: the Fréchet means on Hadamard manifolds and in the Karcher ball on the readings, the charts and the flat tori on the law in the square, and the electrostatics of the sphere on Stolarsky's kernel and the shells.

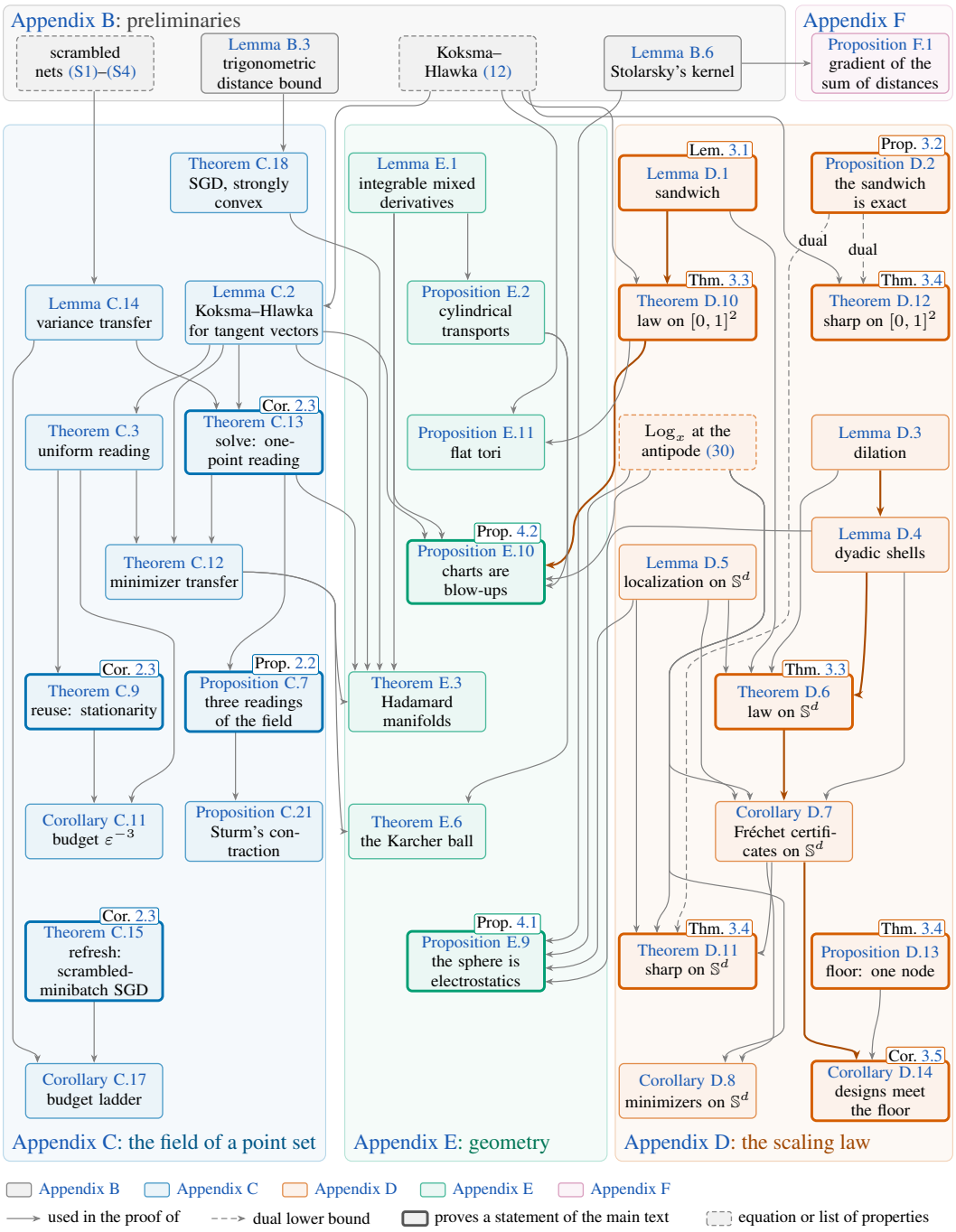

Figure 5: **How the results depend on each other.** An arrow from A to B means that the proof of B uses A; every node is a link to the result it names. The two heavier paths are the backbones of the scaling law on spheres and in the square. A tag on a node gives the statement of the main text that the result proves. The dashed arrows record that the sharpness theorems are the lower-bound half of the duality of Proposition D.2. The geometry column sits between the other two because its results use both. The field column rests on one inequality, (12) in the form Lemma C.2, and on its randomized counterpart Lemma C.14.

B PRELIMINARIES

B.1 RIEMANNIAN OPTIMIZATION

Throughout, $\mathcal{M}$ is a complete connected Riemannian manifold of dimension n with metric $\langle \cdot, \cdot \rangle_x$ on $T_x\mathcal{M}$, norm $\|\cdot\|_x$ (we drop the subscript when no confusion arises), and geodesic distance d. We write $\operatorname{Exp}_x \colon T_x\mathcal{M} \to \mathcal{M}$ for the exponential map, $\operatorname{inj}(\mathcal{M}) \in [0, \infty]$ for the injectivity radius (positive when $\mathcal{M}$ is compact; it can vanish on a complete noncompact manifold, and results that need it positive say so), and, whenever y is joined to x by a unique minimizing geodesic, $\operatorname{Log}_x(y) \in T_x\mathcal{M}$ for its initial velocity, so that $\|\operatorname{Log}_x(y)\| = d(x, y)$. $P_{x \to y}$ denotes parallel transport along that geodesic. A *Hadamard manifold* is a complete simply connected manifold of nonpositive sectional curvature; on it Exp_x is a diffeomorphism for every x, Log_x is globally defined and smooth, $d(\cdot, y)^2$ is smooth, closed geodesic balls are geodesically convex, and the metric projection Π_C onto a closed geodesically convex set C is well defined and nonexpansive, $d(\Pi_C x, \Pi_C y) \le d(x, y)$ (Bačák, 2014). The cut locus $\operatorname{Cut}(x)$ is the set of points beyond which geodesics from x stop being minimizing; Log_x is smooth on $\mathcal{M} \setminus (\operatorname{Cut}(x) \cup \{x\})$ and $d(x, y) < \operatorname{inj}(\mathcal{M})$ implies $y \notin \operatorname{Cut}(x)$. For a C^1 function f, $\operatorname{grad} f(x) \in T_x\mathcal{M}$ is the Riemannian gradient; for $f \in C^2$, $\operatorname{Hess} f(x)$ is the Riemannian Hessian, a self-adjoint operator on $T_x\mathcal{M}$. A point x is *ε-stationary* if $\|\operatorname{grad} f(x)\| \le \varepsilon$; randomized methods will certify the mean-square version $\mathbb{E}\,\|\operatorname{grad} f(x)\|^2 \le \varepsilon^2$ for a random iterate. Standard references are do Carmo (1992); Lee (2018) for the geometry and Absil et al. (2008); Boumal (2023) for optimization.

Definition B.1 (geodesic convexity)**.** A set $C \subset \mathcal{M}$ is *geodesically convex* if any two points of C are joined by a unique minimizing geodesic of $\mathcal{M}$ and that geodesic lies in C. A function $f \colon C \to \mathbb{R}$ is *μ-strongly geodesically convex* ($\mu \ge 0$; *geodesically convex* if $\mu = 0$) if for every such geodesic $\gamma \colon [0, 1] \to C$, $f(\gamma(t)) \le (1 - t) f(\gamma(0)) + t f(\gamma(1)) - \frac{\mu}{2} t(1 - t)\, d(\gamma(0), \gamma(1))^2$. For differentiable f this is equivalent to

$$f(y) \;\ge\; f(x) + \langle \operatorname{grad} f(x), \operatorname{Log}_x(y) \rangle + \tfrac{\mu}{2}\, d(x, y)^2 \qquad \text{for all } x, y \in C. \tag{6}$$

Two consequences of (6) will be used repeatedly. If $x^\star \in C$ satisfies $\operatorname{grad} f(x^\star) = 0$, then for all $y \in C$

$$f(y) - f(x^\star) \;\ge\; \tfrac{\mu}{2}\, d(y, x^\star)^2, \qquad \|\operatorname{grad} f(y)\| \;\ge\; \mu\, d(y, x^\star). \tag{7}$$

The first is (6) with $x = x^\star$. For the second, take $x = y$ and $y = x^\star$ in (6) and use the first consequence:

$$-\big\langle \operatorname{grad} f(y), \operatorname{Log}_y(x^\star) \big\rangle \;\ge\; f(y) - f(x^\star) + \tfrac{\mu}{2} d(y, x^\star)^2 \;\ge\; \mu\, d(y, x^\star)^2,$$

while the left side is at most $\|\operatorname{grad} f(y)\|\; d(y, x^\star)$ by the Cauchy–Schwarz inequality.

Definition B.2 (smoothness)**.** Let $S \subset \mathcal{M}$. A C^1 function f is *L-smooth on S* if

$$f(\operatorname{Exp}_x(v)) \;\le\; f(x) + \langle \operatorname{grad} f(x), v \rangle + \tfrac{L}{2}\, \|v\|^2 \qquad \text{for all } x \in S,\; v \in T_x\mathcal{M}. \tag{8}$$

If C is geodesically convex, f has an *L-Lipschitz gradient on C* if $\|\operatorname{grad} f(y) - P_{x \to y} \operatorname{grad} f(x)\| \le L\, d(x, y)$ for all $x, y \in C$.

If $f \in C^2$ and $\|\operatorname{Hess} f(z)\| \le L$ for all $z \in \mathcal{M}$, then f is L-smooth on $\mathcal{M}$: for the geodesic $\gamma(t) = \operatorname{Exp}_x(tv)$, Taylor's formula with integral remainder gives

$$f(\gamma(1)) - f(x) - \langle \operatorname{grad} f(x), v \rangle \;=\; \int_0^1 (1 - t)\, \langle \operatorname{Hess} f(\gamma(t))\dot\gamma(t), \dot\gamma(t) \rangle \, \mathrm{d}t \;\le\; \frac{L}{2}\, \|v\|^2,$$

and the same Hessian bound on a geodesically convex C gives an L-Lipschitz gradient on C (Boumal, 2023). If $\operatorname{grad} f$ is L-Lipschitz on C and $\operatorname{grad} f(x^\star) = 0$ for some $x^\star \in C$, then $\|\operatorname{grad} f(y)\| \le L\, d(y, x^\star)$ on C. Finally, (8) with $v = -\eta \operatorname{grad} f(x)$ is the *descent lemma*

$$f\big(\operatorname{Exp}_x(-\eta \operatorname{grad} f(x))\big) \;\le\; f(x) - \eta\big(1 - \tfrac{L\eta}{2}\big)\, \|\operatorname{grad} f(x)\|^2. \tag{9}$$

We also need the comparison inequality of Zhang & Sra (2016) that replaces the Euclidean law of cosines in negative curvature.

Lemma B.3 (trigonometric distance bound, Zhang & Sra, 2016, Lemma 5). *Let $\mathcal{M}$ be a Hadamard manifold whose sectional curvatures are bounded below by $-\kappa$, $\kappa \geq 0$. Consider a geodesic triangle with side lengths a, b, c and let A be the angle opposite a, that is, the angle between the sides of lengths b and c. Then*

$$a^2 \;\leq\; \zeta(\kappa, c)\, b^2 + c^2 - 2bc\cos A, \qquad \zeta(\kappa, c) = \frac{\sqrt{\kappa}\, c}{\tanh(\sqrt{\kappa}\, c)} \quad (\zeta(0, c) = 1).$$

$\zeta(\kappa, \cdot)$ is nondecreasing, so $\zeta(\kappa, c) \leq \zeta(\kappa, c')$ whenever $c \leq c'$.

Lastly, the Hessian of the squared distance obeys the classical comparison bounds (Karcher, 1977; Afsari, 2011): if the sectional curvatures of $\mathcal{M}$ lie in $[-\kappa, \Delta]$ with $\kappa \geq 0$, $\Delta > 0$, and y is not in the cut locus of x, $\ell = d(x, y) < \pi/\sqrt{\Delta}$, then the eigenvalues of $\mathrm{Hess}_x\, \frac{1}{2} d(\cdot, y)^2$ lie in

$$\Big[\, \sqrt{\Delta}\ell \cot(\sqrt{\Delta}\ell),\ \sqrt{\kappa}\ell \coth(\sqrt{\kappa}\ell) \,\Big], \tag{10}$$

with the convention $\sqrt{\kappa}\ell \coth(\sqrt{\kappa}\ell) = 1$ if $\kappa = 0$ and $\sqrt{\Delta}\ell \cot(\sqrt{\Delta}\ell) = 1$ if $\Delta = 0$; the radial direction has eigenvalue exactly 1 (Lezcano-Casado, 2020, Thm. 3.11 and proof of Thm. 3.15).

B.2 Quasi-Monte Carlo integration

Let $P_N = \{u_1, \ldots, u_N\} \subset [0, 1)^s$ be a point set (multiset). Its *star discrepancy* is

$$D_N^*(P_N) = \sup_{a \in [0,1]^s} \Big| \frac{1}{N} \#\{i : u_i \in [0, a)\} - \prod_{j=1}^{s} a_j \Big|.$$

For the first N points of the Halton sequence (Halton, 1960), of a Sobol' sequence (Sobol', 1967; Joe & Kuo, 2008) or of any (t, s)-sequence, $D_N^*(P_N) \leq C_s\, N^{-1} (\log N)^s$ for $N \geq 2$; for a (t, m, s)-net in base b with $N = b^m$ points, $D_N^*(P_N) \leq C_{t,s,b}\, N^{-1} (\log N)^{s-1}$ (Niederreiter, 1992; Dick & Pillichshammer, 2010). No point set can do better than $c_s N^{-1} (\log N)^{(s-1)/2}$: this is Roth's bound for the L_2 discrepancy (Roth, 1954), which transfers to D_N^* since D_N^* dominates the L_2 discrepancy (sharper lower bounds for D_N^* are known in dimension $s \geq 3$). We write $\widetilde{O}(\cdot)$ for $O(\cdot)$ up to powers of $\log N$ or $\log(1/\varepsilon)$.

Hardy–Krause variation and the Koksma–Hlawka inequality. For $g \colon [0, 1]^s \to \mathbb{R}$ and a subbox $R = \prod_j [a_j, b_j]$, let $\Delta(g; R)$ be the alternating sum of g over the 2^s vertices of R. The *Vitali variation* $V^{(s)}(g)$ is the supremum of $\sum_R |\Delta(g; R)|$ over all partitions of $[0, 1]^s$ into subboxes, and the *Hardy–Krause variation* is

$$V_{\mathrm{HK}}(g) = \sum_{\emptyset \neq \alpha \subseteq \{1, \ldots, s\}} V^{(|\alpha|)}\big(g(\,\cdot_\alpha, \mathbf{1}_{-\alpha})\big),$$

the sum of the Vitali variations of the restrictions of g to the faces on which the coordinates outside α are set to 1 (Owen, 2005). If all mixed partial derivatives $\partial^\alpha g = \prod_{j \in \alpha} \partial_{u_j} g$ are continuous on the closed cube, then

$$V_{\mathrm{HK}}(g) = \sum_{\emptyset \neq \alpha} \int_{[0,1]^{|\alpha|}} \big|\partial^\alpha g(u_\alpha, \mathbf{1}_{-\alpha})\big|\, \mathrm{d}u_\alpha \;\leq\; \sum_{\emptyset \neq \alpha} \|\partial^\alpha g\|_\infty . \tag{11}$$

The Koksma–Hlawka inequality (Koksma, 1942; Hlawka, 1961; Kuipers & Niederreiter, 1974) states that for every point set P_N and every g of bounded Hardy–Krause variation,

$$\Big| \frac{1}{N} \sum_{i=1}^{N} g(u_i) - \int_{[0,1]^s} g(u)\, \mathrm{d}u \Big| \;\leq\; V_{\mathrm{HK}}(g)\, D_N^*(P_N). \tag{12}$$

It is a deterministic inequality, and this is the property that the uniform reading of the field exploits: it holds simultaneously for every member of a family of integrands whose variations are uniformly bounded.

Scrambled nets. Owen's nested uniform scrambling (Owen, 1995; 1997b; 1998) of a (t, m, s)-net in base b with $N = b^m$ points produces a random point set $\widetilde{P}_N$ which is again a (t, m, s)-net with probability one and each of whose points is uniformly distributed on $[0,1]^s$. Consequently, for $\widehat{I}_N(g) = \frac{1}{N}\sum_i g(u_i)$:

(S1) $\mathbb{E}[\widehat{I}_N(g)] = \int g$ for every $g \in L^1([0,1]^s)$;

(S2) for every $g \in L^2([0,1]^s)$ with variance $\sigma^2(g)$, $\mathrm{Var}(\widehat{I}_N(g)) \le \Gamma_{b,t,s}\,\sigma^2(g)/N$ with $\Gamma_{b,t,s} = b^t\big(\frac{b+1}{b-1}\big)^s$ (Owen, 1998; Owen & Rudolf, 2021) (for $t = 0$ one may take $\Gamma = (b/(b-1))^s$ (Owen, 1997b)), and, for fixed g, $\mathrm{Var}(\widehat{I}_N(g)) = o(1/N)$ as $m \to \infty$ (Owen, 1997b; Gerber, 2015);

(S3) if g is smooth, in the sense that for every coordinate subset $\alpha \subseteq \{1, \dots, s\}$, the mixed derivative $\partial^\alpha g$ (one derivative per coordinate in α) is continuous on the closed cube, then for scrambled (t, m, s)-nets $\mathrm{Var}(\widehat{I}_N(g)) = O\big(N^{-3}(\log N)^{s-1}\big)$ (Owen, 1997a), the hypothesis in this form being due to Owen (2008);

(S4) for an integer $\alpha \ge 1$, take a digital $(t, m, \alpha s)$-net in a fixed prime base with bounded t, apply nested uniform scrambling, and interlace each group of α coordinates. If g has square-integrable mixed partial derivatives of order up to α in each variable, then $\mathrm{Var}(\widehat{I}_N(g)) = O(N^{-2\alpha-1+\delta})$ for every $\delta > 0$ (Dick, 2011; 2026).

Deterministic higher-order digital nets of order α (Dick, 2008) integrate the same class of functions with error $O(N^{-\alpha}(\log N)^{\alpha s})$, the constant being a Sobolev-type norm $\|g\|_\alpha$ of mixed dominating smoothness which reduces to a bound on V_{HK} when $\alpha = 1$. Finally, integrands that are unbounded near the boundary of the cube, such as compositions with the inverse Gaussian distribution function, are covered by Owen's boundary growth condition: if g is smooth on the open cube and $|\partial^\alpha g(u)| \le B\prod_{j=1}^s \min(u_j, 1-u_j)^{-A_j - \mathbf{1}\{j\in\alpha\}}$ for all α, with $0 < A_j < 1$, then for any randomized point set with $\mathbb{E}\,D_N^* = O(N^{-1+\epsilon})$ for all $\epsilon > 0$ (scrambled nets, randomized Halton points) the expected absolute error is $O(N^{-1+\max_j A_j+\epsilon})$ (Owen, 2006a, Thm. 5.7); the deterministic Halton sequence avoids the boundary and obeys analogous bounds (Owen, 2006a;b).

Linear matrix scrambling with a digital shift. Implementations such as `scipy.stats.qmc.Sobol` do not use nested uniform scrambling but the random linear scrambling of Matoušek (1998) followed by a digital shift, called affine matrix scrambling by Owen (2003): in a prime base b, the k-th digit of a scrambled coordinate is $x_k = \sum_{j\le k} M_{kj}a_j + C_k \pmod b$, where (a_j) are the digits of the unscrambled coordinate, the entries M_{kj} $(j < k)$ and C_k are independent and uniform on $\mathbb{Z}_b$, the diagonal entries M_{kk} are independent and uniform on $\mathbb{Z}_b \setminus \{0\}$, and independent copies are used for the s coordinates. Every scrambled point is then uniformly distributed (Owen, 2003, Prop. 3.1), so (S1) holds. The variance properties transfer as well: the variance of an equal-weight rule depends on the joint law of the point set only through the joint laws of pairs of points (Owen, 2003, Sec. 3), and these pair laws coincide for the two scramblings. The equivalence is used in the literature (e.g. Goda & Suzuki, 2026); we record the short argument because the experiments rely on it.

Lemma B.4 (variance equivalence of the two scramblings)**.** *Let $P_N \subset [0,1)^s$ be a point set in a prime base b and let $\widetilde{P}_N$ and $\widetilde{P}'_N$ be obtained from it by affine matrix scrambling and by nested uniform scrambling, respectively, each coordinate scrambled independently. Then $\mathrm{Var}(\widehat{I}_N(g))$ is the same for $\widetilde{P}_N$ and $\widetilde{P}'_N$ for every $g \in L^2([0,1]^s)$. In particular, (S2) and (S3) hold for affine matrix scrambling, and (S4) holds when the interlacing is applied after this scrambling.*

Proof. Since $\mathrm{Var}(\widehat{I}_N(g)) = N^{-2}\sum_{i,i'}\mathrm{Cov}(g(x_i), g(x_{i'}))$, it suffices to show that every pair $(x_i, x_{i'})$ has the same joint law under both scramblings, and since coordinates are scrambled independently, it suffices to treat one coordinate. Let $a = \sum_k a_k b^{-k}$ and $a' = \sum_k a'_k b^{-k}$ be the unscrambled coordinates of the two points and let k_0 be the first index with $a_{k_0} \ne a'_{k_0}$ ($k_0 = \infty$ if $a = a'$). Under nested uniform scrambling (Owen, 1995; 1997b), the digit pairs (x_k, x'_k) are independent across k; for $k < k_0$ they are equal and uniform, for $k = k_0$ they form a uniformly distributed ordered pair of distinct digits (one uniform permutation is applied to two distinct digits), and for $k > k_0$ they are independent and uniform (independent permutations are applied, because the digit prefixes differ).

Under affine matrix scrambling, the rows $(M_{k\cdot}, C_k)$ are independent across k, so the digit pairs are again independent across k; the independent uniform shift C_k makes x_k uniform and independent of $(M_{kj})_j$; and $x'_k - x_k = \sum_{j\le k} M_{kj}(a'_j - a_j)$ vanishes for $k < k_0$, equals $M_{k_0k_0}(a'_{k_0} - a_{k_0})$, which is uniform on the nonzero digits, for $k = k_0$, and for $k > k_0$ contains the summand $M_{kk_0}(a'_{k_0} - a_{k_0})$, with M_{kk_0} uniform on $\mathbb{Z}_b$ and independent of the remaining summands, hence is uniform on $\mathbb{Z}_b$; in every case $x'_k - x_k$ is a function of $(M_{kj})_j$ alone and therefore independent of x_k. The digit-pair laws thus agree for every k, hence so do the joint laws of (x, x') and the variances. The last statement follows because (S2) and (S3) are statements about $\mathrm{Var}(\widehat{I}_N(g))$, and interlacing is a deterministic map of the scrambled net, so the pair laws of interlaced points agree as well. □

The lemma concerns the variance only: statements about the full error distribution, such as concentration bounds or the behavior of median estimators, do not transfer, and they do differ between the two scramblings (Goda & Suzuki, 2026). It also concerns the infinite-precision randomization. An implementation randomizes finitely many digits (`scipy.stats.qmc.Sobol` uses $B = 30$ binary digits by default and at most 64) and sets the remaining digits to zero, so each coordinate of its points is uniform on the grid $\{j2^{-B}\}$ rather than on $[0, 1)$. The next lemma couples such a point set with an ideal scrambling and bounds the difference.

Lemma B.5 (finite-precision scrambling). *Let $\widetilde{P}_N^{(B)} = \{\tilde{u}_i^{(B)}\}$ be the affine matrix scrambling in base 2 of a digital net whose points have at most B nonzero digits, with only the first B digits of each coordinate randomized and the remaining digits set to 0.*

(i) On an extension of the probability space there is an infinite-precision affine matrix scrambling $\widetilde{P}_N = \{\tilde{u}_i\}$ of the same net whose first B digits coincide with those of $\widetilde{P}_N^{(B)}$; in particular $0 \le \tilde{u}_{ij} - \tilde{u}_{ij}^{(B)} \le 2^{-B}$ for all i, j.

(ii) If $G\colon [0,1]^s \to E$ satisfies $\|G(u) - G(u')\|_E \le L_u \|u - u'\|_2$, then $\left\|\widehat{I}_N^{(B)}(G) - \widehat{I}_N(G)\right\|_E \le \epsilon_B := L_u\sqrt{s}\,2^{-B}$ for every realization, and

$$\left(\mathbb{E}\left\|\widehat{I}_N^{(B)}(G) - \int G\right\|_E^2\right)^{1/2} \le \left(\mathbb{E}\left\|\widehat{I}_N(G) - \int G\right\|_E^2\right)^{1/2} + \epsilon_B,$$

$$\left\|\mathbb{E}\,\widehat{I}_N^{(B)}(G) - \int G\right\|_E \le \epsilon_B.$$

(iii) Each coordinate of $\tilde{u}_i^{(B)}$ is uniform on $\{j2^{-B} : 0 \le j < 2^B\}$, so $\mathbb{E}\,\widehat{I}_N^{(B)}(g) = 2^{-B}\sum_j g(j2^{-B})$ for $s = 1$; for $g(u) = u$ this is $\frac{1}{2} - 2^{-B-1}$, and the truncated rule is biased.

Proof. (i) Row k of a lower-triangular scrambling matrix acts only on digits $j \le k$, so the first B scrambled digits depend only on the first B rows and shift digits. Extend each coordinate's matrix by independent rows $k > B$ (entries M_{kj}, $j < k$, uniform on $\{0, 1\}$, $M_{kk} = 1$) and independent shift digits C_k, $k > B$. The result is an infinite-precision affine matrix scrambling of the same net, with independent rows as in Lemma B.4, whose first B digits are those of $\widetilde{P}_N^{(B)}$; truncation to B digits changes each coordinate by at most 2^{-B}. (ii) By (i), $\left\|\tilde{u}_i - \tilde{u}_i^{(B)}\right\|_2 \le \sqrt{s}\,2^{-B}$, so each summand of the two rules differs by at most ϵ_B; the first inequality is Minkowski's inequality in $L^2(\mathbb{P}; E)$ and the second follows from $\mathbb{E}\,\widehat{I}_N(G) = \int G$. (iii) $\tilde{u}_{ij}^{(B)} = 2^{-B}\lfloor 2^B\tilde{u}_{ij}\rfloor$ with $\tilde{u}_{ij}$ uniform on $[0, 1)$, and $2^{-B}\sum_{j<2^B} j2^{-B} = \frac{1}{2} - 2^{-B-1}$. □

For a Lipschitz integrand the effect of finite precision is therefore a deterministic perturbation of size at most ϵ_B, added to the root mean squared error of the ideal scrambling; Theorem C.13(b) holds for the implemented rule with $\sigma_N(x^\star) + \epsilon_B$ in place of $\sigma_N(x^\star)$, and Remark C.16 treats fresh minibatches. For the bounded models of Appendix G the paired comparison of Appendix G.9 measures the effect directly: no reported root mean squared error changes by more than a relative $5 \cdot 10^{-5}$. The Gaussian transports are not Lipschitz, and a coordinate equal to 0 would make Φ^{-1} infinite; Appendix G.10 states the endpoint policy and its audit.

QMC on spheres. Let σ be the normalized surface measure on $\mathbb{S}^d \subset \mathbb{R}^{d+1}$ and $\{Y_{\ell k}\}$ an orthonormal basis of spherical harmonics, $\ell \geq 0$ the degree. For $\tau \geq 0$ the Sobolev space $H^\tau(\mathbb{S}^d)$ consists of the $a \in L^2(\mathbb{S}^d)$ with

$$\|a\|^2_{H^\tau(\mathbb{S}^d)} = \sum_{\ell \geq 0} \big(1 + \ell(\ell + d - 1)\big)^\tau \sum_k |\hat{a}_{\ell k}|^2 < \infty;$$

for $\tau > d/2$ it embeds continuously into $C(\mathbb{S}^d)$, so that the *worst-case error* $\mathrm{wce}_{H^\tau}(Q) = \sup\{|Qa - \sigma a| : \|a\|_{H^\tau} \leq 1\}$ of a rule Q is finite. Following Brauchart et al. (2014), a sequence (Y_N) of N-point sets is a *QMC design of order* τ if $\mathrm{wce}_{H^\tau}(Y_N) \leq c_\tau N^{-\tau/d}$ for the equal-weight rules, the optimal order (Brauchart et al., 2014, Def. 2 and Thm. 3). A QMC design of order τ is one of every order $\tau' \in (d/2, \tau]$ (Brauchart et al., 2014, Thm. 4); there are spherical t-designs (Delsarte et al., 1977) with $t \asymp N^{1/d}$ (Bondarenko et al., 2013) that form QMC designs of every order $\tau > d/2$ (Brauchart et al., 2014, Thm. 6); and N-point sets maximizing the sum of distances are QMC designs of order $(d+1)/2$ (Brauchart et al., 2014, Thm. 16).

The L_2 *spherical cap discrepancy* of $X_N = \{x_1, \dots, x_N\} \subset \mathbb{S}^d$ is $D_{L_2}(X_N)^2 = \int_{-1}^1 \int_{\mathbb{S}^d} \big|\frac{1}{N}\#\{i : x_i \in C(z,t)\} - \sigma(C(z,t))\big|^2 \,\mathrm{d}\sigma(z)\,\mathrm{d}t$, where $C(z,t) = \{x : \langle x, z\rangle \geq t\}$. For a finite signed measure ϱ on $\mathbb{S}^d$ put $\mathcal{D}(\varrho)^2 = \int_{-1}^1 \int_{\mathbb{S}^d} \varrho(C(a,t))^2 \,\mathrm{d}\sigma(a)\,\mathrm{d}t$, so that $D_{L_2}(X_N) = \mathcal{D}(Q_N - \sigma)$ for the empirical measure Q_N of X_N.

Lemma B.6 (Stolarsky's kernel). *Let $C_d = d\sqrt{\pi}\,\Gamma(d/2)/\Gamma((d+1)/2)$, so $C_2 = 4$. The kernel*

$$K_d(y,z) = \int_{\mathbb{S}^d} \int_{-1}^1 \mathbf{1}\{\langle a, y\rangle \geq t\}\mathbf{1}\{\langle a, z\rangle \geq t\}\,\mathrm{d}t\,\mathrm{d}\sigma(a) = 1 - \frac{\|y - z\|}{C_d}$$

is continuous and positive semidefinite, and for every finite signed measure ϱ with $\varrho(\mathbb{S}^d) = 0$

$$C_d\,\mathcal{D}(\varrho)^2 = -\iint \|y - z\|\;\mathrm{d}\varrho(y)\,\mathrm{d}\varrho(z) = C_d \sup\Big\{\Big|\int a\,\mathrm{d}\varrho\Big|^2 : \|a\|_{H_{K_d}} \leq 1\Big\}, \tag{13}$$

where H_{K_d} is the reproducing kernel Hilbert space of K_d. In particular $D_{L_2}(X_N)$ is the worst-case error of the equal-weight rule in H_{K_d}, and for $\varrho = Q_N - \sigma$

$$C_d\,D_{L_2}(X_N)^2 \;=\; \int_{\mathbb{S}^d}\int_{\mathbb{S}^d} \|x - y\|\;\mathrm{d}\sigma(x)\,\mathrm{d}\sigma(y) \;-\; \frac{1}{N^2}\sum_{i,j=1}^N \|x_i - x_j\|\,. \tag{14}$$

Proof. For fixed a, $\int_{-1}^1 \mathbf{1}\{\langle a, y\rangle \geq t\}\mathbf{1}\{\langle a, z\rangle \geq t\}\,\mathrm{d}t = 1 + \min(\langle a, y\rangle, \langle a, z\rangle)$, and $\min(\beta, \beta') = \frac{1}{2}(\beta + \beta' - |\beta - \beta'|)$. Since $\int \langle a, w\rangle\;\mathrm{d}\sigma(a) = 0$ and, by rotation invariance, $\int |\langle a, w\rangle|\,\mathrm{d}\sigma(a) = \|w\| \int |a_1|\,\mathrm{d}\sigma(a)$ with $\int |a_1|\,\mathrm{d}\sigma = \Gamma(\frac{d+1}{2})/(\sqrt{\pi}\,\Gamma(\frac{d}{2}+1)) = 2/C_d$, we get $K_d(y,z) = 1 - \|y - z\|/C_d$. The kernel is a Gram kernel, $K_d(y,z) = \big\langle \mathbf{1}_{\{(a,t):\langle a,y\rangle \geq t\}}, \mathbf{1}_{\{(a,t):\langle a,z\rangle \geq t\}}\big\rangle_{L^2(\sigma \otimes \mathrm{d}t)}$, hence positive semidefinite, and it is continuous. By Fubini's theorem $\mathcal{D}(\varrho)^2 = \iint K_d\,\mathrm{d}\varrho\,\mathrm{d}\varrho$, and the constant 1 does not contribute when $\varrho(\mathbb{S}^d) = 0$; this is the first equality in (13). The second is the standard identity $\iint K\,\mathrm{d}\varrho\,\mathrm{d}\varrho = \big\|\int K(\cdot, y)\,\mathrm{d}\varrho(y)\big\|^2_{H_K} = \sup_{\|a\|_{H_K} \leq 1} |\int a\,\mathrm{d}\varrho|^2$ for a bounded continuous positive semidefinite kernel. Formula (14) is the case $\varrho = Q_N - \sigma$, using that $\int \|x - y\|\;\mathrm{d}\sigma(y)$ does not depend on x. □

Equation (14) is Stolarsky's invariance principle (Stolarsky, 1973; Brauchart & Dick, 2013); on $\mathbb{S}^2$ the double integral equals $4/3$, so $4D_{L_2}(X_N)^2 = \frac{4}{3} - \frac{1}{N^2}\sum_{i,j}\|x_i - x_j\|$. Brauchart & Dick (2013) show that $H_{K_d} = H^{(d+1)/2}(\mathbb{S}^d)$ with equivalent norms, so there is c_0 with $\mathrm{wce}_{H^{(d+1)/2}}(Q) \leq c_0\,D_{L_2}$ for every equal-weight rule, and D_{L_2} is the certificate of $H^{(d+1)/2}(\mathbb{S}^d)$. The maximal sum of distances falls short of $N^2 \iint \|x - y\|\;\mathrm{d}\sigma\,\mathrm{d}\sigma$ by exactly the order $N^{1-1/d}$ (Alexander, 1972; Stolarsky, 1973; Beck, 1984), so every N-point set has $D_{L_2}(X_N) \geq c\,N^{-1/2-1/(2d)}$ and sets with $D_{L_2}(X_N) = O(N^{-1/2-1/(2d)})$ exist; i.i.d. uniform points have $\mathbb{E}D^2_{L_2} \asymp N^{-1}$.

C THE FIELD OF A POINT SET: READINGS, BUDGETS AND PROOFS

This section proves Proposition 2.2 and derives from it every optimization statement of the paper: the uniform reading for a reused point set (Appendix C.3), the one-point reading for minimizers (Appendix C.4), the mean-square reading for refreshed scrambled nets (Appendix C.5) and Sturm's contraction (Appendix C.6). The optimization arguments are the standard ones (Boumal et al., 2019; Ghadimi & Lan, 2013; Hosseini & Sra, 2020; Zhang & Sra, 2016); what is specific is which reading each of them needs.

C.1 SETTING, POTENTIAL AND FIELD

Let $F\colon \mathcal{M} \times [0,1]^s \to \mathbb{R}$ and

$$f(x) = \int_{[0,1]^s} F(x,u)\,\mathrm{d}u, \qquad f_N(x) = \frac{1}{N}\sum_{i=1}^{N} F(x,u_i), \qquad P_N = \{u_1,\dots,u_N\} \subset [0,1)^s. \tag{15}$$

The point set P_N is arbitrary in this section; it is a low-discrepancy set in Appendix C.3 and a scrambled net in Appendix C.5. We refer to u as the *integration variable* and to x as the *decision variable*.

Assumption C.1 (regularity)**.** For every $u \in [0,1]^s$ the function $F(\cdot,u)$ is C^1 on $\mathcal{M}$, and the maps $(x,u) \mapsto F(x,u)$ and $(x,u) \mapsto \operatorname{grad}_x F(x,u) \in T_x\mathcal{M}$ are continuous on $\mathcal{M} \times [0,1]^s$.

Under Assumption C.1, f_N is C^1 with $\operatorname{grad} f_N(x) = \frac{1}{N}\sum_i \operatorname{grad}_x F(x,u_i)$, and f is C^1 with

$$\operatorname{grad} f(x) = \int_{[0,1]^s} \operatorname{grad}_x F(x,u)\,\mathrm{d}u. \tag{16}$$

Indeed, let γ be a C^1 curve with $\gamma(0) = x$ and $\dot\gamma(0) = v$. By the fundamental theorem of calculus and Fubini's theorem,

$$f(\gamma(t)) - f(x) \;=\; \int_0^t \phi(\tau)\,\mathrm{d}\tau, \qquad \phi(\tau) := \int_{[0,1]^s} \langle \operatorname{grad}_x F(\gamma(\tau),u), \dot\gamma(\tau)\rangle\;\mathrm{d}u.$$

The integrand defining ϕ is continuous, hence bounded, on the compact set $\gamma([0,t]) \times [0,1]^s$, so ϕ is continuous by dominated convergence. Therefore

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t))\Big|_{t=0} = \phi(0) = \int_{[0,1]^s} \langle \operatorname{grad}_x F(x,u), v\rangle\;\mathrm{d}u,$$

which is (16); continuity of $\operatorname{grad} f$ follows in the same way.

In the language of the main text, the *discrepancy charge* of the rule is $\rho_N = Q_N - P$ with $Q_N = \frac{1}{N}\sum_i \delta_{u_i}$, the *potential* is $\Phi_N = f_N - f = \rho_N F(\cdot,u)$ and the *field* is $e_N = \operatorname{grad}\Phi_N$. By (16) and the linearity of Q_N,

$$e_N(x) = \operatorname{grad} f_N(x) - \operatorname{grad} f(x) = \rho_N\, G_x, \qquad G_x = \operatorname{grad}_x F(x,\cdot)\colon [0,1]^s \to T_x\mathcal{M}. \tag{17}$$

Nothing in this section uses that $\Omega = [0,1]^s$ or that ρ_N comes from N points: all statements hold for a probability space (Ω, P) and any probability measure Q in place of Q_N, with the certificate of the corresponding function class.

The gradient integrand $u \mapsto \operatorname{grad}_x F(x,u)$ takes values in the fixed inner-product space $T_x\mathcal{M}$; for fixed x it is simply a vector-valued function on the cube, and its u-derivatives require no connection. What is needed is a notion of variation for such functions that controls the Riemannian norm of integration errors and does not depend on a choice of basis in $T_x\mathcal{M}$.

Vector-valued integrands.

Definition C.1 (frame-free Hardy–Krause variation)**.** Let E be a finite-dimensional real inner-product space and $G\colon [0,1]^s \to E$. Define

$$V_{\mathrm{HK}}(G) \;:=\; \sup_{v\in E,\ \|v\|=1} V_{\mathrm{HK}}\big(\,\langle G(\cdot), v\rangle\,\big) \;\in [0,\infty].$$

Lemma C.2 (Koksma–Hlawka inequality for vector-valued integrands)**.** *Let $G\colon [0,1]^s \to E$ with $V_{\mathrm{HK}}(G) < \infty$ and let $(e_a)_{a=1}^m$ be an orthonormal basis of E, $G^a = \langle G, e_a\rangle$. Then:*

(i) $V_{\mathrm{HK}}(G)$ is invariant under linear isometries of E; in particular it does not depend on the basis;

(ii) $\max_a V_{\mathrm{HK}}(G^a) \;\le\; V_{\mathrm{HK}}(G) \;\le\; \big(\sum_{a=1}^m V_{\mathrm{HK}}(G^a)^2\big)^{1/2}$*;*

(iii) for every point set P_N, $\Big\| \dfrac{1}{N}\sum_{i=1}^N G(u_i) - \displaystyle\int_{[0,1]^s} G(u)\,\mathrm{d}u \Big\|_E \;\le\; V_{\mathrm{HK}}(G)\, D_N^*(P_N)$*;*

(iv) if all mixed partial derivatives $\partial^\alpha G$, $\alpha \subseteq \{1,\dots,s\}$, exist and are continuous on the closed cube, then $V_{\mathrm{HK}}(G) \le \sum_{\emptyset\neq\alpha} \|\partial^\alpha G\|_\infty$, where $\|\partial^\alpha G\|_\infty = \sup_u \|\partial^\alpha G(u)\|_E$.

Proof. (i) Frame invariance. If A is a linear isometry of E, then $\langle AG, v\rangle = \langle G, A^* v\rangle$, and A^* maps the unit sphere of E onto itself; taking the supremum over v gives $V_{\mathrm{HK}}(AG) = V_{\mathrm{HK}}(G)$.

(ii) The sandwich. The lower bound is the choice $v = e_a$ in the definition. For the upper bound, write $\langle G, v\rangle = \sum_a v_a G^a$ and use that V_{HK} is a seminorm together with the Cauchy–Schwarz inequality:

$$V_{\mathrm{HK}}(\langle G, v\rangle) \;\le\; \sum_a |v_a|\, V_{\mathrm{HK}}(G^a) \;\le\; \|v\| \Big(\sum_a V_{\mathrm{HK}}(G^a)^2\Big)^{1/2}.$$

(iii) Koksma–Hlawka in the norm of E. Let $e := \frac{1}{N}\sum_i G(u_i) - \int_{[0,1]^s} G$. If $e = 0$ there is nothing to prove; otherwise put $v := e/\|e\|$. Then, by the scalar inequality (12) applied to $g = \langle G, v\rangle$,

$$\begin{aligned}\|e\| = \langle e, v\rangle &= \frac{1}{N}\sum_{i=1}^N \langle G(u_i), v\rangle - \int_{[0,1]^s} \langle G(u), v\rangle\ \mathrm{d}u \\ &\le V_{\mathrm{HK}}(\langle G, v\rangle)\, D_N^*(P_N) \le V_{\mathrm{HK}}(G)\, D_N^*(P_N).\end{aligned}$$

(iv) Smooth integrands. For a unit vector v, $\partial^\alpha \langle G, v\rangle = \langle \partial^\alpha G, v\rangle$ is continuous on the closed cube, so (11) gives $V_{\mathrm{HK}}(\langle G, v\rangle) \le \sum_{\alpha\neq\emptyset} \|\langle \partial^\alpha G, v\rangle\|_\infty \le \sum_{\alpha\neq\emptyset} \|\partial^\alpha G\|_\infty$; take the supremum over v. □

The lemma transports the Koksma–Hlawka inequality from $\mathbb{R}$ to every fibre $T_x\mathcal{M}$ at once, with a constant that does not depend on a choice of orthonormal frame. For a general certificate H we put $\|G\|_H = \sup_{\|v\|=1} \|\langle G(\cdot), v\rangle\|_H$ in the same way; parts (i) and (iii) of the lemma hold verbatim, with D_H in place of D_N^*.

The uniform reading for Hardy–Krause variation.

Assumption C.2 (uniformly bounded variation on a set $S \subseteq \mathcal{M}$)**.**

$$V_0(S) := \sup_{x\in S} V_{\mathrm{HK}}\big(F(x,\cdot)\big) < \infty, \qquad V_1(S) := \sup_{x\in S} V_{\mathrm{HK}}\big(\operatorname{grad}_x F(x,\cdot)\big) < \infty,$$

where the second variation is that of Definition C.1 with $E = T_x\mathcal{M}$.

Theorem C.3 (uniform reading for Hardy–Krause variation)**.** *Under Assumptions C.1 and C.2 on S, for every point set $P_N \subset [0,1)^s$,*

$$\begin{aligned}\sup_{x\in S} \big|f_N(x) - f(x)\big| &\le V_0(S)\, D_N^*(P_N), \\ \sup_{x\in S} \|\operatorname{grad} f_N(x) - \operatorname{grad} f(x)\|_x &\le V_1(S)\, D_N^*(P_N).\end{aligned} \tag{18}$$

Proof. Fix $x \in S$. The first bound is (12) applied to $g = F(x, \cdot)$, whose variation is at most $V_0(S)$. For the second, (16) gives

$$\operatorname{grad} f_N(x) - \operatorname{grad} f(x) = \frac{1}{N}\sum_{i=1}^{N} G(u_i) - \int_{[0,1]^s} G(u)\,\mathrm{d}u,$$
$$G := \operatorname{grad}_x F(x, \cdot)\colon [0,1]^s \to T_x\mathcal{M}.$$

and Lemma C.2(iii) bounds the norm of the right side by $V_{\mathrm{HK}}(G)\, D_N^*(P_N) \le V_1(S)\, D_N^*(P_N)$. Both bounds hold for every $x \in S$ with the same right-hand side, hence for the suprema. □

In the degenerate case $\mathcal{M} = \mathbb{R}$, $s = 1$, the theorem is the Koksma–Hlawka inequality itself; the content is that nothing is lost on the way to a general manifold. Three features of the theorem carry the rest of the paper. It is *deterministic*: no probability is involved, so no union bound over x is needed to make it uniform. It is *decoupled*: the manifold enters only through the definition of grad_x and the norm on $T_x\mathcal{M}$, the point set only through D_N^*. And it is *rate-agnostic*: every applicable star-discrepancy bound is inherited by the gradient error. Higher-order rules use their corresponding mixed-smoothness norm and quadrature bound in the same scalarization argument; higher-order accuracy does not follow from star discrepancy alone.

Remark C.4 (sufficient conditions). By Lemma C.2(iv), Assumption C.2 holds on a compact S as soon as the mixed derivatives $\partial_u^\alpha F(x,u)$ and $\partial_u^\alpha \operatorname{grad}_x F(x,u)$ exist for all $\alpha \subseteq \{1,\dots,s\}$ and are continuous on $S \times [0,1]^s$, since continuous functions on a compact set are bounded. In a chart φ around x (or in an embedding $\mathcal{M} \subset \mathbb{R}^m$), $\partial_u^\alpha \operatorname{grad}_x F(x,u)$ is computed by differentiating the coordinate (or ambient) expression of the gradient in u; the metric coefficients depend on x only and are not differentiated. When $\mathcal{M}$ is compact and F and $\operatorname{grad}_x F$ have continuous mixed u-derivatives on $\mathcal{M} \times [0,1]^s$, Assumption C.2 holds on $S = \mathcal{M}$.

Remark C.5 (second-order transfer). If $F(\cdot, u)$ is C^2 with $(x,u) \mapsto \operatorname{Hess}_x F(x,u)$ continuous, the same argument, now with E the space of self-adjoint operators on $T_x\mathcal{M}$ with the Frobenius inner product, gives $\sup_{x\in S} \|\operatorname{Hess} f_N(x) - \operatorname{Hess} f(x)\| \le V_2(S) D_N^*(P_N)$. Combined with the transfer of gradients this transports second-order stationarity guarantees; we do not pursue this here.

Example C.6 (principal subspaces of a random covariance)**.** Let $\mathcal{M} = \mathrm{Gr}(n,p)$, the quotient of the Stiefel manifold $\mathrm{St}(n,p)$ by $O(p)$ with the quotient metric, represented by matrices $X \in \mathbb{R}^{n\times p}$ with orthonormal columns, and let $a\colon [0,1]^s \to \mathbb{R}^n$ be a random vector with $\mathbb{E}[aa^\top] = \Sigma$. Put $F(X,u) = -\frac{1}{2}\left\|a(u)^\top X\right\|_F^2$, so that $f(X) = -\frac{1}{2}\operatorname{tr}(X^\top \Sigma X)$ is minimized by the dominant p-dimensional eigenspace. The Riemannian gradient is $\operatorname{grad}_X F(X,u) = -(I - XX^\top)a(u)a(u)^\top X$, an element of the horizontal space $\{H : X^\top H = 0\}$, and its mixed u-derivatives are those of the matrix $a(u)a(u)^\top$, which do not involve X beyond the fixed projector $(I - XX^\top)$. If a is a bounded transformation of u with continuous mixed derivatives (for instance $a(u) = W(2u - \mathbf{1})$, uniform noise), Assumption C.2 holds on all of $\mathrm{Gr}(n,p)$ with $V_1 \le \sum_\alpha \left\|\partial^\alpha (aa^\top)\right\|_\infty$. For Gaussian noise $a(u) = W\Phi^{-1}(u)$ the integrand is unbounded, Assumption C.2 fails, and one is in the regime of Owen's boundary growth condition, see Appendix B.2; this is the case in the Gaussian Grassmannian experiment of Appendix G, while the bounded transport $a(u) = \sqrt{3}\,W(2u - \mathbf{1})$ of Appendix G.2 keeps the same Σ and satisfies Assumption C.2 on all of $\mathrm{Gr}(n,p)$. Model (B) of Appendix G.3, a bounded smooth transport on the cone of positive definite matrices, satisfies the hypotheses of the local minimizer-transfer results of Appendix E (Theorem E.3): its gradient integrand has finite variation uniformly over the ball containing the data and the minimizers. Bounded data do not by themselves give the global hypothesis $V_0(\mathcal{M}) < \infty$ of Theorem C.9, which concerns every $x \in \mathcal{M}$. Already for $\mathcal{M} = \mathbb{R}$, $T(u) = u$ and $F(x,u) = \frac{1}{2}(x-u)^2$ the transport is bounded and smooth, yet $V_{\mathrm{HK}}(F(x,\cdot)) = \int_0^1 |x - u|\,\mathrm{d}u \to \infty$ as $|x| \to \infty$; on a manifold with unbounded squared distance the same growth occurs. The global stationarity theorem therefore needs a compact decision set, the localized version of Remark C.10, or a separate verification of $V_0(\mathcal{M})$.

C.2 THE THREE READINGS

Proposition C.7 (three readings of the field)**.** *Let (Ω, P) be a probability space, F satisfy Assumption C.1 with Ω in place of $[0,1]^s$, Q a probability measure on Ω, and $e(x) = \operatorname{grad} f_Q(x) - \operatorname{grad} f(x) = (Q-P)G_x$ the field, where $f_Q = QF$.*

(a) (Uniform.) *For every $S \subseteq \mathcal{M}$, $\sup_{x\in S} \|e(x)\| \le V_1(S)\, D_H(Q)$ with $V_1(S) = \sup_{x\in S} \|G_x\|_H$.*

(b) (Mean square.) *If Q is random with $\mathbb{E}\, Qg = Pg$ for all $g \in L^1(P)$, and $\int \|G_x\|^2 \, \mathrm{d}P < \infty$, then $\mathbb{E}\, \|e(x)\|^2 = \sum_a \operatorname{Var}\left(Q\, \langle G_x, e_a\rangle\right)$ for every orthonormal basis (e_a) of $T_x\mathcal{M}$.*

(c) (One point.) *Let $\operatorname{grad} f(x^\star) = 0$, let $C \ni x^\star$ be closed and geodesically convex, let f_Q be μ_0-strongly geodesically convex on C and C^1 on an open set containing C, and let $\hat{x}$ minimize f_Q over C. Then*

$$\mu_0\, d(\hat{x}, x^\star) \le \|e(x^\star)\|\,. \tag{19}$$

If moreover $\operatorname{grad} f_Q(\hat{x}) = 0$, the minimizing geodesic $\gamma(t) = \operatorname{Exp}_{x^\star}(tw)$, $w = \operatorname{Log}_{x^\star} \hat{x}$, $t \in [0,1]$, lies in an open set on which f_Q is C^2, and Π_t denotes parallel transport from $\gamma(t)$ to $x^\star$ along γ, then with $\bar{H} = \int_0^1 \Pi_t \circ \operatorname{Hess} f_Q(\gamma(t)) \circ \Pi_t^{-1}\, \mathrm{d}t$,

$$\bar{H}\, \operatorname{Log}_{x^\star} \hat{x} = -e(x^\star). \tag{20}$$

If $\mu_0 \le \operatorname{Hess} f_Q \le L_0$ along γ, then $\bar{H}$ is self-adjoint with spectrum in $[\mu_0, L_0]$ and $\|e(x^\star)\|/L_0 \le d(\hat{x}, x^\star) \le \|e(x^\star)\|/\mu_0$.

Proof. (a) Fix $x \in S$ with $e(x) \neq 0$ and put $v = e(x)/\|e(x)\|$. By (16) for f and for f_Q, $\langle e(x), v\rangle = (Q-P)\langle G_x, v\rangle$, a scalar integration error, so $\|e(x)\| = \langle e(x), v\rangle \le \|\langle G_x, v\rangle\|_H\, D_H(Q) \le V_1(S)\, D_H(Q)$.

(b) $e(x) = \sum_a \big((Q-P)\langle G_x, e_a\rangle\big) e_a$, so $\|e(x)\|^2 = \sum_a \big(Q\langle G_x, e_a\rangle - P\langle G_x, e_a\rangle\big)^2$; take expectations and use $\mathbb{E}\, Q\langle G_x, e_a\rangle = P\langle G_x, e_a\rangle$.

(c) The inequality is the argument of Theorem C.13(a) below, which uses only strong convexity of f_Q on C and first-order optimality of $\hat{x}$: $\mu_0 d^2 \le -\langle \operatorname{grad} f_Q(x^\star), \operatorname{Log}_{x^\star} \hat{x}\rangle \le \|\operatorname{grad} f_Q(x^\star)\|\, d$ with $d = d(\hat{x}, x^\star)$, and $\operatorname{grad} f_Q(x^\star) = e(x^\star)$ because $\operatorname{grad} f(x^\star) = 0$. For the identity, let $X(t) = \Pi_t \operatorname{grad} f_Q(\gamma(t)) \in T_{x^\star}\mathcal{M}$. Parallel transport commutes with covariant differentiation along γ, and $\dot{\gamma}(t) = \Pi_t^{-1} w$ because the velocity of a geodesic is parallel, so

$$\dot{X}(t) = \Pi_t \nabla_{\dot{\gamma}(t)} \operatorname{grad} f_Q = \Pi_t\, \operatorname{Hess} f_Q(\gamma(t))\big[\Pi_t^{-1} w\big].$$

Integrating from 0 to 1 gives $X(1) - X(0) = \bar{H}w$, with $X(1) = \Pi_1 \operatorname{grad} f_Q(\hat{x}) = 0$ and $X(0) = \operatorname{grad} f_Q(x^\star) = e(x^\star)$. Each $\Pi_t \operatorname{Hess} f_Q(\gamma(t))\Pi_t^{-1}$ is self-adjoint with the spectrum of $\operatorname{Hess} f_Q(\gamma(t))$, because Π_t is a linear isometry, so the average $\bar{H}$ is self-adjoint with spectrum in $[\mu_0, L_0]$, invertible, and $\|w\| = \big\|\bar{H}^{-1} e(x^\star)\big\|$ lies between $\|e(x^\star)\|/L_0$ and $\|e(x^\star)\|/\mu_0$. □

Remark C.8 (first order). If along a sequence of rules $d(\hat{x}, x^\star) \to 0$, $\sup_{B(x^\star, r)} \|\operatorname{Hess} f_Q - \operatorname{Hess} f\| \to 0$ for some $r > 0$ (for instance by the second-order reading of Remark C.5), and $\operatorname{Hess} f$ is continuous and invertible at $x^\star$, then $\bar{H} \to \operatorname{Hess} f(x^\star)$ and (20) gives $d(\hat{x}, x^\star) = \big\|\operatorname{Hess} f(x^\star)^{-1} e(x^\star)\big\|\,(1 + o(1))$. The minimizer error is then the image of one vector, the field at the minimizer, under a fixed linear map; Table 10 checks this replicate by replicate.

C.3 REUSE: ONE POINT SET ALONG THE OPTIMIZATION PATH

Fix a low-discrepancy point set P_N and minimize f_N by Riemannian gradient descent (Algorithm 1). The point set is chosen once and the algorithm is deterministic; the iterates depend on the points, so only the uniform reading Proposition C.7(a) applies, and it turns every statement about f_N into a statement about f.

Assumption C.3 (uniform smoothness)**.** For every $u \in [0,1]^s$, $F(\cdot, u)$ is L-smooth on $\mathcal{M}$ in the sense of (8).

Algorithm 1 QMC sample average approximation with Riemannian gradient descent (QMC-SAA-RGD)

Require: point set $P_N = \{u_1, \dots, u_N\}$, initial point $x_0 \in \mathcal{M}$, smoothness constant L, iteration budget K

1: **for** $k = 0, 1, \dots, K-1$ **do**
2: $\quad g_k \leftarrow \frac{1}{N} \sum_{i=1}^{N} \operatorname{grad}_x F(x_k, u_i)$ $\quad \triangleright = \operatorname{grad} f_N(x_k)$
3: $\quad x_{k+1} \leftarrow \operatorname{Exp}_{x_k}\big(-\frac{1}{L} g_k\big)$
4: **end for**
5: **return** $x_{k^\star}$ with $k^\star \in \arg\min_{k<K} \|g_k\|$

Under Assumptions C.1 and C.3, f and every f_N are L-smooth on $\mathcal{M}$, since (8) is preserved under averages (for f, average (8) over u and use (16)).

Stationarity.

Theorem C.9 (stationarity transfer)**.** *Let Assumptions C.1 and C.3 hold, let f be bounded below, $\Delta_0 := f(x_0) - \inf_{\mathcal{M}} f$, and suppose $V_0 := V_0(\mathcal{M}) < \infty$. Let $S_0 := \{x \in \mathcal{M} : f(x) \le f(x_0) + 2V_0\}$ and $V_1 := V_1(S_0) < \infty$. Then all iterates of Algorithm 1 lie in S_0, and for every $K \ge 1$*

$$\min_{0 \le k < K} \|\operatorname{grad} f(x_k)\| \;\le\; \sqrt{\frac{2L\,(\Delta_0 + 2V_0 D_N^*(P_N))}{K}} \;+\; V_1\, D_N^*(P_N). \tag{21}$$

Proof. Write $D := D_N^*(P_N) \le 1$.

Step 1 (descent for f_N). Since f_N is L-smooth, the descent lemma (9) with $\eta = 1/L$ gives, for every k,

$$f_N(x_{k+1}) \;\le\; f_N(x_k) - \frac{1}{2L} \|\operatorname{grad} f_N(x_k)\|^2, \tag{22}$$

so in particular $f_N(x_k) \le f_N(x_0)$ for all k.

Step 2 (the iterates stay in S_0). By Theorem C.3 with $S = \mathcal{M}$ and by Step 1,

$$f(x_k) \;\le\; f_N(x_k) + V_0 D \;\le\; f_N(x_0) + V_0 D \;\le\; f(x_0) + 2V_0 D \;\le\; f(x_0) + 2V_0,$$

that is, $x_k \in S_0$ for every k.

Step 3 (summing the descent inequality). Summing (22) over $k < K$ and using $\inf f_N \ge \inf f - V_0 D$ (Theorem C.3 again),

$$\frac{1}{2L} \sum_{k<K} \|\operatorname{grad} f_N(x_k)\|^2 \;\le\; f_N(x_0) - f_N(x_K) \;\le\; \big(f(x_0) + V_0 D\big) - \big(\inf_{\mathcal{M}} f - V_0 D\big) \;=\; \Delta_0 + 2V_0 D,$$

hence $\min_{k<K} \|\operatorname{grad} f_N(x_k)\|^2 \le 2L(\Delta_0 + 2V_0 D)/K$.

Step 4 (transfer to f). Since $x_k \in S_0$, the gradient bound of (18) with $S = S_0$ gives $\|\operatorname{grad} f(x_k)\| \le \|\operatorname{grad} f_N(x_k)\| + V_1 D$ for every k. Combining this with Step 3 yields (21). □

The same bound holds for the returned iterate $x_{k^\star}$, since the proof bounds $\min_k \|\operatorname{grad} f_N(x_k)\|$ and $\|\operatorname{grad} f(x_{k^\star})\| \le \|\operatorname{grad} f_N(x_{k^\star})\| + V_1 D_N^*(P_N)$; this matters because $\min_k \|\operatorname{grad} f(x_k)\|$ itself is not computable. The proof has the structure of every optimization result in this paper: the Riemannian argument (here the descent lemma (Boumal et al., 2019; Bento et al., 2017)) is applied verbatim to f_N, and a reading of the field converts the conclusion into one about f. The only geometric price of uniformity is the enlargement of the sublevel set by $2V_0 D_N^*$.

If $\Delta_0 = 0$, the initial point is already optimal and can be returned. If $V_1 = 0$, every sample gradient is exact on S_0, and one fixed sample suffices. The asymptotic budget below concerns fixed positive Δ_0 and V_1 as $\varepsilon \downarrow 0$.

Remark C.10 (localizing the value transfer). The proof uses $V_0(\mathcal{M})$ only to compare f_N with f at the iterates, which lie in the sublevel set $M_0 := \{x : f_N(x) \le f_N(x_0)\}$. If a set $S \supseteq M_0$ has $V_0(S) < \infty$ (for instance $S = \mathcal{M}$ when $\mathcal{M}$ is compact, or a ball containing M_0 when the sublevel

sets of f_N are bounded), then Steps 2 and 3 hold with $V_0(S)$ in place of V_0, since they evaluate $f_N - f$ only at x_0, at the iterates, and at x_K; the iterates then lie in $S \cap \{f \le f(x_0) + 2V_0(S)\}$, and (21) holds with $V_0(S)$ and $V_1(S \cap S_0)$. Bounded data alone do not give $V_0(\mathcal{M}) < \infty$ on a noncompact manifold (Example C.6), so this localization is what makes the theorem applicable to, for instance, Fréchet means of bounded data on the positive definite cone, whose sample objectives are strongly convex and have bounded sublevel sets. For a budget in the sense of Corollary C.11 the set S must not depend on N, or its constants must be bounded uniformly in N. For Fréchet functionals with data in $\bar{B}(p,\rho)$ on a Hadamard manifold one set serves every N: $f_N(x) \ge \frac{1}{2}(d(x,p) - \rho)^2$ when $d(x,p) \ge \rho$ and $f_N(x_0) \le \frac{1}{2}(d(x_0,p) + \rho)^2$, so every sublevel set $\{f_N \le f_N(x_0)\}$ lies in $S = \bar{B}(p, d(x_0,p) + 2\rho)$.

Corollary C.11 (gradient-evaluation budget)**.** *In the setting of Theorem C.9, let P_N be the first N points of a Halton or Sobol' sequence, so that $D_N^*(P_N) \le C_s N^{-1}(\log N)^s$. Given $\varepsilon > 0$, choose N with $V_1 C_s N^{-1}(\log N)^s \le \varepsilon/2$, that is $N = \widetilde{O}(V_1/\varepsilon)$, and $K = \lceil 8L(\Delta_0 + 2V_0)/\varepsilon^2 \rceil$. Then $\min_{k<K} \|\mathrm{grad}\, f(x_k)\| \le \varepsilon$, and the total number of evaluations of $\mathrm{grad}_x F$ is*

$$KN \;=\; \widetilde{O}\Big(\frac{L(\Delta_0 + V_0)\, V_1}{\varepsilon^3}\Big).$$

If instead P_N is a higher-order digital net of order α (Dick, 2008) and the family $\{\langle \mathrm{grad}_x F(x,\cdot), v\rangle\}$ is bounded in the corresponding norm of mixed smoothness α uniformly in $x \in S_0$ and unit v, then $N = \widetilde{O}(\varepsilon^{-1/\alpha})$ and $KN = \widetilde{O}(\varepsilon^{-2-1/\alpha})$.

Proof. With these choices the first term in (21) is at most $\sqrt{2L(\Delta_0 + 2V_0)/K} \le \varepsilon/2$ and the second is at most $\varepsilon/2$. The higher-order case replaces (12) by the higher-order error bound in the proof of Theorem C.3, which is unchanged otherwise. □

For comparison, Riemannian SGD with i.i.d. samples needs $O(\varepsilon^{-4})$ stochastic gradients (Ghadimi & Lan, 2013; Hosseini & Sra, 2020), and no method using only unbiased stochastic gradients with bounded variance can do better in general (Arjevani et al., 2023, Thm. 3). Under the additional assumption that the stochastic gradients are mean-squared smooth, variance-reduced methods of SPIDER type reach $O(\varepsilon^{-3})$ with i.i.d. samples, in the Euclidean case (Arjevani et al., 2023) and on manifolds (Zhou et al., 2019); the QMC route reaches the same ε^{-3} without variance reduction, and with scrambled nets it goes below it (Appendix C.5). The lower bound does not cover the QMC method because its oracle is different: it evaluates $\mathrm{grad}_x F(x,u)$ at chosen points u and uses regularity of the integrand in u, which an unbiased bounded-variance oracle does not offer. The improvement does not come from giving up unbiasedness: the scrambled-net estimators of Appendix C.5 are unbiased in their ideal form and improve the exponent further. What buys the improvement is that the integration error of N chosen points decays like $\widetilde{O}(N^{-1})$, or faster for randomized nets and smooth integrands, instead of the $N^{-1/2}$ standard deviation of an average of N independent samples. The constants $C_s(\log N)^s$ are those of the Koksma–Hlawka bound; they are pessimistic at moderate N once s exceeds a handful, a known feature of QMC error bounds that the experiments of Appendix G illustrate and that weighted-space theory addresses (Appendix H.2). Smoothness of F in u is what buys the improvement, and higher smoothness buys more: as $\alpha \to \infty$ the budget $\varepsilon^{-2-1/\alpha}$ approaches the ε^{-2} iteration count of exact gradient descent.

C.4 Solve: minimizers of the sampled objective

Under strong geodesic convexity of the population objective, the uniform reading yields a bound on the distance between the sample minimizer and the population minimizer that is linear in the discrepancy. In parts (b) and (c) the sample objective f_N need not be convex: only f is assumed convex. Part (b) uses only stationarity of f_N; part (c) uses minimality of $\hat{x}_N$ over C and needs neither interiority nor V_0, which makes it the natural form for constrained sample average approximation.

Theorem C.12 (minimizer transfer under strong geodesic convexity)**.** *Let $C \subseteq \mathcal{M}$ be closed and geodesically convex, let Assumption C.1 hold on an open set $U \supseteq C$ (that is, with $\mathcal{M}$ replaced by U), let f be μ-strongly geodesically convex on C with minimizer $x^\star$ over C, and write $D = D_N^*(P_N)$, $V_0 = V_0(C)$, $V_1 = V_1(C)$ (Assumption C.2); V_0 may be infinite in parts (b) and (c).*

(a) If $V_0 < \infty$, every minimizer $\hat{x}_N$ of f_N over C satisfies $f(\hat{x}_N) - f(x^\star) \le 2V_0 D$ and $d(\hat{x}_N, x^\star) \le 2\sqrt{V_0 D/\mu}$.

(b) If $x^\star$ lies in the interior of C, so that $\operatorname{grad} f(x^\star) = 0$, every stationary point $\bar{x}$ of f_N in the interior of C satisfies

$$d(\bar{x}, x^\star) \;\le\; \frac{V_{\mathrm{HK}}\big(\operatorname{grad}_x F(\bar{x}, \cdot)\big)}{\mu}\, D \;\le\; \frac{V_1}{\mu}\, D. \tag{23}$$

In particular, if $V_0 < \infty$, (23) holds for $\hat{x}_N$ as soon as $2\sqrt{V_0 D/\mu} < d(x^\star, \partial C)$.

(c) Whether or not $x^\star$ is interior, every minimizer $\hat{x}_N$ of f_N over C satisfies $d(\hat{x}_N, x^\star) \le V_1 D/\mu$.

Proof. *(a)* Since $\hat{x}_N$ minimizes f_N over C and $x^\star \in C$, two applications of Theorem C.3 give

$$f(\hat{x}_N) \;\le\; f_N(\hat{x}_N) + V_0 D \;\le\; f_N(x^\star) + V_0 D \;\le\; f(x^\star) + 2V_0 D.$$

The distance bound follows from quadratic growth: (6) at $x^\star$ towards $\hat{x}_N$ and the first-order optimality condition $\langle \operatorname{grad} f(x^\star), \operatorname{Log}_{x^\star}(\hat{x}_N)\rangle \ge 0$ of $x^\star$ over C give $\frac{\mu}{2}\, d(\hat{x}_N, x^\star)^2 \le f(\hat{x}_N) - f(x^\star) \le 2V_0 D$.

(b) Let $\bar{x}$ be interior to C with $\operatorname{grad} f_N(\bar{x}) = 0$. The second inequality in (7), formula (16) at $\bar{x} \in U$, and Lemma C.2(iii) give

$$\mu\, d(\bar{x}, x^\star) \;\le\; \|\operatorname{grad} f(\bar{x})\| \;=\; \|\operatorname{grad} f(\bar{x}) - \operatorname{grad} f_N(\bar{x})\| \;\le\; V_{\mathrm{HK}}\big(\operatorname{grad}_x F(\bar{x}, \cdot)\big)\, D.$$

For the last claim: if $2\sqrt{V_0 D/\mu} < d(x^\star, \partial C)$, then by (a) the point $\hat{x}_N$ is interior to C, hence stationary for f_N, and (b) applies to it.

(c) Write $\hat{x} = \hat{x}_N$, $d = d(\hat{x}, x^\star) > 0$ and $\ell = \operatorname{Log}_{\hat{x}}(x^\star)$. First-order optimality over the geodesically convex set C, as in the proof of Theorem C.13, gives $\langle \operatorname{grad} f(x^\star), \operatorname{Log}_{x^\star}(\hat{x})\rangle \ge 0$ and $\langle \operatorname{grad} f_N(\hat{x}), \ell\rangle \ge 0$. Applying (6) for f at $\hat{x}$ towards $x^\star$ and at $x^\star$ towards $\hat{x}$ and adding, $\mu d^2 \le -\langle \operatorname{grad} f(\hat{x}), \ell\rangle$. Adding the nonnegative term $\langle \operatorname{grad} f_N(\hat{x}), \ell\rangle$ and using the Cauchy–Schwarz inequality, formula (16) at $\hat{x} \in U$ and Lemma C.2(iii) as in part (b),

$$\mu d^2 \;\le\; \langle \operatorname{grad} f_N(\hat{x}) - \operatorname{grad} f(\hat{x}), \ell\rangle \;\le\; V_{\mathrm{HK}}\big(\operatorname{grad}_x F(\hat{x}, \cdot)\big) D\, d \;\le\; V_1 D\, d.$$

Neither V_0 nor strong convexity of f_N is used. □

With a Halton or Sobol' point set (23) reads $d(\hat{x}_N, x^\star) = \widetilde{O}(N^{-1})$, deterministically. The i.i.d. sample average approximation gives $d(\hat{x}_N, x^\star) = O_{\mathbb{P}}(N^{-1/2})$, the rate of the central limit theorem for M-estimators on manifolds (Bhattacharya & Patrangenaru, 2003; 2005), and in the Euclidean setting of Shapiro et al. (2009) the constant V_1/μ plays the role of the ratio of the Lipschitz constant of the sample gradient to the strong convexity modulus. When f_N is itself strongly geodesically convex and smooth on C, its minimizer is computed by Riemannian gradient descent at a linear rate (Zhang & Sra, 2016; Boumal, 2023), so that the overall cost of ε-accuracy in distance is $\widetilde{O}(N(\varepsilon)\log(1/\varepsilon))$ gradient evaluations with $N(\varepsilon) = \widetilde{O}(V_1/(\mu\varepsilon))$.

The roles of the two objectives can be exchanged: if the *sample* objective is strongly convex, only the integrand at the single point $x^\star$ needs a quadrature bound. This is the situation of Fréchet means on Hadamard manifolds (Appendix E), where every sample objective is 1-strongly convex, and it is the regime of a sample average approximation solved once for one fixed, possibly randomized, point set.

Theorem C.13 (minimizer transfer under sample strong convexity)**.** *Let $C \subseteq \mathcal{M}$ be closed and geodesically convex, let Assumption C.1 hold on an open set $U \supseteq C$, and let $x^\star \in C$ satisfy $\operatorname{grad} f(x^\star) = 0$.*

(a) If f_N is μ_0-strongly geodesically convex on C and $\hat{x}_N$ minimizes f_N over C, then

$$\begin{aligned} \mu_0\, d(\hat{x}_N, x^\star) \;&\le\; \|\operatorname{grad} f_N(x^\star)\| \;=\; \|\operatorname{grad} f_N(x^\star) - \operatorname{grad} f(x^\star)\| \\ &\le\; V_{\mathrm{HK}}\big(\operatorname{grad}_x F(x^\star, \cdot)\big)\, D_N^*(P_N). \end{aligned} \tag{24}$$

(b) Let the point set be one scrambled net $\widetilde{P}_N = \{\tilde{u}_i\}$, drawn once and kept fixed, and suppose that almost surely f_N is μ_0-strongly geodesically convex on C, with μ_0 not depending on the draw,

and $\hat{x}_N$ minimizes f_N over C. Then

$$\mathbb{E}\, d(\hat{x}_N, x^\star)^2 \;\le\; \frac{\sigma_N^2(x^\star)}{\mu_0^2}, \qquad \sigma_N^2(x^\star) := \mathbb{E}\Big\| \frac{1}{N}\sum_{i=1}^{N} \operatorname{grad}_x F(x^\star, \tilde{u}_i) - \operatorname{grad} f(x^\star)\Big\|^2. \tag{25}$$

Under (S2), $\sigma_N^2(x^\star) \le \Gamma_{b,t,s}\, \sigma_F^2(x^\star)/N$, where

$$\sigma_F^2(x^\star) = \int \|\operatorname{grad}_x F(x^\star, u) - \operatorname{grad} f(x^\star)\|^2 \, \mathrm{d}u,$$

and if $u \mapsto \langle \operatorname{grad}_x F(x^\star, u), v\rangle$ satisfies the smoothness of (S3) for every unit $v \in T_{x^\star}\mathcal{M}$ with a common constant, then $\sigma_N^2(x^\star) = O(N^{-3}(\log N)^{s-1})$ (Lemma C.14 below, at the single point $x^\star$).

Proof. Write $\hat{x} = \hat{x}_N$ and $d = d(\hat{x}, x^\star)$, and assume $d > 0$. Since $\hat{x}$ minimizes f_N over the geodesically convex set C and f_N is C^1 on U, first-order optimality gives $\langle \operatorname{grad} f_N(\hat{x}), \operatorname{Log}_{\hat{x}}(y)\rangle \ge 0$ for all $y \in C$: the geodesic $t \mapsto \operatorname{Exp}_{\hat{x}}(t \operatorname{Log}_{\hat{x}}(y))$ stays in C for $t \in [0,1]$, so $t \mapsto f_N(\operatorname{Exp}_{\hat{x}}(t \operatorname{Log}_{\hat{x}}(y)))$ is minimal at $t = 0$, and the left-hand side is its derivative there. Apply (6) for f_N twice, at $\hat{x}$ towards $x^\star$ and at $x^\star$ towards $\hat{x}$:

$$\begin{aligned} f_N(x^\star) \;&\ge\; f_N(\hat{x}) + \langle \operatorname{grad} f_N(\hat{x}), \operatorname{Log}_{\hat{x}}(x^\star)\rangle + \tfrac{\mu_0}{2} d^2 \;\ge\; f_N(\hat{x}) + \tfrac{\mu_0}{2} d^2, \\ f_N(\hat{x}) \;&\ge\; f_N(x^\star) + \langle \operatorname{grad} f_N(x^\star), \operatorname{Log}_{x^\star}(\hat{x})\rangle + \tfrac{\mu_0}{2} d^2. \end{aligned}$$

Adding the two inequalities gives $\mu_0 d^2 \le -\langle \operatorname{grad} f_N(x^\star), \operatorname{Log}_{x^\star}(\hat{x})\rangle \le \|\operatorname{grad} f_N(x^\star)\|\, d$ by the Cauchy–Schwarz inequality, which is the first inequality in (24). The middle equality holds because $\operatorname{grad} f(x^\star) = 0$, and the last inequality is Lemma C.2(iii) applied to $G = \operatorname{grad}_x F(x^\star, \cdot)$ through formula (16) at $x^\star \in U$. For (b), the first inequality in (24) holds for every draw; square it and take expectations. The two bounds on $\sigma_N^2(x^\star)$ are Lemma C.14 at the single point $x = x^\star$. □

Theorem C.12(b) uses strong convexity of f and stationarity of f_N at $\hat{x}_N$; Theorem C.13 uses strong convexity of f_N and stationarity of f at $x^\star$. The second form asks more of the sample objective and less of the quadrature: no uniformity in x, no expected supremum over a family of integrands, and no fresh randomization at every iteration. Part (b) is the guarantee for a sample average approximation solved for one scrambled point set, which is how the experiments of Appendix G.3 are run: there every sample objective is 1-strongly convex, and the observed ratio $d(\hat{x}_N, x^\star)/\|\operatorname{grad} f_N(x^\star)\|$ lies between 0.977 and 0.984, consistent with (24) at $\mu_0 = 1$.

C.5 Refresh: Riemannian SGD with scrambled minibatches

A deterministic point set produces an inexact gradient, whose error is a fixed function of the point set. Owen's scrambling, or the affine matrix scrambling of Lemma B.4 in its infinite-precision form, makes the estimator unbiased while keeping the low discrepancy, and it does so *independently at every iteration*, which is what a stochastic-approximation analysis needs; the finite-precision implementations are treated in Remark C.16. Algorithm 2 is Riemannian SGD in which the minibatch at iteration k is a freshly scrambled digital net.

Algorithm 2 Riemannian SGD with scrambled-net minibatches (RQMC-RSGD)

Require: a (t, m, s)-net P_N in base b, $N = b^m$; initial point x_0; step size η; iteration budget K; optionally a closed geodesically convex set X for projection

1: **for** $k = 0, 1, \dots, K-1$ **do**
2: draw an Owen scrambling $\widetilde{P}_N^{(k)} = \{u_1^{(k)}, \dots, u_N^{(k)}\}$ of P_N, independent of the past
3: $g_k \leftarrow \frac{1}{N}\sum_{i=1}^{N} \operatorname{grad}_x F(x_k, u_i^{(k)})$
4: $x_{k+1} \leftarrow \operatorname{Exp}_{x_k}(-\eta\, g_k)$ (or $x_{k+1} \leftarrow \Pi_X\big(\operatorname{Exp}_{x_k}(-\eta\, g_k)\big)$ in the projected variant)
5: **end for**
6: **return** x_ι with ι uniform on $\{0, \dots, K-1\}$ (nonconvex case) or x_K (strongly convex case)

Let $\widehat{g}_N(x) = \frac{1}{N}\sum_i \operatorname{grad}_x F(x, \tilde{u}_i)$ denote the estimator built from one scrambled net $\{\tilde{u}_i\}$, scrambled by nested uniform scrambling or by the affine matrix scrambling of Lemma B.4, which

has the same pair law and therefore the same first two moments. Property (S1) and (16) give $\mathbb{E}[\widehat{g}_N(x)] = \operatorname{grad} f(x)$ whenever $\int \|\operatorname{grad}_x F(x,u)\|\, \mathrm{d}u < \infty$. Its mean squared error is controlled by scalar scrambled-net theory through the following lemma, which is the randomized counterpart of Lemma C.2.

Lemma C.14 (from scalar variance bounds to tangent vectors)**.** *Let $x \in \mathcal{M}$ and assume $\sigma_F^2(x) := \int_{[0,1]^s} \|\operatorname{grad}_x F(x,u) - \operatorname{grad} f(x)\|_x^2 \, \mathrm{d}u < \infty$.*

(i) For every scrambled (t,m,s)-net in base b, $\mathbb{E}\,\|\widehat{g}_N(x) - \operatorname{grad} f(x)\|^2 \le \Gamma_{b,t,s}\, \sigma_F^2(x)/N$.

(ii) If for every unit $v \in T_x\mathcal{M}$ the scalar function $u \mapsto \langle \operatorname{grad}_x F(x,u), v\rangle$ satisfies the hypothesis of (S3) (respectively (S4)) with a constant c_x independent of v, then $\mathbb{E}\,\|\widehat{g}_N(x) - \operatorname{grad} f(x)\|^2 \le n\, c_x\, N^{-3}(\log N)^{s-1}$ (respectively $n\, c_x\, N^{-2\alpha-1+\delta}$), where $n = \dim \mathcal{M}$.

Proof. Fix an orthonormal basis $(e_a)_{a=1}^n$ of $T_x\mathcal{M}$ and put $G^a(u) := \langle \operatorname{grad}_x F(x,u), e_a\rangle$, so that $\widehat{g}_N(x) = \sum_a \widehat{I}_N(G^a)\, e_a$.

Step 1 (decomposition of the mean squared error). By (S1), $\mathbb{E}\,\widehat{I}_N(G^a) = \int_{[0,1]^s} G^a = \langle \operatorname{grad} f(x), e_a\rangle$, hence

$$\mathbb{E}\,\|\widehat{g}_N(x) - \operatorname{grad} f(x)\|^2 \;=\; \sum_{a=1}^{n} \operatorname{Var}\big(\widehat{I}_N(G^a)\big).$$

Step 2 (part (i)). By (S2) applied to each G^a,

$$\begin{aligned}\sum_a \operatorname{Var}(\widehat{I}_N(G^a)) &\le \frac{\Gamma}{N}\sum_a \sigma^2(G^a)\\ &= \frac{\Gamma}{N}\int_{[0,1]^s}\sum_a \langle \operatorname{grad}_x F(x,u) - \operatorname{grad} f(x), e_a\rangle^2 \,\mathrm{d}u\\ &= \frac{\Gamma}{N}\,\sigma_F^2(x).\end{aligned}$$

Step 3 (part (ii)). Each of the n terms in Step 1 is bounded by (S3) (respectively (S4)) applied to G^a, with the constant c_x. □

Part (i) is dimension-free: the sum of componentwise variances is the total variance in the Riemannian norm. Note also that the bounds are again invariant under the choice of frame, since the left-hand sides are.

Assumption C.4 (estimator variance)**.** There is a set $S \subseteq \mathcal{M}$ and a number σ_N^2 such that $\mathbb{E}\,\|\widehat{g}_N(x) - \operatorname{grad} f(x)\|^2 \le \sigma_N^2$ for all $x \in S$.

By Lemma C.14, Assumption C.4 holds on S with $\sigma_N^2 = \Gamma_{b,t,s}\,\sigma_F^2(S)/N$, $\sigma_F^2(S) = \sup_{x\in S}\sigma_F^2(x)$, for any square-integrable gradient integrand, and with $\sigma_N^2 = n\,c\,N^{-3}(\log N)^{s-1}$ when the smoothness of (S3) holds uniformly on S. We keep σ_N^2 abstract so that any present or future scrambled-net variance bound can be inserted.

Nonconvex objectives. The descent argument uses only the population objective and the first two conditional moments of the search direction. We therefore state it in that generality and then verify the hypotheses for Algorithm 2; in particular, the components $F(\cdot,u)$ need not share a smoothness constant.

Assumption C.5 (population smoothness and conditional moments)**.** f is L-smooth on $\mathcal{M}$ in the sense of (8) and bounded below, and the search directions g_k of the iteration $x_{k+1} = \operatorname{Exp}_{x_k}(-\eta g_k)$ satisfy, for the sigma-field $\mathcal{F}_k$ generated by the randomness of iterations $0,\dots,k-1$,

$$\mathbb{E}[g_k \mid \mathcal{F}_k] = \operatorname{grad} f(x_k), \qquad \mathbb{E}\big[\,\|g_k - \operatorname{grad} f(x_k)\|^2 \,\big|\, \mathcal{F}_k\big] \le \sigma_N^2 \qquad \text{for every } k.$$

Theorem C.15 (nonconvex case). *Under Assumption C.5, the iteration $x_{k+1} = \mathrm{Exp}_{x_k}(-\eta g_k)$ with $0 < \eta \le 1/L$ satisfies, with $\Delta_0 = f(x_0) - \inf_{\mathcal{M}} f$ and ι uniform on $\{0, \dots, K-1\}$ and independent of the run,*

$$\mathbb{E}\,\|\mathrm{grad}\, f(x_\iota)\|^2 \;=\; \frac{1}{K}\sum_{k=0}^{K-1} \mathbb{E}\,\|\mathrm{grad}\, f(x_k)\|^2 \;\le\; \frac{2\Delta_0}{\eta K} + L\eta\,\sigma_N^2. \tag{26}$$

Algorithm 2 (unprojected) satisfies Assumption C.5 whenever the gradient formula (16) holds, f is L-smooth on $\mathcal{M}$ and bounded below, $\int \|\mathrm{grad}_x F(x,u)\|^2\, \mathrm{d}u < \infty$ for all x, and Assumption C.4 holds on $S = \mathcal{M}$. Formula (16) holds under Assumption C.1, and also, by dominated differentiation, when $F(\cdot, u)$ is C^1 for almost every u and $\sup_x \|\mathrm{grad}_x F(x,u)\| \le \Lambda(u)$ with Λ integrable, which covers transports that are unbounded near the boundary of the cube. Assumption C.3 implies the L-smoothness of f but is not otherwise needed.

Proof. Let $\mathcal{F}_k$ be as in Assumption C.5. For Algorithm 2, $\mathcal{F}_k$ is generated by the scramblings of iterations $0, \dots, k-1$, x_k is $\mathcal{F}_k$-measurable, and the scrambling at iteration k is independent of $\mathcal{F}_k$.

Step 1 (conditional moments of the search direction). Under Assumption C.5,

$$\begin{aligned} \mathbb{E}[g_k \mid \mathcal{F}_k] \;&=\; \mathrm{grad}\, f(x_k), \\ \mathbb{E}\big[\,\|g_k\|^2 \,\big|\, \mathcal{F}_k\big] \;&=\; \|\mathrm{grad}\, f(x_k)\|^2 + \mathbb{E}\big[\,\|g_k - \mathrm{grad}\, f(x_k)\|^2 \,\big|\, \mathcal{F}_k\big] \;\le\; \|\mathrm{grad}\, f(x_k)\|^2 + \sigma_N^2. \end{aligned}$$

For Algorithm 2 these two relations are (S1) (with (16), the square integrability ensuring that the moments exist) and Assumption C.4 at the $\mathcal{F}_k$-measurable point x_k, because the fresh scrambling is independent of $\mathcal{F}_k$; this verifies the last claim of the theorem, the L-smoothness of f under Assumption C.3 having been noted in Appendix C.3.

Step 2 (one step of descent). Since f is L-smooth on $\mathcal{M}$, (8) with $v = -\eta g_k$ gives

$$f(x_{k+1}) \;\le\; f(x_k) - \eta\,\langle \mathrm{grad}\, f(x_k), g_k\rangle + \frac{L\eta^2}{2}\,\|g_k\|^2 .$$

Step 3 (expectations). Taking $\mathbb{E}[\,\cdot \mid \mathcal{F}_k]$ in Step 2, inserting Step 1, and then taking expectations,

$$\begin{aligned} \mathbb{E} f(x_{k+1}) &\le \mathbb{E} f(x_k) - \eta\Big(1 - \frac{L\eta}{2}\Big)\mathbb{E}\,\|\mathrm{grad}\, f(x_k)\|^2 + \frac{L\eta^2}{2}\sigma_N^2 \\ &\le \mathbb{E} f(x_k) - \frac{\eta}{2}\,\mathbb{E}\,\|\mathrm{grad}\, f(x_k)\|^2 + \frac{L\eta^2}{2}\sigma_N^2 \end{aligned},$$

where the second inequality uses $L\eta \le 1$.

Step 4 (telescoping). Summing Step 3 over $k < K$ and using $\mathbb{E} f(x_K) \ge \inf f$,

$$\frac{\eta}{2}\sum_{k<K} \mathbb{E}\,\|\mathrm{grad}\, f(x_k)\|^2 \;\le\; \Delta_0 + K\,\frac{L\eta^2}{2}\,\sigma_N^2,$$

and dividing by $\eta K/2$ gives (26). □

The bound (26) is the standard one for SGD with a constant step, in the Euclidean case (Ghadimi & Lan, 2013) and on manifolds (Hosseini & Sra, 2020, Lemma 7 and Theorem 8); the point is that σ_N^2 is now the mean squared error of a scrambled-net estimator and decays faster than $1/N$. Only the population objective needs to be smooth. In the Gaussian Grassmann model of Appendix G.1, $f(X) = -\frac{1}{2}\,\mathrm{tr}(X^\top \Sigma X)$ is smooth on the compact Grassmannian while the smoothness constants of the components grow with $\|a(u)\|^2$; Assumption C.1 fails on the closed cube, but $\|\mathrm{grad}_X F(X,u)\|_F \le \|a(u)\|^2$ is integrable, so (16) holds by dominated differentiation, and the theorem applies with the variance bound of (S2), $\sigma_N^2 \le \Gamma\sigma_F^2/N$, because the gradient integrand is square integrable, whereas the N^{-3} rate of (S3) needs the mixed smoothness that the Gaussian transport lacks and that the bounded transport of Appendix G.2 has.

Remark C.16 (finite-precision scrambling). Implementations randomize B digits only (Lemma B.5), so their estimator has a deterministic bias of norm at most ϵ_B for integrands that are L_u-Lipschitz in u. If, in place of the first condition of Assumption C.5, $\|\mathbb{E}[g_k \mid \mathcal{F}_k] - \mathrm{grad}\, f(x_k)\| \le \beta$, and σ_N^2

bounds $\mathbb{E}[\|g_k - \mathbb{E}[g_k \mid \mathcal{F}_k]\|^2 \mid \mathcal{F}_k]$, then (26) holds with β^2 added to its right-hand side. Indeed, writing $b_k = \mathbb{E}[g_k \mid \mathcal{F}_k] - \operatorname{grad} f(x_k)$ and using $L\eta \le 1$, the conditional expectation of the right side of Step 2 is at most

$$\begin{aligned} &f(x_k) - \eta \langle \operatorname{grad} f(x_k), \operatorname{grad} f(x_k) + b_k \rangle + \frac{\eta}{2} \|\operatorname{grad} f(x_k) + b_k\|^2 + \frac{L\eta^2}{2}\sigma_N^2 \\ &\quad = f(x_k) - \frac{\eta}{2} \|\operatorname{grad} f(x_k)\|^2 + \frac{\eta}{2} \|b_k\|^2 + \frac{L\eta^2}{2}\sigma_N^2, \end{aligned}$$

and telescoping as in Step 4 gives the claim. For the implemented estimator one may take $\beta = \epsilon_B$ and replace σ_N by $\sigma_N + \epsilon_B$, provided L_u bounds the Lipschitz constant in u uniformly in x (Lemma B.5(ii)). With $B = 30$ both changes are of order $\beta_{30} \le 3.6 \cdot 10^{-7}$ for the bounded Grassmann integrand of Appendix G.2.

For higher-order randomized rates below, $\delta > 0$ denotes an arbitrarily small exponent loss. It may be chosen afresh in each bound; its constant may depend on δ. This loss is inherited from property (S4) and is not a logarithmic factor. Integer iteration counts add at most N component-gradient evaluations to the unrounded budget. If $\Delta_0 = 0$, return the already optimal x_0; the iteration budgets below concern $\Delta_0 > 0$. When $\sigma_N^2 = 0$, take $\eta = 1/L$.

Corollary C.17 (the budget ladder)**.** *In the setting of Theorem C.15, given $\varepsilon > 0$, choose $\eta = \min\{1/L,\ \varepsilon^2/(2L\sigma_N^2)\}$ and $K = \lceil 4\Delta_0/(\eta\varepsilon^2)\rceil$. Then $\mathbb{E}\,\|\operatorname{grad} f(x_\iota)\|^2 \le \varepsilon^2$, and the total number KN of gradient evaluations is as in rows 1, 2, 4 and 6 of Table 4; the deterministic rows 3 and 5 of the table are recorded there for comparison and follow from Theorem C.9 and Corollary C.11 under Assumption C.2 instead. In particular (we write σ_F^2 for $\sigma_F^2(\mathcal{M})$):*

- *(i) (i.i.d. minibatches) $\sigma_N^2 = \sigma_F^2/N$ and $KN \le 8L\Delta_0\sigma_F^2\varepsilon^{-4} + N$ for every $N \le 2\sigma_F^2/\varepsilon^2$ (for larger N the budget grows linearly in N, since $K \ge 4L\Delta_0/\varepsilon^2$ regardless);*
- *(ii) (scrambled nets, square-integrable integrand) $KN \le 8\Gamma_{b,t,s}L\Delta_0\sigma_F^2\varepsilon^{-4} + N$ for every $N \le 2\Gamma_{b,t,s}\sigma_F^2/\varepsilon^2$: never worse than i.i.d. sampling by more than the factor $\Gamma_{b,t,s} = b^t\big(\frac{b+1}{b-1}\big)^s$, which for $b = 2$ is 2^t3^s and for $t = 0$ can be replaced by $(b/(b-1))^s$ ((S2)), that is, 2^s in base 2;*
- *(iii) (scrambled (t, m, s)-nets, smooth integrand) with $N = N(\varepsilon)$ the smallest net size with $\sigma_N^2 \le \varepsilon^2/2$, $N(\varepsilon) = \widetilde{O}(\varepsilon^{-2/3})$, $\eta = 1/L$, $K = \lceil 4L\Delta_0/\varepsilon^2\rceil$ and $KN = \widetilde{O}(L\Delta_0\,\varepsilon^{-8/3})$;*
- *(iv) (higher-order scrambled nets of order α, for every $\delta > 0$) $N(\varepsilon) = O(\varepsilon^{-2/(2\alpha+1)-\delta})$ and $KN = O(L\Delta_0\,\varepsilon^{-2-2/(2\alpha+1)-\delta})$.*

Proof. There are two regimes, according to which term of the minimum defines η.

- If $\sigma_N^2 \ge \varepsilon^2/2$, then $\eta = \varepsilon^2/(2L\sigma_N^2) \le 1/L$ and the noise term of (26) equals $L\eta\sigma_N^2 = \varepsilon^2/2$; the choice $K = \lceil 8L\Delta_0\sigma_N^2/\varepsilon^4\rceil$ makes the first term $2\Delta_0/(\eta K) \le \varepsilon^2/2$ as well, and

$$KN \;\le\; 8L\Delta_0\,(N\sigma_N^2)\,\varepsilon^{-4} + N, \qquad N\sigma_N^2 = \sigma_F^2 \text{ (i.i.d.)}, \quad N\sigma_N^2 \le \Gamma\sigma_F^2 \text{ (scrambled nets)}.$$

- If $\sigma_N^2 \le \varepsilon^2/2$, then $\eta = 1/L$, the noise term is at most $\varepsilon^2/2$, and $K = \lceil 4L\Delta_0/\varepsilon^2\rceil$ makes the first term at most $\varepsilon^2/2$, so $KN \le 4L\Delta_0 N/\varepsilon^2 + N$.

In cases (iii) and (iv) the smallest N with $\sigma_N^2 \le \varepsilon^2/2$ is $\widetilde{O}(\varepsilon^{-2/3})$, respectively $O(\varepsilon^{-2/(2\alpha+1)-\delta})$, by Lemma C.14(ii), and the second regime applies. □

Strong geodesic convexity on Hadamard manifolds: curvature and discrepancy separate. We now switch from gradient norms to distances and from L-smoothness to an L-Lipschitz gradient (Definition B.2). Under strong geodesic convexity the classical Riemannian SGD analysis of Zhang & Sra (2016) bounds distances rather than gradient norms, and curvature enters through the constant ζ of Lemma B.3. The following theorem shows how this constant and the discrepancy interact: ζ multiplies the variance term, and the variance is the scrambled-net variance.

Table 4: Gradient-evaluation budgets for ε-stationarity on an L-smooth objective on a manifold, as functions of the sampler: rows 3 and 5 certify $\min_{k<K} \|\mathrm{grad}\, f(x_k)\| \le \varepsilon$ deterministically (Corollary C.11), the other rows certify $\mathbb{E}\,\|\mathrm{grad}\, f(x_\iota)\|^2 \le \varepsilon^2$ for a random iterate (Corollary C.17). σ_N^2 is the mean squared error of the N-point gradient estimator (bias^2 for deterministic sets). Logarithmic factors are suppressed. In the higher-order randomized row, $\delta > 0$ is an arbitrarily small loss that may differ between entries. The Riemannian constants L, Δ_0 and, in rows 4 and 6, the dimension $\dim \mathcal{M}$ multiply every entry; for integrands of uniformly bounded variation the geometry does not change the exponent (Appendix D describes what happens at the cut locus).

sampler	σ_N^2	minibatch $N(\varepsilon)$	total KN
i.i.d. Monte Carlo	σ_F^2/N	$\le 2\sigma_F^2/\varepsilon^2$	ε^{-4}
scrambled net, L^2 integrand	$\le \Gamma\sigma_F^2/N$	$\le 2\Gamma\sigma_F^2/\varepsilon^2$	$\le \Gamma\varepsilon^{-4}$
deterministic Halton/Sobol' (Theorem C.9)	$V_1^2 N^{-2}(\log N)^{2s}$	ε^{-1}	ε^{-3}
scrambled (t,m,s)-net, smooth integrand	$N^{-3}(\log N)^{s-1}$	$\varepsilon^{-2/3}$	$\varepsilon^{-8/3}$
higher-order net, order α (deterministic)	$N^{-2\alpha}$	$\varepsilon^{-1/\alpha}$	$\varepsilon^{-2-1/\alpha}$
higher-order scrambled net, order α	$N^{-2\alpha-1+\delta}$	$\varepsilon^{-2/(2\alpha+1)-\delta}$	$\varepsilon^{-2-2/(2\alpha+1)-\delta}$
exact gradient	0	–	ε^{-2} iterations

Theorem C.18 (RQMC-RSGD, strongly geodesically convex)**.** *Let $\mathcal{M}$ be a Hadamard manifold with sectional curvatures bounded below by $-\kappa$, $\kappa \ge 0$, and let $X \subset \mathcal{M}$ be closed, geodesically convex, of diameter at most D_X. Suppose $x_0 \in X$, that f has a global minimizer $x^\star \in X$, is μ-strongly geodesically convex on X, has an L-Lipschitz gradient on X, and that Assumptions C.1 and C.4 hold on $S = X$. Let $\zeta = \zeta(\kappa, D_X)$. Then the projected Algorithm 2 with $0 < \eta \le \mu/(\zeta L^2)$ satisfies, for all $k \ge 0$,*

$$\mathbb{E}\, d(x_{k+1}, x^\star)^2 \;\le\; (1-\mu\eta)\, \mathbb{E}\, d(x_k, x^\star)^2 + \zeta\eta^2\sigma_N^2, \qquad \textit{and hence} \tag{27}$$

$$\mathbb{E}\, d(x_K, x^\star)^2 \;\le\; (1-\mu\eta)^K d(x_0, x^\star)^2 + \frac{\zeta\eta\,\sigma_N^2}{\mu}. \tag{28}$$

Proof. Let $y_{k+1} := \mathrm{Exp}_{x_k}(-\eta g_k)$, so that $x_{k+1} = \Pi_X(y_{k+1})$.

Step 1 (projection). Since $x^\star \in X$ and Π_X is nonexpansive with $\Pi_X x^\star = x^\star$, we have $d(x_{k+1}, x^\star) \le d(y_{k+1}, x^\star)$.

Step 2 (trigonometric bound). Apply Lemma B.3 to the geodesic triangle with vertices $x_k, y_{k+1}, x^\star$. The sides adjacent to x_k have lengths $b = \eta\,\|g_k\|$ and $c = d(x_k, x^\star) \le D_X$, the angle A between them satisfies $bc\cos A = \langle -\eta g_k, \mathrm{Log}_{x_k}(x^\star)\rangle$, and the third side is $a = d(y_{k+1}, x^\star)$. Hence, with $\zeta = \zeta(\kappa, D_X) \ge \zeta(\kappa, c)$,

$$d(y_{k+1}, x^\star)^2 \;\le\; \zeta\eta^2\,\|g_k\|^2 + d(x_k, x^\star)^2 + 2\eta\,\langle g_k, \mathrm{Log}_{x_k}(x^\star)\rangle\,.$$

Step 3 (conditional expectations). Condition on $\mathcal{F}_k$. By (S1), strong convexity (6) and the first inequality in (7),

$$\begin{aligned}\mathbb{E}\big[\,\langle g_k, \mathrm{Log}_{x_k}(x^\star)\rangle \,\big|\, \mathcal{F}_k\big] &= \langle \mathrm{grad}\, f(x_k), \mathrm{Log}_{x_k}(x^\star)\rangle \\ &\le f(x^\star) - f(x_k) - \frac{\mu}{2}\, d(x_k, x^\star)^2, \\ &\le -\mu\, d(x_k, x^\star)^2\end{aligned}$$

and by Assumption C.4 together with the Lipschitz gradient, which gives $\|\mathrm{grad}\, f(x_k)\| \le L\, d(x_k, x^\star)$,

$$\mathbb{E}\big[\,\|g_k\|^2 \,\big|\, \mathcal{F}_k\big] \;\le\; \|\mathrm{grad}\, f(x_k)\|^2 + \sigma_N^2 \;\le\; L^2\, d(x_k, x^\star)^2 + \sigma_N^2.$$

Step 4 (contraction). Inserting Step 3 into Step 2 and using Step 1,
$\mathbb{E}\big[d(x_{k+1}, x^\star)^2 \,\big|\, \mathcal{F}_k\big] \;\le\; \big(1 - 2\mu\eta + \zeta\eta^2 L^2\big)\, d(x_k, x^\star)^2 + \zeta\eta^2\sigma_N^2 \;\le\; (1-\mu\eta)\, d(x_k, x^\star)^2 + \zeta\eta^2\sigma_N^2$,
since $\zeta\eta L^2 \le \mu$. Taking expectations gives (27); unrolling the recursion and bounding the geometric sum $\sum_{j<K}(1-\mu\eta)^j$ by $1/(\mu\eta)$ gives (28). □

Corollary C.19 (budgets under strong geodesic convexity). *In the setting of Theorem C.18, take $\eta = \mu/(\zeta L^2)$, a net size N with $\sigma_N^2 \le L^2\varepsilon^2/2$, and $K = \max\{0, \lceil (\zeta L^2/\mu^2)\log(2d(x_0,x^\star)^2/\varepsilon^2)\rceil\}$ (return x_0 if its distance is already at most ε). Then $\mathbb{E}\, d(x_K, x^\star)^2 \le \varepsilon^2$ with*

$$KN = \widetilde{O}\Big(\zeta\,\frac{L^2}{\mu^2}\,N(\varepsilon)\Big), \qquad \textit{where}$$

$$N(\varepsilon) = \begin{cases} 2\Gamma\sigma_F^2/(L^2\varepsilon^2) & \textit{(i.i.d.: } \Gamma = 1\textit{; general scrambling: } \Gamma\textit{)},\\ \widetilde{O}\big((L\varepsilon)^{-2/3}\big) & \textit{(scrambled nets, smooth integrand)},\\ O\big((L\varepsilon)^{-2/(2\alpha+1)-\delta}\big) & \textit{(higher-order scrambled nets of order } \alpha\textit{)}. \end{cases}$$

The curvature constant $\zeta = \sqrt{\kappa}D_X/\tanh(\sqrt{\kappa}D_X)$ multiplies the iteration count, exactly as in Zhang & Sra (2016), while the sampler determines $N(\varepsilon)$: the two ingredients separate. (The iteration count carries $(L/\mu)^2$ rather than the L/μ of exact gradient descent; this is the condition-number dependence of the present contraction argument, which absorbs the variance term with a constant step, and it is independent of the sampler.) For i.i.d. minibatches the total cost is $\widetilde{O}(\zeta\sigma_F^2/(\mu^2\varepsilon^2))$, the classical ε^{-2}; with scrambled nets and a smooth integrand it becomes $\widetilde{O}(\zeta(L/\mu)^2(L\varepsilon)^{-2/3})$. It is worth comparing the two routes of this paper on the same problem: the deterministic route of Appendix C.4 costs $\widetilde{O}(V_1/(\mu\varepsilon))$ gradient evaluations for ε-accuracy in distance, the randomized route $\widetilde{O}(\varepsilon^{-2/3})$, so under strong geodesic convexity the scrambled net dominates the deterministic point set, because the variance of a scrambled net decays faster than the square of the discrepancy. A third route, in which one scrambled net is drawn once and the resulting sample objective is minimized exactly, is covered by Theorem C.13(b) whenever the sample objective is strongly convex; it needs the variance bound only at $x^\star$, and no iteration-wise randomization.

Remark C.20 (retractions and vector transports). The results of Appendices C.3 and C.5 were stated for the exponential map. For a general retraction R, Definition B.2 is replaced by L-smoothness with respect to R (Boumal, 2023), under which the descent lemma and Theorem C.15 hold verbatim; Theorem C.18 uses the trigonometric comparison and therefore the exponential map. The readings of the field are independent of the retraction.

C.6 Sturm's contraction as a reading of the field

Proposition C.21 (barycentres contract in nonpositive curvature). *Let $\mathcal{M}$ be a Hadamard manifold and ν, ν' Borel probability measures on $\mathcal{M}$ with finite second moments, with barycentres $b(\nu), b(\nu')$, the minimizers of $f_\nu(x) = \frac{1}{2}\int d(x,y)^2\,\mathrm{d}\nu(y)$ and $f_{\nu'}$. Then*

$$d\big(b(\nu), b(\nu')\big) \le W_1(\nu,\nu') \qquad \textit{and} \qquad d\big(b(\nu), b(\nu')\big) \le \sup_{\|v\|=1} \Big\|\big\langle \mathrm{Log}_{b(\nu)}, v\big\rangle\Big\|_H\, D_H(\nu')$$

for every certificate H of functions on $\mathcal{M}$ with $D_H(\nu') = \sup\{|\nu' g - \nu g| : \|g\|_H \le 1\}$.

Proof. On a Hadamard manifold $\frac{1}{2}d(\cdot,y)^2$ is smooth and 1-strongly geodesically convex with gradient $-\mathrm{Log}_x(y)$ of norm $d(x,y)$ (Appendix B.1); with finite second moments, dominated differentiation gives $f_{\nu'} \in C^1$, 1-strongly convex, with $\mathrm{grad}\, f_{\nu'}(x) = -\int \mathrm{Log}_x\,\mathrm{d}\nu'$, and the same for f_ν. Proposition C.7(c) with $C = \mathcal{M}$, $f = f_\nu$, $x^\star = b(\nu)$, $f_Q = f_{\nu'}$ and $\mu_0 = 1$ gives $d(b(\nu'), b(\nu)) \le \Big\|\int \mathrm{Log}_{b(\nu)}\,\mathrm{d}(\nu' - \nu)\Big\|$, and scalarizing in the direction of this vector gives the second bound. For the first, the map $y \mapsto \mathrm{Log}_x(y)$ is 1-Lipschitz on a Hadamard manifold: in the geodesic triangle (x,y,y') the angle at x is at most the Euclidean comparison angle (Sturm, 2003; Bačák, 2014), so $d(y,y')^2 \ge \|\mathrm{Log}_x\, y\|^2 + \|\mathrm{Log}_x\, y'\|^2 - 2\langle \mathrm{Log}_x\, y, \mathrm{Log}_x\, y'\rangle = \|\mathrm{Log}_x\, y - \mathrm{Log}_x\, y'\|^2$. Hence $y \mapsto \big\langle \mathrm{Log}_{b(\nu)}\, y, v\big\rangle$ is 1-Lipschitz for unit v, and the Kantorovich–Rubinstein duality bounds its integral against $\nu' - \nu$ by $W_1(\nu,\nu')$. □

The first inequality is the fundamental contraction property of Sturm (2003), proved there for all spaces of nonpositive curvature by a different argument; here it is the one-point reading with the Lipschitz certificate. The second, with H the Hardy–Krause variation of $\mathrm{Log}_{b(\nu)} \circ T$ for a transport

T, is the QMC rate of Theorem E.3. The Lipschitz certificate pays the curse of dimension of W_1; the Hardy–Krause certificate pays regularity in the integration variable instead.

D THE SCALING LAW: PROOFS

Throughout, (Ω, P) is a probability space, H a class of functions on Ω with a seminorm $\|\cdot\|_H$ vanishing on constants, extended to $H + \mathbb{R}$ by $\|a + c\|_H = \|a\|_H$, and $D_H(Q) = \sup\{|Qa - Pa| : \|a\|_H \le 1\}$ for a probability measure Q. A *positive rule* is a probability measure with finite support; for equal weights it is the empirical measure of its nodes. Since Q and P are probability measures, $(Q - P)(a + c) = (Q - P)a$, so $|(Q - P)g| \le \|g\|_H D_H(Q)$ for $g \in H + \mathbb{R}$.

D.1 THE SANDWICH AND ITS EXACTNESS

Lemma D.1 (sandwich). *Let Q be a probability measure with $D_H(Q) < \infty$, g bounded and measurable, and $g_\pm \in H + \mathbb{R}$ with $g_- \le g \le g_+$. Then*

$$-\|g_-\|_H D_H(Q) - P(g - g_-) \;\le\; Qg - Pg \;\le\; \|g_+\|_H D_H(Q) + P(g_+ - g).$$

Consequently $|Qg - Pg| \le \mathcal{S}_g(D_H(Q))$, where $\mathcal{S}_g = \max\{\mathcal{S}_g^+, \mathcal{S}_g^-\}$,

$$\mathcal{S}_g^+(t) = \inf\left\{t\,\|g_+\|_H + P(g_+ - g) :\; g_+ \in H + \mathbb{R},\; g_+ \ge g\right\}, \qquad \mathcal{S}_g^- = \mathcal{S}_{-g}^+.$$

Proof. Since $Q \ge 0$ and $g \le g_+$, $Qg \le Qg_+ = Pg_+ + (Q - P)g_+ \le Pg_+ + \|g_+\|_H D_H(Q)$; subtract Pg. The lower bound is the upper bound for $-g$. □

Positivity is used exactly once, in $Qg \le Qg_+$. Without it the certificate says nothing about an integrand outside H: a signed rule can match every function of H and still integrate g badly.

Proposition D.2 (the sandwich is exact). *Let Ω be a compact metric space, P a Borel probability measure, and H a reproducing kernel Hilbert space on Ω with continuous kernel K, so that $D_H(\nu) = \left\|\int K(\cdot, y)\,\mathrm{d}(\nu - P)(y)\right\|_H$. Let g be bounded and upper semicontinuous and $t > 0$. Then*

$$\max\left\{\nu g - Pg : \nu \in \mathcal{P}(\Omega), D_H(\nu) \le t\right\} = \inf\left\{t\,\|a\|_H + P(a + c - g) : a \in H, c \in \mathbb{R}, a + c \ge g\right\}. \tag{29}$$

Moreover, for every $\epsilon > 0$ there is an equal-weight rule Q, with nodes counted with multiplicity, such that $D_H(Q) \le t$ and $Qg - Pg$ is within ϵ of the maximum.

Proof. Equip $\mathcal{P}(\Omega)$ with the weak topology, in which it is a compact convex subset of the locally convex space of finite signed measures with the weak-$*$ topology. The map $\nu \mapsto D_H(\nu)^2 = \iint K\,\mathrm{d}(\nu - P)\,\mathrm{d}(\nu - P)$ is continuous, since K is bounded and continuous on $\Omega \times \Omega$ and $\nu_n \otimes \nu_n \to \nu \otimes \nu$ weakly when $\nu_n \to \nu$; and $\nu \mapsto \nu g$ is upper semicontinuous for bounded upper semicontinuous g. The feasible set is therefore compact and nonempty ($\nu = P$), and the maximum on the left of (29) is attained. Call it V.

Weak duality. If $D_H(\nu) \le t$ and $a + c \ge g$, then $\nu g - Pg \le \nu(a + c) - Pg = (\nu - P)a + P(a + c - g) \le t\,\|a\|_H + P(a + c - g)$.

Strong duality. Let $L(\nu, a) = (\nu - P)(g - a) + t\,\|a\|_H$ on $\mathcal{P}(\Omega) \times H$. For fixed a, $L(\cdot, a)$ is affine and upper semicontinuous; for fixed ν, $L(\nu, \cdot)$ is convex and continuous on H, because $|(\nu - P)a| \le \|a\|_H D_H(\nu)$ and $D_H(\nu) \le 2\sup_x K(x, x)^{1/2}$. Sion's minimax theorem (Sion, 1958), applied to $-L$ with the compact convex set $\mathcal{P}(\Omega)$, gives $\max_\nu \inf_a L = \inf_a \max_\nu L$. On the left, $\inf_a\left(t\,\|a\|_H - (\nu - P)a\right)$ is 0 if $D_H(\nu) \le t$ and $-\infty$ otherwise (take a a multiple of the representer $\int K(\cdot, y)\,\mathrm{d}(\nu - P)(y)$), so $\max_\nu \inf_a L = V$. On the right, the maximum over probability measures of $\nu(g - a)$ is $\max_\Omega(g - a)$, attained at a Dirac mass since $g - a$ is upper semicontinuous, so $\max_\nu L(\nu, a) = t\,\|a\|_H + \max_\Omega(g - a) - P(g - a)$, which is the right side of (29) at $c = \max_\Omega(g - a)$, the smallest admissible c. Hence the two sides agree.

Equal weights. Let $\nu^\star$ attain V and $\nu_s = (1 - s)\nu^\star + sP$, $s \in (0, 1)$, so that $D_H(\nu_s) = (1 - s)D_H(\nu^\star) < t$ (or $= 0$) and $\nu_s g - Pg = (1 - s)V$. Partition Ω into finitely many Borel cells

$A_1, \dots, A_M$ of diameter at most δ and let x_k maximize g over the compact set $\overline{A_k}$, which is possible because g is upper semicontinuous. Then $\mu_\delta = \sum_k \nu_s(A_k)\delta_{x_k}$ satisfies $\mu_\delta g \geq \nu_s g$, and $|\mu_\delta\varphi - \nu_s\varphi| \leq \delta \operatorname{Lip}(\varphi)$ for Lipschitz φ, so $\mu_\delta \to \nu_s$ weakly as $\delta \to 0$ and $D_H(\mu_\delta) \to D_H(\nu_s)$ by the continuity shown above. Round the weights $\nu_s(A_k)$ to multiples n_k/n of $1/n$ with $\sum_k |n_k/n - \nu_s(A_k)| \leq M/n$ and let Q be the equal-weight rule with node x_k repeated n_k times. Since $\left\| \int K(\cdot, y)\, \mathrm{d}\eta(y) \right\|_H^2 = \iint K \,\mathrm{d}\eta \,\mathrm{d}\eta \leq \|K\|_\infty \|\eta\|_{\mathrm{TV}}^2$ for signed η, $D_H(Q) \leq D_H(\mu_\delta) + \|K\|_\infty^{1/2} M/n$ and $Qg \geq \mu_\delta g - \|g\|_\infty M/n$. Choosing δ small and then n large makes Q feasible with $Qg - Pg \geq (1-s)V - \epsilon/2$; let $s \downarrow 0$. □

The proposition identifies the certificate exponent of an integrand with the scaling exponent of its one-sided modulus, and it makes the two halves of every sharp statement below dual to each other: sandwiches give upper bounds, measures give lower bounds, and at the optimum they meet. For the discontinuous integrands of this paper we construct the lower-bound measures explicitly (Theorem D.11).

D.2 Dilation

On $\mathbb{R}^d$ we use $\hat{a}(\xi) = \int a(x) e^{-ix\cdot\xi}\,\mathrm{d}x$ and $\|a\|_{H^\tau(\mathbb{R}^d)}^2 = \int (1 + \|\xi\|^2)^\tau |\hat{a}(\xi)|^2\,\mathrm{d}\xi$.

Lemma D.3 (dilation). *For $\varphi \in H^\tau(\mathbb{R}^d)$, $\tau \geq 0$, and $\lambda > 0$, the dilation $\varphi_\lambda = \varphi(\cdot/\lambda)$ satisfies $\|\varphi_\lambda\|_{H^\tau(\mathbb{R}^d)} \leq \max\{\lambda^{d/2}, \lambda^{d/2-\tau}\}\, \|\varphi\|_{H^\tau(\mathbb{R}^d)}$.*

Proof. $\widehat{\varphi_\lambda}(\xi) = \lambda^d \hat{\varphi}(\lambda\xi)$, so $\|\varphi_\lambda\|^2 = \lambda^d \int (1 + \|\eta\|^2/\lambda^2)^\tau |\hat{\varphi}(\eta)|^2\,\mathrm{d}\eta$, and $1 + \|\eta\|^2/\lambda^2 \leq \max\{1, \lambda^{-2}\}(1 + \|\eta\|^2)$. □

Lemma D.4 (dyadic shells). *Let $m \geq 0$, $\tau \geq 0$, and let $k \geq \tau$ be an integer. Let $\omega \in C^k(\mathbb{S}^{d-1})$, $b \in C_c^k(\mathbb{R}^d)$ with support in $B(0, R_0)$, and $\chi \in C^\infty([0,\infty); [0,1])$ with $\chi = 0$ on $[0,1]$ and $\chi = 1$ on $[2,\infty)$. For $0 < r \leq 1$ put $g_r(z) = \chi(\|z\|/r)\, b(z)\, \|z\|^m\, \omega(z/\|z\|)$. Then, with $\alpha = m + d/2 - \tau$,*

$$\|g_r\|_{H^\tau(\mathbb{R}^d)} \leq C\, \ell_\alpha(r), \qquad \ell_\alpha(r) = \begin{cases} r^\alpha, & \alpha < 0, \\ 1 + \log(1/r), & \alpha = 0, \\ 1, & \alpha > 0, \end{cases}$$

where C depends only on d, m, τ, k, R_0, χ and the C^k norms of b and ω. If $\alpha = 0$ and $k > 2\tau + d$, the bound improves to $C\,(1 + \log(1/r))^{1/2}$.

Proof. Fix $\eta \in C_c^\infty(\{1/2 < \|w\| < 2\})$ with $\sum_{j\in\mathbb{Z}} \eta(2^{-j}w) = 1$ for $w \neq 0$. The terms $g_r(z)\eta(z/(2^j r))$ vanish for $j \leq -1$, whose support lies in $\{\|z\| < r\}$ where $\chi(\|z\|/r) = 0$, and for $2^{j-1}r \geq R_0$, whose support lies outside $B(0, R_0)$; so $g_r = \sum_{j=0}^{J} g_r\, \eta(\cdot/(2^j r))$ with $2^{J-1}r < R_0$. The j-th term is $(2^j r)^m \Phi_j(z/(2^j r))$ with

$$\Phi_j(w) = \chi(2^j \|w\|)\, b(2^j r w)\, \|w\|^m\, \omega(w/\|w\|)\, \eta(w),$$

supported in $\{1/2 \leq \|w\| \leq 2\}$. On this annulus $\|w\|^m\, \omega(w/\|w\|)\eta(w)$ is a fixed C^k function; $b(2^j r\,\cdot)$ has C^k norm at most $\max\{1, 4R_0\}^k \|b\|_{C^k}$ because $2^j r < 4R_0$; and $\chi(2^j\|w\|)$ equals 1 there for $j \geq 2$ and is one of two fixed smooth functions for $j \in \{0, 1\}$. Hence $\|\Phi_j\|_{H^\tau} \leq \|\Phi_j\|_{H^k} \leq C_1$ uniformly in j and r, and Lemma D.3 with $\lambda = 2^j r$ bounds the j-th term by $C_1(2^j r)^\alpha$ if $2^j r \leq 1$ and by $C_1(4R_0)^{m+d/2}$ otherwise. There are at most $\log_2(4R_0) + 2$ terms of the second kind. For the first kind, $\sum_{j\geq 0,\, 2^j r \leq 1} (2^j r)^\alpha$ is at most $r^\alpha/(1 - 2^\alpha)$ if $\alpha < 0$, at most $1 + \log_2(1/r)$ if $\alpha = 0$, and at most $1/(1 - 2^{-\alpha})$ if $\alpha > 0$.

If $\alpha = 0$ and $k > 2\tau + d$, the terms of the first kind are almost orthogonal. Write $\varphi_j = \lambda_j^m \Phi_j(\cdot/\lambda_j)$ with $\lambda_j = 2^j r$, so that $\hat{\varphi}_j(\xi) = \lambda_j^{m+d} \hat{\Phi}_j(\lambda_j \xi)$, and let $\lambda_j \leq \lambda_{j'} \leq 1$. Since $(1 + \|\xi\|^2)^\tau \leq$

$2^\tau(1+\|\xi\|^{2\tau})$, the substitution $\eta=\lambda_{j'}\xi$ and $\left\|\hat\Phi_j\right\|_\infty \le \|\Phi_j\|_{L^1}$ give

$$|\langle\varphi_j,\varphi_{j'}\rangle_{H^\tau}| \le 2^\tau\Big(\frac{\lambda_j}{\lambda_{j'}}\Big)^{m+d}\|\Phi_j\|_{L^1}\Big(\lambda_{j'}^{2m+d}\left\|\hat\Phi_{j'}\right\|_{L^1}+\lambda_{j'}^{2m+d-2\tau}\int\|\eta\|^{2\tau}\,|\hat\Phi_{j'}(\eta)|\,\mathrm{d}\eta\Big) \le C\,2^{-(m+d)|j-j'|},$$

because $2m+d-2\tau=2\alpha=0$, $\lambda_{j'}\le 1$, and $|\hat\Phi_{j'}(\eta)|\le C(1+\|\eta\|)^{-k}$ with $k>2\tau+d$. Hence $\left\|\sum_j\varphi_j\right\|_{H^\tau}^2 \le \sum_{j,j'}|\langle\varphi_j,\varphi_{j'}\rangle_{H^\tau}| \le C(1+\log_2(1/r))\sum_{i\in\mathbb{Z}}2^{-(m+d)|i|}$ over the terms of the first kind. □

The geometric sum is the whole mechanism: each shell is a dilation of one profile, so the shell at scale ρ costs ρ^α, and the cost of cutting at scale r is dominated by the innermost shell when $\alpha<0$ and by the outermost when $\alpha>0$. When $\alpha=0$ every shell costs the same, and what decides is how shells add: almost orthogonally in a Hilbert norm, which gives the square root of the number of shells, and linearly in Hardy–Krause variation, an L^1 norm of the mixed derivative, which gives the logarithm itself (Theorem D.10(a)). Both orders are attained (Theorems D.11 and D.12).

Lemma D.5 (localization on spheres). *Let $\tau\ge 0$ and $k\ge\tau$ an integer. There is C such that for every $p\in\mathbb{S}^d$ and every a supported in the geodesic ball $B(p,\pi/2)$, $\|a\|_{H^\tau(\mathbb{S}^d)}\le C\left\|a\circ\mathrm{Exp}_p\right\|_{H^\tau(\mathbb{R}^d)}$, where $a\circ\mathrm{Exp}_p$ is extended by zero outside $B(0,\pi/2)\subset T_p\mathbb{S}^d\cong\mathbb{R}^d$; and $\|a\|_{H^\tau(\mathbb{S}^d)}\le C\,\|a\|_{C^k(\mathbb{S}^d)}$ for every $a\in C^k(\mathbb{S}^d)$. Conversely, for a fixed $\chi\in C_c^\infty(B(0,1))$ there is C with $\left\|\chi\cdot(a\circ\mathrm{Exp}_p)\right\|_{H^\tau(\mathbb{R}^d)}\le C\,\|a\|_{H^\tau(\mathbb{S}^d)}$ for every $p\in\mathbb{S}^d$ and $a\in H^\tau(\mathbb{S}^d)$.*

Proof. On a compact manifold the Sobolev spaces defined by the spectral norm coincide with those defined through a finite atlas and a subordinate partition of unity, with equivalent norms (Triebel, 1992, Ch. 7). The normal-coordinate chart Exp_p on $B(0,\pi/2)$ is one such chart, with all derivatives bounded independently of p because $\mathrm{Exp}_{Rp}(Rz)=R\,\mathrm{Exp}_p(z)$ for rotations R, and the spectral norm is rotation invariant; hence C does not depend on p. The second bound follows from the first applied to a partition of unity and from $\|\cdot\|_{H^\tau(\mathbb{R}^d)}\le\|\cdot\|_{H^k(\mathbb{R}^d)}$ for compactly supported functions. The converse is the other half of the same equivalence: include Exp_p on $B(0,\pi/2)$ in the atlas with a partition-of-unity function equal to 1 on $B(p,1)$, and use that multiplication by χ is bounded on $H^\tau(\mathbb{R}^d)$. □

D.3 The law on spheres

Theorem D.6 (scaling law on $\mathbb{S}^d$). *Let $d\ge 1$, $\tau>d/2$, $0\le m\le\tau-d/2$, and $k\ge\tau$ an integer, with $k>2\tau+d$ if $m=\tau-d/2$. Let $\mathcal{G}$ be a family of bounded functions on $\mathbb{S}^d$ with the following property: every $g\in\mathcal{G}$ is $g=a_g+s_g$ with $\|a_g\|_{H^\tau(\mathbb{S}^d)}\le A$, and for some $p_g\in\mathbb{S}^d$*

$$s_g\circ\mathrm{Exp}_{p_g}(z)=\sum_{j=1}^{J}b_{g,j}(z)\,\|z\|^{m_j}\,\omega_{g,j}(z/\|z\|),\qquad m_j\ge m,$$

with $b_{g,j}\in C_c^k(B(0,\pi/4))$, $\omega_{g,j}\in C^k(\mathbb{S}^{d-1})$, all C^k norms at most B, $s_g=0$ outside $B(p_g,\pi/4)$, and $|s_g(p_g)|\le JB^2$ (the formula leaves the value at p_g free when some $m_j=0$). Then there is C, depending on d,τ,k,J,B,m, the m_j and the fixed cutoffs χ,ϑ,η of the proof only, such that every probability measure Q on $\mathbb{S}^d$ with finite support and $D=\mathrm{wce}_{H^\tau}(Q)\le 1$ satisfies

$$\sup_{g\in\mathcal{G}}\big|Qg-\sigma g\big|\le A\,D+C\,D^{(m+d)/(\tau+d/2)}\,\ell(D),\quad \ell(D)=\begin{cases}1, & m<\tau-d/2,\\ (1+\log\frac{1}{D})^{1/2}, & m=\tau-d/2.\end{cases}$$

At $m=\tau-d/2$ the exponent is 1 and the factor $(1+\log(1/D))^{1/2}$ is the price of the borderline; Theorem D.11 shows that it cannot be removed.

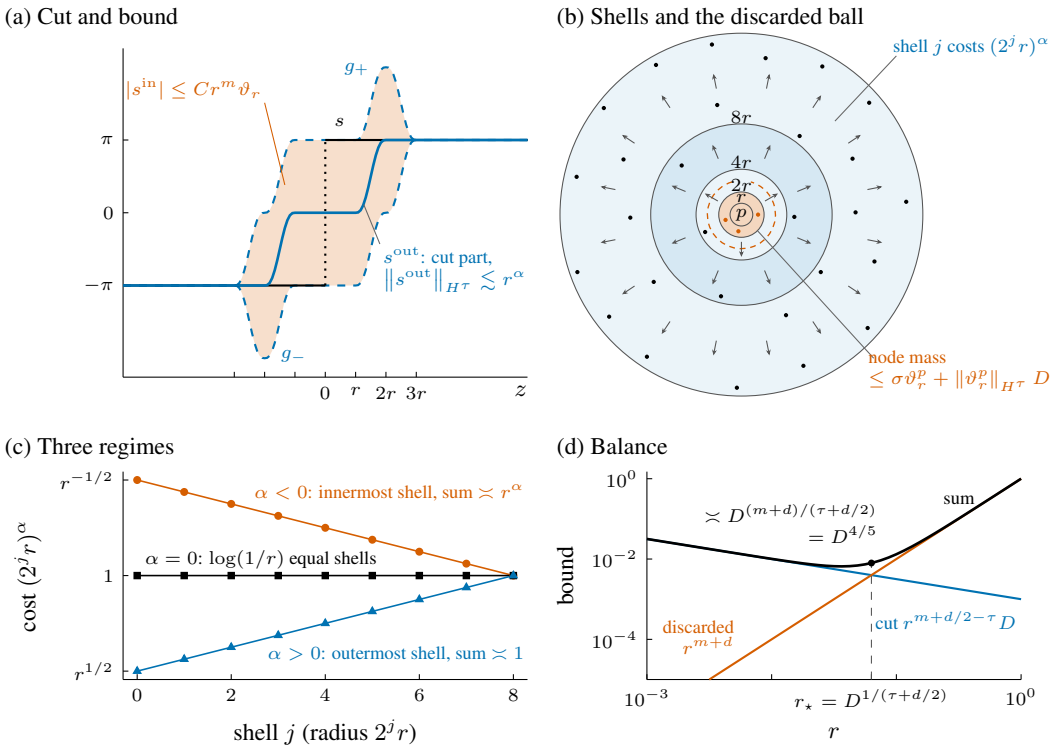

Figure 6: **The scaling law is a balance between a cut and a bump.** (a) The sandwich of Lemma D.1 in the proof of Theorem D.6, for $d = 1$, $s = \pi \operatorname{sgn} z$ ($m = 0$), $r = 0.15$: the singular part s is cut at distance r; the cut part s^{out} (blue; $= s$ for $|z| \ge 2r$) lies in H^τ, the discarded part is bounded by $Cr^m\vartheta_r$, so $g_\pm = s^{\text{out}} \pm Cr^m\vartheta_r$ enclose s; positivity turns the bump into a bound on the node mass near p. (b) Lemma D.4: the cut part is a sum over dyadic shells, each a dilation of one profile (Lemma D.3); the dashed circle encloses $\operatorname{supp}\vartheta_r^p$. (c) With $\alpha = m + d/2 - \tau$, the innermost shell dominates if $\alpha < 0$, the outermost if $\alpha > 0$; at $\alpha = 0$ the shells add almost orthogonally in H^τ, giving $(\log(1/r))^{1/2}$, and linearly in Hardy–Krause variation, giving $\log(1/r)$ (Theorem D.10). (d) The costs balance at $r_\star$, giving the exponent $(m+d)/(\tau+d/2)$: $4/5$ for $d = 2$, $\tau = 3/2$, $m = 0$ ($D = 10^{-3}$), the field exponent of the cap discrepancy (Corollary D.7).

Proof. Fix g and write $s = s_g$, $p = p_g$, $b_j = b_{g,j}$, $\omega_j = \omega_{g,j}$. The part a_g contributes at most AD. Let $r_0 = \pi/24$ and $0 < r \le r_0$, and let χ be as in Lemma D.4. Split $s = s^{\text{out}} + s^{\text{in}}$ with $s^{\text{out}} \circ \operatorname{Exp}_p = \sum_j \chi(\|z\|/r) b_j \|z\|^{m_j} \omega_j(z/\|z\|)$; then s^{in} vanishes outside $B(p, 2r)$ and $|s^{\text{in}}| \le JB^2(2r)^m \le C_1 r^m$, because $m_j \ge m$ and $2r \le 1$.

The cut part. By Lemmas D.4 and D.5, $\|s^{\text{out}}\|_{H^\tau(\mathbb{S}^d)} \le C_2 \sum_j \ell_{m_j+d/2-\tau}(r) \le C_3 r^{m+d/2-\tau}\ell(r^{\tau+d/2})$: for $m_j < \tau - d/2$ this uses $m_j \ge m$ and $r \le 1$; if $m < \tau - d/2$ and $m_j \ge \tau - d/2$, it uses that $\ell_{m_j+d/2-\tau}(r) \le 1 + \log(1/r)$ is at most a constant times $r^{m+d/2-\tau}$, whose exponent is negative; if $m = \tau - d/2$, every term is at most $C(1+\log(1/r))^{1/2} \le C\ell(r^{\tau+d/2})$ by the improved bound of Lemma D.4. Hence $|(Q-\sigma)s^{\text{out}}| \le C_3 r^{m+d/2-\tau}\ell(r^{\tau+d/2})D$.

The discarded part. Let $\vartheta \in C_c^\infty(B(0,3))$ with $0 \le \vartheta \le 1$ and $\vartheta = 1$ on $B(0,2)$, and $\vartheta_r^p = \vartheta(\operatorname{Exp}_p^{-1}(\cdot)/r)$ on $B(p, \pi/2)$, zero elsewhere. Then $\mathbf{1}_{B(p,2r)} \le \vartheta_r^p$, $\sigma\vartheta_r^p \le \sigma(B(p,3r)) \le C_4 r^d$, and $\|\vartheta_r^p\|_{H^\tau(\mathbb{S}^d)} \le C_5 r^{d/2-\tau}$ by Lemmas D.3 and D.5. Because $Q \ge 0$,

$$|Qs^{\text{in}}| \le C_1 r^m\, Q\mathbf{1}_{B(p,2r)} \le C_1 r^m\, Q\vartheta_r^p \le C_1 r^m \big(\sigma\vartheta_r^p + \|\vartheta_r^p\|_{H^\tau} D\big) \le C_1 r^m \big(C_4 r^d + C_5 r^{d/2-\tau} D\big),$$

and $|\sigma s^{\text{in}}| \le C_1 C_4 r^{m+d}$.

Balance. Altogether $|(Q-\sigma)s| \le C_6\big(r^{m+d/2-\tau}\ell(r^{\tau+d/2})D + r^{m+d}\big)$. If $D \le r_0^{\tau+d/2}$, take $r = D^{1/(\tau+d/2)}$; both terms are at most $D^{(m+d)/(\tau+d/2)}\ell(D)$. Otherwise $|(Q-\sigma)s| \le 2\sup|s| \le C_7 \le C_7 r_0^{-(m+d)} D^{(m+d)/(\tau+d/2)}$; here $\sup|s| \le JB^2$ uses the convention at p_g. □

Figure 6 draws the proof: the cut and the bump, the dyadic shells and their three regimes, and the balance of the two costs.

On $\mathbb{S}^d$ the field integrand of the Fréchet loss has the required form. For $x \in \mathbb{S}^d$ and $z \in T_{-x}\mathbb{S}^d = x^\perp$ with $0 < \|z\| < \pi$,

$$\operatorname{Log}_x\big(\operatorname{Exp}_{-x} z\big) = (\pi - \|z\|)\,\frac{z}{\|z\|} = \pi\frac{z}{\|z\|} - z. \tag{30}$$

Indeed $y = \operatorname{Exp}_{-x} z = -\cos\|z\|\, x + \sin\|z\|\, z/\|z\|$ lies on the great circle through $-x$ and x in the plane of x and z, at distance $\pi - \|z\| < \pi$ from x, and writing $y = \cos\theta\, x + \sin\theta\, u$ with $\theta = \pi - \|z\|$ gives $u = z/\|z\|$.

Corollary D.7 (the Fréchet certificates on spheres)**.** *Let $\tau > d/2$, $k \ge \tau+1$ an integer, with $k > 2\tau + d$ if $\tau = \frac{d}{2}+1$, $h \in C^k(\mathbb{S}^d)$, $F(x,y) = h(y)\frac{1}{2}d(x,y)^2$, and let Q be a positive rule with $D = \operatorname{wce}_{H^\tau}(Q) \le 1$. Write $\Phi(x) = (Q-\sigma)F(x,\cdot)$ and $e(x) = \operatorname{grad}\Phi(x)$ for $x \notin -\operatorname{supp} Q$. Then*

$$\sup_{x\notin -\operatorname{supp} Q}\|e(x)\| \le C_h\, D^{d/(\tau+d/2)}, \qquad \sup_{x\in\mathbb{S}^d}|\Phi(x)| \le \begin{cases} C_h\, D, & \tau < \frac{d}{2}+1, \\ C_h\, D\,(1+\log\frac{1}{D})^{1/2}, & \tau = \frac{d}{2}+1, \\ C_h\, D^{(d+1)/(\tau+d/2)}, & \tau > \frac{d}{2}+1. \end{cases}$$

For the cap discrepancy, $\tau = (d+1)/2$ and $D \le c_0 D_{L_2}$, the exponents are $2d/(2d+1)$ for the field and 1 for the value; on $\mathbb{S}^2$ the field bound is $C_h D_{L_2}^{4/5}$.

Proof. For x and a unit vector $v \in T_x\mathbb{S}^d$, $\langle e(x), v\rangle = (Q-\sigma)g_{x,v}$ with $g_{x,v} = -h\,\langle \operatorname{Log}_x, v\rangle$, a bounded function (Appendix C). Let $p = -x$ and $\psi \in C^\infty([0,\infty);[0,1])$ with $\psi = 1$ on $[0,\pi/8]$ and $\psi = 0$ on $[\pi/5,\infty)$. By (30), with $\tilde h = h\circ\operatorname{Exp}_p$ and $T_x\mathbb{S}^d = T_p\mathbb{S}^d$,

$$g_{x,v}\circ\operatorname{Exp}_p(z) = \underbrace{-\pi\psi(\|z\|)\tilde h(z)\,\langle z/\|z\|, v\rangle}_{s} + \underbrace{\psi(\|z\|)\tilde h(z)\,\langle z, v\rangle + \big(1-\psi(\|z\|)\big)g_{x,v}(\operatorname{Exp}_p z)}_{a}.$$

The first part has the form of Theorem D.6 with $J=1$, $m_1 = 0$, $b = -\pi\psi\tilde h$ and $\omega(\theta) = \langle\theta, v\rangle$; the second is in $C^k(\mathbb{S}^d)$, because Log_x is smooth on $\mathbb{S}^d\setminus\{-x\}$, and by compactness and rotation invariance its C^k norm, and the C^k norms of b and ω, are bounded uniformly in x and v. Theorem D.6 and Lemma D.5 give $|\langle e(x), v\rangle| \le AD + CD^{d/(\tau+d/2)} \le C_h D^{d/(\tau+d/2)}$, since $D \le 1$ and $d/(\tau+d/2) < 1$; take $v = e(x)/\|e(x)\|$. For the value, $F(x,\cdot)\circ\operatorname{Exp}_p(z) = \tilde h(z)\big(\frac{\pi^2}{2} + \frac{1}{2}\|z\|^2\big) - \pi\tilde h(z)\|z\|$ has a singularity of order $m=1$: if $1 > \tau - d/2$ it lies in $H^\tau(\mathbb{S}^d)$, with norm bounded uniformly in x, by Lemma D.4 with $\alpha > 0$ and $r \to 0$, and the value error is at most $C_h D$; if $1 \le \tau - d/2$, Theorem D.6 applies with $m = 1$, and at $1 = \tau - d/2$ it carries the factor $(1+\log(1/D))^{1/2}$, which Theorem D.11 shows cannot be removed; for $d = 1$ the construction of that theorem is evaluated exactly in Appendix G.8. □

Corollary D.8 (minimizers from the uniform field bound)**.** *In the setting of Corollary D.7, let $C \subseteq \mathbb{S}^d$ be closed and geodesically convex, $f = \sigma F$ μ-strongly geodesically convex on C with minimizer $x^\star$ in the interior of C, and $\hat x \in \operatorname{int} C \setminus (-\operatorname{supp} Q)$ a stationary point of QF. Then $d(\hat x, x^\star) \le C_h D^{d/(\tau+d/2)}/\mu$; for the cap discrepancy on $\mathbb{S}^2$ and nodes with $D_{L_2} = O(N^{-3/4})$ this is $O(N^{-3/5})$. If $h(y_i) > 0$ at every node, no point of $-\operatorname{supp} Q$ is a local minimizer of QF.*

Proof. By (7), $\mu\, d(\hat x, x^\star) \le \|\operatorname{grad} f(\hat x)\| = \|e(\hat x)\|$, and Corollary D.7 bounds the right side. For the last claim, let $x_0 = -y_i$ with $h(y_i) > 0$. By (30), $\frac{1}{2}d(x,y_i)^2 = \frac{1}{2}(\pi - d(x,x_0))^2$ near x_0, so this term of QF decreases at rate $\pi h(y_i)Q(\{y_i\})$ along every geodesic leaving x_0, while the other terms are differentiable at x_0; moving from x_0 against the gradient of the differentiable part (in any direction if it vanishes) decreases QF. □

Remark D.9 (softer routes give smaller exponents). Positivity can be used more crudely. With nonnegative weights and $h \geq 0$, QF is semiconcave on $\mathbb{S}^d$ with constant $\|h\|_\infty$, because $\frac{1}{2}d(\cdot, y)^2$ is 1-semiconcave (its Hessian is at most the identity off the cut point, where it has a concave cone), and σF is smooth; so Φ is Λ-semiconcave for some Λ. If $|\Phi| \leq \varepsilon$, then at every point of differentiability and for unit v, $\langle e(x), v\rangle \geq (\Phi(\mathrm{Exp}_x(tv)) - \Phi(x))/t - \Lambda t/2 \geq -2\varepsilon/t - \Lambda t/2$, and $t = 2\sqrt{\varepsilon/\Lambda}$ gives $\|e(x)\| \leq 2\sqrt{\Lambda\varepsilon}$. With $\varepsilon = O(D)$ this is $D^{1/2}$: positivity through one-sided second derivatives yields the exponent $1/2$, positivity through node masses in small balls yields $d/(\tau + d/2)$.

D.4 The law on the square

Theorem D.10 (scaling law on $[0,1]^2$)**.** *Let P_N be a point set in $[0,1)^2$, Q its equal-weight rule, λ Lebesgue measure and $D^* = D_N^*(P_N)$.*

(a) Point singularity. *Let $u_0 \in (0,1)^2$, $U \subset (0,1)^2$ an open neighbourhood of u_0, $Z\colon U \to \mathbb{R}^2$ a C^2 diffeomorphism onto its image with $Z(u_0) = 0$, $b \in C_c^2(U)$ and $\omega \in C^2(\mathbb{S}^1)$. If $g = a + b\,\omega(Z/\|Z\|)$ with $V_{\mathrm{HK}}(a) \leq A$, the value of the second term at u_0 being any number of modulus at most $\|b\|_\infty \|\omega\|_\infty$, then $|Qg - \lambda g| \leq AD^* + C\,D^*(1 + \log(1/D^*))$.*

(b) Polygonal jumps. *Let $\Pi_1, \dots, \Pi_K \subseteq [0,1]^2$ be convex polygons and $b_1, \dots, b_K \in C^2([0,1]^2)$. If $g = a + \sum_k b_k \mathbf{1}_{\Pi_k}$ with $V_{\mathrm{HK}}(a) \leq A$, then $|Qg - \lambda g| \leq AD^* + C\,(D^*)^{1/2}$, where C depends only on K and $\max_k \|b_k\|_{C^2}$ (a convex subset of $[0,1]^2$ has perimeter at most 4). This covers jumps across segments oblique to the axes, segments that reach the boundary of the square, and junctions where several segments meet.*

Proof. In part (a) the singular term s has compact support in the open square, so every function built from it below vanishes on the faces $\{u_j = 1\}$, and by (11) its Hardy–Krause variation is $\int_{[0,1]^2} |\partial_1\partial_2(\cdot)|$ whenever it is C^2; in part (b) the face terms are bounded separately.

(a) Since $\operatorname{supp} b$ is a compact subset of U, there are $0 < \kappa_1 \leq \kappa_2$ with $\kappa_1\|u - u_0\| \leq \|Z(u)\| \leq \kappa_2\|u - u_0\|$ on $\operatorname{supp} b$. For $\Theta = Z/\|Z\|$ the chain rule gives $\partial_i\Theta = (I - \Theta\Theta^\top)\partial_i Z/\|Z\|$ and a sum of terms of order $\|Z\|^{-2}$ for $\partial_1\partial_2\Theta$, so $|\partial^\alpha(\omega \circ \Theta)| \leq C\|u - u_0\|^{-|\alpha|}$ for $\alpha \subseteq \{1,2\}$. Let $\vartheta \in C_c^\infty(B(0,3))$ with $0 \leq \vartheta \leq 1$, $\vartheta = 1$ on $B(0,2)$, and for $r \leq r_0$ (with $B(u_0, 5r_0) \subset U$) put $s^{\mathrm{out}} = (1 - \vartheta((u - u_0)/r))s$ and $s^{\mathrm{in}} = s - s^{\mathrm{out}}$. Expanding $\partial_1\partial_2 s^{\mathrm{out}}$ by the product rule, the term with both derivatives on s is at most $C\|u - u_0\|^{-2}$ on $\{\|u - u_0\| \geq 2r\}$ and integrates to at most $C(1 + \log(1/r))$; each term with a derivative on the cutoff is $O(r^{-2})$ on an annulus of area $O(r^2)$. Hence $V_{\mathrm{HK}}(s^{\mathrm{out}}) \leq C(1 + \log(1/r))$. The bump $\Xi_r(u) = \vartheta(2(u - u_0)/(3r))$ satisfies $\Xi_r \geq \mathbf{1}_{B(u_0,3r)} \geq |s^{\mathrm{in}}|/\|s\|_\infty$, $\lambda\Xi_r \leq Cr^2$, and $V_{\mathrm{HK}}(\Xi_r) = \int |\partial_1\partial_2\vartheta|$, independent of r: Hardy–Krause variation is invariant under this dilation. By Lemma D.1 in the form $|(Q - \lambda)s| \leq |(Q - \lambda)s^{\mathrm{out}}| + \|s\|_\infty (Q\Xi_r + \lambda\Xi_r)$ and the Koksma–Hlawka inequality,

$$|(Q - \lambda)s| \leq C(1 + \log(1/r))D^* + \|s\|_\infty \left(2\lambda\Xi_r + V_{\mathrm{HK}}(\Xi_r)D^*\right) \leq C'\big((1 + \log(1/r))D^* + r^2\big),$$

and $r = (D^*)^{1/2}$ gives the claim when $D^* \leq r_0^2$; otherwise the trivial bound $2\|s\|_\infty$ suffices.

(b) By linearity it suffices to treat $s = b\,\mathbf{1}_\Pi$ for one convex polygon Π and $b \in C^2([0,1]^2)$. Fix a radial mollifier $\varphi \in C_c^\infty(B(0,1))$, $\varphi \geq 0$, $\int\varphi = 1$, put $\varphi_r = r^{-2}\varphi(\cdot/r)$, and let Π^{+t} and Π^{-t} be the outer and inner parallel sets of Π at distance t, both convex. The mollified indicator $M_r = \mathbf{1}_\Pi * \varphi_r$ equals $\mathbf{1}_\Pi$ outside the strip $S_r = \Pi^{+r} \setminus \Pi^{-r}$, so $s = b\,M_r + E_r$ with $|E_r| \leq \|b\|_\infty \mathbf{1}_{S_r}$. Let $\Xi_r = \mathbf{1}_{\Pi^{+2r}\setminus\Pi^{-2r}} * \varphi_r$, which is at least $\mathbf{1}_{S_r}$, and note $\lambda(\Pi^{+3r}\setminus\Pi^{-3r}) \leq Cr$ with C depending on the perimeter. *Variation of bM_r.* Its mixed derivative is $(\partial_1\partial_2 b)M_r + \partial_1 b\,\partial_2 M_r + \partial_2 b\,\partial_1 M_r + b\,\partial_1\partial_2 M_r$. The first term is bounded; $\partial_i M_r = \mathbf{1}_\Pi * \partial_i\varphi_r$ and $\partial_1\partial_2 M_r = \mathbf{1}_\Pi * \partial_1\partial_2\varphi_r$ vanish outside S_r, because there $\mathbf{1}_\Pi$ is constant on the support of the mollifier and the derivatives of φ_r integrate to zero, and they are bounded by $\|\partial_i\varphi_r\|_{L^1} \leq Cr^{-1}$ and $\|\partial_1\partial_2\varphi_r\|_{L^1} \leq Cr^{-2}$. Since $\lambda(S_r) \leq Cr$, the integral of the mixed derivative is at most $C(1 + 1 + r^{-1})$. On a face $\{u_j = 1\}$, M_r restricts to $\int \varphi_r(w)\mathbf{1}_{I(w)}\,\mathrm{d}w$ with intervals $I(w)$ (because Π is convex), a mixture of interval indicators whose one-dimensional variation is at most 2; so the face terms of bM_r are bounded, and those of Ξ_r, a difference of two such mixtures, likewise. Hence $V_{\mathrm{HK}}(bM_r) \leq C/r$ by (11), and in the same way $V_{\mathrm{HK}}(\Xi_r) \leq C/r$ and $\lambda\Xi_r \leq Cr$. *Sandwich.* As in (a), $|(Q - \lambda)s| \leq V_{\mathrm{HK}}(bM_r)D^* + \|b\|_\infty (Q\Xi_r + \lambda\Xi_r) \leq$

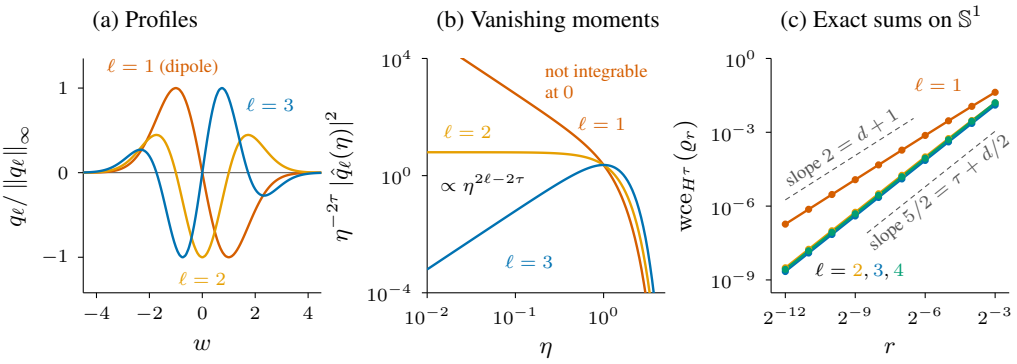

Figure 7: **Multipoles are the lower bounds on spheres.** Theorem D.11 places at the singular point a signed measure ϱ_r of width r whose profile q has vanishing moments of all orders below an integer $\ell > \tau - d/2$; it moves the integral of a singularity of order $m < \tau - d/2$ by order r^{m+d}, while its discrepancy is only of order $r^{\tau+d/2}$. Here $d = 1$ and $\tau = 2$, and the profiles are $q_\ell = \psi^{(\ell)}$ with $\psi(w) = e^{-w^2/2}$, whose moments of all orders below ℓ vanish; the theorem needs $\ell \geq 2$. (a) The profiles, normalized. (b) The integrand of $I = \int \|\eta\|^{-2\tau} |\hat{q}(\eta)|^2 \, \mathrm{d}\eta$ for $q = q_\ell$: the vanishing moments make $\hat{q}_\ell$ small where the weight $\|\eta\|^{-2\tau}$ is large, so $I < \infty$ for $\ell \geq 2$, and the dilation $q \mapsto q(\cdot/r)$ gives the order $r^{\tau+d/2}$; for the dipole $I = \infty$, because $\tau > d/2 + 1$. (c) The discrepancy, by the exact Fourier sums of Appendix G.8; the slopes fitted over $r = 2^{-7}, \dots, 2^{-12}$ are 2.00, 2.50, 2.50 and 2.50 for $\ell = 1, \dots, 4$, which match $\min\{\tau + \frac{1}{2}, \ell + 1\}$.

$(C/r)D^* + \|b\|_\infty \left(2\lambda\Xi_r + (C/r)D^*\right) \leq C'(D^*/r + r)$, and $r = (D^*)^{1/2}$ gives the claim, since $D^* \leq 1$. □

In part (a) the dilation invariance of Hardy–Krause variation at a point is $\beta = 0$, and the logarithm is the sum over the dyadic shells of the order-0 singularity, each of which costs a constant. In part (b) a strip of width r costs r^{-1}, so $\beta = 1$ and $c = 1$; this recovers the classical order $(D^*)^{1/2}$ for half-planes in the unit square (Kuipers & Niederreiter, 1974; Niederreiter, 1992). For an axis-parallel segment, $n_1 n_2 = 0$, the mixed derivative of $b\,\mathbf{1}\{\ell > 0\}$ is a finite measure and the variation is finite: the jump costs nothing. Both orders are attained (Theorem D.12).

D.5 Sharpness

The lower bounds are the other half of the duality of Proposition D.2: signed measures concentrated at scale r next to the singular set. On spheres their discrepancy is the dilation computation of Lemma D.3 read in Fourier variables, and it is small exactly when enough of their moments vanish. In the square the dual object of Hardy–Krause variation is the discrepancy function, and the measure is designed through it.

Theorem D.11 (the law is sharp on spheres). *Let $d \geq 1$, $\tau > d/2$, $0 \leq m < \tau - d/2$ and $D = \mathrm{wce}_{H^\tau}$. Let $p \in \mathbb{S}^d$ and $g = a + s$ with $a \in H^\tau(\mathbb{S}^d)$ and $s \circ \mathrm{Exp}_p(z) = b(z)\,\|z\|^m\,\omega(z/\|z\|)$, supported in $B(p, \pi/4)$ and bounded at p when $m = 0$, where $b \in C_c^1(B(0,\pi/4))$ with $b(0) \neq 0$, $\omega \in C(\mathbb{S}^{d-1})$, and $w \mapsto \|w\|^m\,\omega(w/\|w\|)$ does not agree almost everywhere on $B(0,1)$ with a polynomial. Then for every $\gamma > (m+d)/(\tau + d/2)$ there are node sets Y_{N_k} of pairwise distinct nodes, none equal to p, with $D(Y_{N_k}) \to 0$ and $|(Q_{N_k} - \sigma)g| / D(Y_{N_k})^\gamma \to \infty$. At the borderline $m = \tau - d/2$, under the same hypotheses, there are such node sets with $|(Q_{N_k} - \sigma)g| \geq c\,D(Y_{N_k})\big(\log(1/D(Y_{N_k}))\big)^{1/2}$. In particular, for $h \equiv 1$ and $F = \frac{1}{2}d^2$, so that f is constant, and $x_0 = -p$:*

(i) for every $\gamma > d/(\tau + d/2)$ there are node sets with $D(Y_{N_k}) \to 0$ and $\|e_{N_k}(x_0)\| / D(Y_{N_k})^\gamma \to \infty$;

(ii) if $\tau > d/2 + 1$, then for every $\gamma > (d+1)/(\tau + d/2)$ there are node sets with $D(Y_{N_k}) \to 0$ and $|\Phi_{N_k}(x_0)| / D(Y_{N_k})^\gamma \to \infty$;

(iii) if $\tau = d/2 + 1$, there are node sets with $D(Y_{N_k}) \to 0$ and $|\Phi_{N_k}(x_0)| \geq c\, D(Y_{N_k})\big(\log(1/D(Y_{N_k}))\big)^{1/2}$.

Hence the exponent of Theorem D.6 cannot be improved for such integrands, nor its borderline factor, nor the field and value bounds of Corollary D.7; since D_{L_2} and $\mathrm{wce}_{H^{(d+1)/2}}$ are equivalent (Brauchart & Dick, 2013), no bound $\sup_x \|e_N(x)\| \leq C\, D_{L_2}(Y_N)^\gamma$ with $\gamma > 2d/(2d+1)$ holds for all node sets.

Proof. Fix an integer $\ell > \tau - d/2$.

A profile. On $C_c^\infty(B(0,1))$ consider the linear functionals $L_\alpha(q) = \int w^\alpha q(w)\,\mathrm{d}w$ for multi-indices $|\alpha| \leq \ell - 1$, and $\Lambda(q) = \int \|w\|^m \omega(w/\|w\|) q(w)\,\mathrm{d}w$. If Λ were a linear combination of the L_α, then $\|w\|^m \omega(w/\|w\|)$ would agree almost everywhere on $B(0,1)$ with a polynomial; so it is not, and since a linear functional that vanishes on the common kernel of finitely many functionals is a linear combination of them, there is a real $q \in C_c^\infty(B(0,1))$ with $L_\alpha(q) = 0$ for $|\alpha| \leq \ell - 1$ and $\Lambda(q) \neq 0$. The Taylor coefficients of $\hat{q}$ at the origin are multiples of the moments $L_\alpha(q)$, so $|\hat{q}(\eta)| \leq C\,\|\eta\|^\ell$, and $I = \int \|\eta\|^{-2\tau} |\hat{q}(\eta)|^2\,\mathrm{d}\eta < \infty$ because $2\ell - 2\tau > -d$ and $\hat{q}$ decays rapidly.

A multipole. Let $r \in (0, \frac{1}{2}]$. In normal coordinates z at p, σ has density $J(z) = c_d(\sin\|z\| / \|z\|)^{d-1} \geq c_d(\sin 1)^{d-1}$ on $B(0,1)$, with $c_d = 1/|\mathbb{S}^d|$. Let ϱ_r be the image under Exp_p of the signed measure $\epsilon\, q(z/r)\,\mathrm{d}z$, with $\epsilon = c_d(\sin 1)^{d-1}/(2\,\|q\|_\infty)$. Then $\nu_r = \sigma + \varrho_r$ is nonnegative, and it is a probability measure because $L_0(q) = 0$.

The discrepancy is of order $r^{\tau+d/2}$. Fix $\chi \in C_c^\infty(B(0,1))$ equal to 1 on $B(0, \frac{1}{2})$, which contains the support of $q(\cdot/r)$. For $a \in H^\tau(\mathbb{S}^d)$ and $\tilde{a} = a \circ \mathrm{Exp}_p$, Plancherel's identity and Lemma D.5 give

$$|\varrho_r a| = \epsilon \Big| \int_{\mathbb{R}^d} \chi\tilde{a}\, q(\cdot/r)\,\mathrm{d}z \Big| \leq \frac{\epsilon}{(2\pi)^d} \|\chi\tilde{a}\|_{H^\tau(\mathbb{R}^d)} \|q(\cdot/r)\|_{H^{-\tau}(\mathbb{R}^d)} \leq C\epsilon\, \|a\|_{H^\tau(\mathbb{S}^d)} \|q(\cdot/r)\|_{H^{-\tau}(\mathbb{R}^d)},$$

where $\|v\|^2_{H^{-\tau}(\mathbb{R}^d)} = \int (1+\|\xi\|^2)^{-\tau} |\hat{v}(\xi)|^2\,\mathrm{d}\xi$. Since $\widehat{q(\cdot/r)}(\xi) = r^d \hat{q}(r\xi)$ and $(1+\|\eta\|^2/r^2)^{-\tau} \leq r^{2\tau}\,\|\eta\|^{-2\tau}$,

$$\|q(\cdot/r)\|^2_{H^{-\tau}(\mathbb{R}^d)} = r^d \int \Big(1 + \frac{\|\eta\|^2}{r^2}\Big)^{-\tau} |\hat{q}(\eta)|^2\,\mathrm{d}\eta \leq r^{2\tau+d}\, I,$$

so $D(\nu_r) \leq C\epsilon I^{1/2}\, r^{\tau+d/2}$. The vanishing moments enter only here: they make $\hat{q}$ small where the weight $(1 + \|\xi\|^2)^{-\tau}$ is large.

The integral moves by order r^{m+d}. By the definition of the worst-case error, $|\varrho_r a| \leq \|a\|_{H^\tau}\, D(\nu_r) \leq C\,\|a\|_{H^\tau}\, r^{\tau+d/2}$. With $z = rw$,

$$\varrho_r s = \epsilon r^{m+d} \int b(rw)\,\|w\|^m\,\omega(w/\|w\|)\,q(w)\,\mathrm{d}w = \epsilon r^{m+d}\big(b(0)\Lambda(q) + R_r\big),$$

where $|R_r| \leq r\,\|\nabla b\|_\infty\,\|\omega\|_\infty\,\|q\|_{L^1}$ because $\|w\| \leq 1$ on the support of q. Since $m + d < \tau + d/2$, $|\varrho_r g| \geq \epsilon |b(0)\Lambda(q)| r^{m+d}/2$ for r small.

Point sets. Let $y_1, y_2, \ldots$ be i.i.d. with law ν_r and Q_n the empirical measure of the first n. Almost surely the points are distinct, none equals p, because ν_r has a density, and $Q_n \to \nu_r$ weakly. For $\tau > d/2$, $H^\tau(\mathbb{S}^d)$ has a continuous reproducing kernel K_τ, so $D(Q_n)^2 = \iint K_\tau\,\mathrm{d}(Q_n - \sigma)\,\mathrm{d}(Q_n - \sigma) \to D(\nu_r)^2$; and g is bounded and continuous off the ν_r-null set $\{p\}$, so $(Q_n - \sigma)g \to \varrho_r g$. Choose $n = n(r)$ with $D(Q_n) \leq 2C\epsilon I^{1/2} r^{\tau+d/2}$ and $|(Q_n - \sigma)g| \geq \epsilon |b(0)\Lambda(q)| r^{m+d}/4$. Along $r_i = 2^{-i}$ the ratio is at least $c\, r_i^{m+d-\gamma(\tau+d/2)} \to \infty$ for $\gamma > (m+d)/(\tau+d/2)$, while $D \to 0$.

The borderline. Let $m = \tau - d/2$, so that $m + d = \tau + d/2$, let q and $\ell > m$ be as above, $L \geq 1$, $\lambda_j = 2^{-j-1}$ for $0 \leq j \leq L$, and let ϱ be the image under Exp_p of $\sum_{j=0}^L \epsilon_j\, q(z/\lambda_j)\,\mathrm{d}z$ with $\epsilon_j = \epsilon\lambda_j^{-m-d}$. At a point with $\|z\| = \rho$ only the pieces with $\lambda_j > \rho$ are nonzero, and $\sum_j \epsilon_j \leq \epsilon\lambda_L^{-m-d}/(1-2^{-m-d})$, so $\epsilon = c_d(\sin 1)^{d-1}(1-2^{-m-d})\lambda_L^{m+d}/(2\,\|q\|_\infty)$ keeps $\nu = \sigma + \varrho$

nonnegative. Every scale moves the integral by the same amount: $\varrho s = \sum_j \epsilon_j \lambda_j^{m+d}(b(0)\Lambda(q) + R_{\lambda_j}) = \epsilon\big((L+1)b(0)\Lambda(q) + O(1)\big)$, since $\sum_j \lambda_j \le 1$. The scales are almost orthogonal in $H^{-\tau}$: for $\lambda_j \le \lambda_k$, since $|\hat q(\zeta)| \le C\,\|\zeta\|^{\ell}$,

$$\epsilon_j\epsilon_k\big|\big\langle q(\cdot/\lambda_j), q(\cdot/\lambda_k)\big\rangle_{H^{-\tau}}\big| \le \epsilon_j\epsilon_k\lambda_j^d\lambda_k^d \int \|\xi\|^{-2\tau}\,|\hat q(\lambda_j\xi)||\hat q(\lambda_k\xi)|\,\mathrm{d}\xi \le C\epsilon^2\Big(\frac{\lambda_j}{\lambda_k}\Big)^{\ell-m},$$

where $\int \|\eta\|^{\ell-2\tau}\,|\hat q(\eta)|\,\mathrm{d}\eta < \infty$ for the same reason as $I < \infty$. With the first step of the discrepancy bound, $D(\nu)^2 \le C\epsilon^2 \sum_{j,k} 2^{-(\ell-m)|j-k|} \le C\epsilon^2(L+1)$, while $|\varrho g| \ge |\varrho s| - \|a\|_{H^\tau}\, D(\nu) \ge c\,\epsilon(L+1)$ for L large. Since $t\,(\log(1/t))^{1/2}$ increases for $t < e^{-1/2}$ and $\log(1/(C\epsilon(L+1)^{1/2})) \le C'(L+1)$, this gives $|\varrho g| \ge c\,D(\nu)(\log(1/D(\nu)))^{1/2}$, and sampling realizes ν by node sets as before, with $D(Q_n) \le 2C\epsilon(L+1)^{1/2}$.

The field and the value. Let $\psi \in C^\infty([0,\infty);[0,1])$ equal 1 near 0 and 0 on $[\pi/5,\infty)$. Since f is constant, $\langle e_N(x_0), e_1\rangle = (Q_N - \sigma)g_1$ for $g_1 = -\big\langle \mathrm{Log}_{x_0}, e_1\big\rangle$, and by (30) $g_1 \circ \mathrm{Exp}_p(z) = -\pi\psi(\|z\|)\,\langle z/\|z\|, e_1\rangle + a_1$ with a_1 smooth, as in the proof of Corollary D.7: here $m = 0$, $b = -\pi\psi$ and $\|w\|^0\,\omega(w/\|w\|) = w_1/\|w\|$, which is not a polynomial; since $\|e_N(x_0)\| \ge |\langle e_N(x_0), e_1\rangle|$, this gives (i). For the value, $\Phi_N(x_0) = (Q_N - \sigma)g_2$ with $g_2 = F(x_0,\cdot)$, and $\frac12 d(x_0, \mathrm{Exp}_p z)^2 = \frac{\pi^2}{2} - \pi\|z\| + \frac12\|z\|^2$ for $\|z\| < \pi$, so $g_2 = -\pi\psi(\|z\|)\,\|z\| + a_2$ with a_2 smooth: here $m = 1$, which is below $\tau - d/2$ when $\tau > d/2+1$ and equal to it when $\tau = d/2+1$, and $\|w\|$ is not a polynomial; this gives (ii) and (iii). □

For the cap discrepancy, $\tau - d/2 = \frac12$, so a dipole suffices ($\ell = 1$), and its discrepancy needs no Fourier analysis: points of $B(p,r)$ are at chordal distance at most $2r$, so by (13) $\mathcal{D}(\varrho_r)^2 = -\frac{1}{C_d}\iint \|y-z\|\,\mathrm{d}\varrho_r\,\mathrm{d}\varrho_r \le \frac{2r}{C_d}|\varrho_r|(\mathbb{S}^d)^2 \le cr^{2d+1}$: the energy of a dipole is its width times its squared mass. For $\tau > d/2+1$ no dipole suffices. Since $H^\tau \subset C^1$, its first moment, of order r^{d+1}, pairs with the gradient of a test function at its centre, so its discrepancy is of order r^{d+1} rather than $r^{\tau+d/2}$ and it excludes only exponents above $d/(d+1)$; this is why higher moments must vanish: the proof takes $\ell > \tau - d/2$, and $\ell = \tau - d/2$ would cost only a factor $(\log(1/r))^{1/2}$. A linear bound cannot be improved either: for a fixed smooth signed density ϱ of mass zero with $\varrho F(x_0,\cdot) \ne 0$, the measures $\sigma + t\varrho$, $t > 0$ small, have value error and discrepancy both proportional to t, and sampling realizes them as above; so the exponent 1 of the value for $\tau < d/2+1$ is sharp as well. The multipoles sit at a scale unrelated to N, so the theorem says nothing about node sets of optimal order; for those the floor and the designs below are the relevant bounds. A quadrature evaluation confirms the dipole scalings $r^{d+1/2}$ and r^d for the cap discrepancy on $\mathbb{S}^2$ and $\mathbb{S}^3$, and an exact Fourier computation on $\mathbb{S}^1$ confirms that profiles with ℓ vanishing moments have discrepancy of order $r^{\min\{\tau+1/2,\,\ell+1\}}$ (Appendix G.8). Figure 7 shows such profiles, why their vanishing moments matter, and the exact scalings.

In the square. For a probability measure ν on $[0,1]^2$ with a density, the *discrepancy function* $\Delta_\nu(a) = \nu([0,a)) - a_1a_2$ gives $D^*(\nu) = \sup_a |\Delta_\nu(a)|$. It is the dual object of Hardy–Krause variation: for $\Delta \in C_c^2((0,1)^2)$ with $\partial_1\partial_2\Delta \ge -1$, the measure $\nu = (1 + \partial_1\partial_2\Delta)\lambda$ is a probability measure with $\Delta_\nu = \Delta$, because Δ vanishes near the axes and near $(1,1)$, and for s that is C^2 on a neighbourhood of the support of Δ, two integrations by parts give

$$(\nu - \lambda)s = \int s\,\partial_1\partial_2\Delta\,\mathrm{d}u = \int \partial_1\partial_2 s\,\Delta\,\mathrm{d}u, \tag{31}$$

the Hlawka–Zaremba identity for a compactly supported discrepancy function. Positivity allows $|\partial_1\partial_2\Delta|$ up to 1, so a discrepancy function of height ϵ can vary on the scale $\epsilon^{1/2}$: the scale $r = (D^*)^{1/2}$ of the proof of Theorem D.10, now forced from the other side. Figure 8 draws the two discrepancy functions of the next theorem.

Theorem D.12 (the law is sharp in the square)**.** *In the setting of Theorem D.10:*

(a) if $b(u_0) \ne 0$ and ω is not constant, there are $c > 0$ and point sets with $D^* \to 0$ and $|Qg - \lambda g| \ge c\,D^*\log(1/D^*)$;

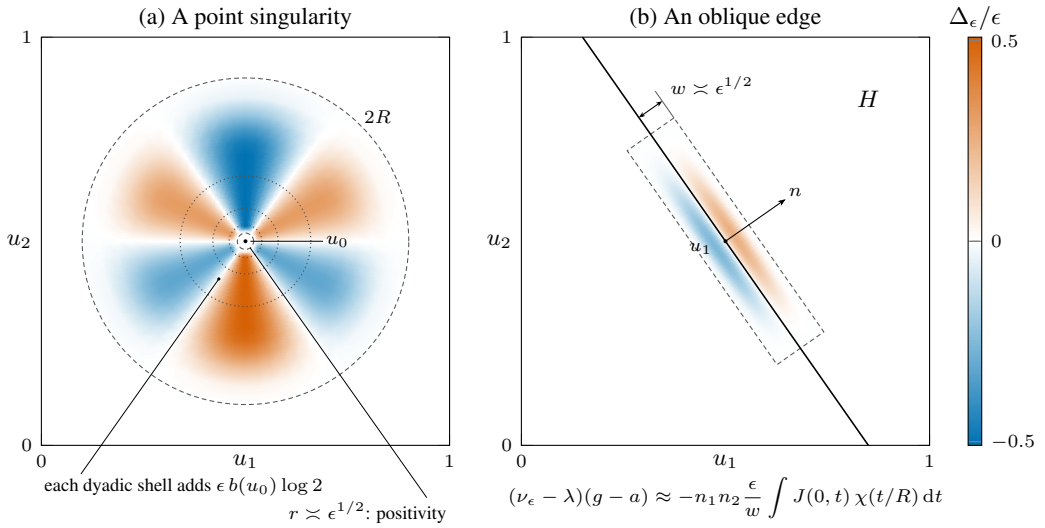

Figure 8: **Discrepancy functions are the lower bounds in the square.** A probability density $1 + \partial_1\partial_2\Delta$ with $\Delta \in C_c^2((0,1)^2)$ has discrepancy function Δ, so its star discrepancy is $\sup|\Delta|$, and by the Hlawka–Zaremba identity (31) it integrates s with error $\int \partial_1\partial_2 s\, \Delta$. Positivity allows $|\partial_1\partial_2\Delta|$ up to 1, so a function of height ϵ can vary only on the scale $\epsilon^{1/2}$. (a) For a point singularity of order 0, $s = b\cos\theta$, Δ_ϵ follows the angular profile of $\partial_1\partial_2 s \propto F(\theta)/\rho^2$, $F = \sin\theta\,(3\cos^2\theta - 1)$, on the annulus $r < \rho < 2R$, $r \asymp \epsilon^{1/2}$; each dyadic shell contributes $\epsilon\, b(u_0)\log 2$, and the error is of order $\epsilon\log(1/\epsilon)$, the logarithm of Theorem D.10(a). (b) For an oblique edge, Δ_ϵ lives within distance $w \asymp \epsilon^{1/2}$ of the edge, and the error is of order $\epsilon^{1/2}$, the square root of Theorem D.10(b). Both are the constructions of Theorem D.12, with $u_0 = u_1 = (\frac{1}{2}, \frac{1}{2})$, $r = 0.02$, $R = 0.2$ in (a) and $n = (\cos 35°, \sin 35°)$, $w = 0.07$, $R = 0.32$ in (b). Dashed: support boundaries; dotted: $\rho = 2r, 4r, 8r$.

(b) if on a disc around a point u_1 of the open square $g - a = c_0 + J\mathbf{1}_H$, with $c_0, J \in C^2$, $J(u_1) \neq 0$ and H a half-plane whose boundary passes through u_1 with unit normal n, $n_1 n_2 \neq 0$, there are $c > 0$ and point sets with $D^* \to 0$ and $|Qg - \lambda g| \geq c\,(D^*)^{1/2}$.

The logarithm in Theorem D.10(a) and the square root in (b) therefore cannot be improved.

Proof. In both parts we construct $\Delta_\epsilon \in C_c^\infty((0,1)^2)$ with $|\partial_1\partial_2\Delta_\epsilon| \leq \frac{1}{2}$, $\sup|\Delta_\epsilon| \leq C\epsilon$ and $|(\nu_\epsilon - \lambda)(g - a)|$ large, where $\nu_\epsilon = (1 + \partial_1\partial_2\Delta_\epsilon)\lambda$, and then realize ν_ϵ by point sets.

(a) Let $s = g - a = b\,\omega \circ \Theta$ with $\Theta = Z/\|Z\|$, let $\rho = \|u - u_0\|$ and θ be polar coordinates of $u - u_0$, $A = DZ(u_0)$ and $\Theta_A(u) = A(u - u_0)/\|A(u - u_0)\|$. The function $\omega \circ \Theta_A$ is homogeneous of degree 0 and, since A is invertible and ω is not constant, not constant; its mixed derivative is $F(\theta)\rho^{-2}$ with F continuous, and $F \not\equiv 0$. Indeed, on a half-plane avoiding the origin a C^2 function with vanishing mixed derivative is $f(u_1) + h(u_2)$; homogeneity of degree 0 and continuity make f and h constant, so $\omega \circ \Theta_A$ would be constant on each of the four open half-planes, hence everywhere. Since $\Theta - \Theta_A = O(\rho)$, $\partial_i(\Theta - \Theta_A) = O(1)$ and $\partial_1\partial_2(\Theta - \Theta_A) = O(\rho^{-1})$, because $Z - A(u - u_0)$ and its derivatives are $O(\rho^2)$, $O(\rho)$, $O(1)$ and $\|Z\| \geq \kappa_1\rho$, and since $\omega \in C^2$ and $b = b(u_0) + O(\rho)$, the chain rule gives

$$\partial_1\partial_2 s = b(u_0)F(\theta)\rho^{-2} + o(\rho^{-2}) \qquad \text{uniformly in } \theta \text{ as } \rho \to 0,$$

on $B(u_0, 4R) \subset U$ for some $R > 0$. Choose $G \in C^\infty(\mathbb{S}^1)$ with $\int_0^{2\pi} FG\,\mathrm{d}\theta = 1$ and $\beta_0 \in C^\infty([0,\infty);[0,1])$ vanishing on $[0,1]$ and equal to 1 on $[2,\infty)$, and put

$$\Delta_\epsilon = \epsilon\, G(\theta)\big(\beta_0(\rho/r) - \beta_0(\rho/R)\big), \qquad r = (2M\epsilon)^{1/2},$$

where M bounds $\|u\|^2\,|\partial_1\partial_2[G(\theta)\beta_0(\|u\|)]|$; by scaling, $|\partial_1\partial_2\Delta_\epsilon| \leq \epsilon M/r^2 = \frac{1}{2}$ once $r \leq R/2$. Then $\sup|\Delta_\epsilon| \leq \epsilon\,\|G\|_\infty$, and Δ_ϵ vanishes on $B(u_0, r)$, where s is singular. By (31) and Frullani's

integral $\int_0^\infty (\beta_0(\rho/r) - \beta_0(\rho/R))\,\rho^{-1}\,\mathrm{d}\rho = \log(R/r)$,

$$(\nu_\epsilon - \lambda)s = \epsilon\, b(u_0) \log \frac{R}{r} + o\Big(\epsilon \log \frac{1}{r}\Big) + O(\epsilon) = \frac{\epsilon}{2}\, b(u_0) \log \frac{1}{\epsilon} + o\Big(\epsilon \log \frac{1}{\epsilon}\Big).$$

(b) Let $\xi = n \cdot (u - u_1)$ and $t = n^\perp \cdot (u - u_1)$ with $n^\perp = (-n_2, n_1)$. After replacing (c_0, J) by $(c_0 + J, -J)$ if necessary, $H = \{\xi > 0\}$, and

$$\partial_1 \partial_2 = n_1 n_2\, \partial_\xi^2 + (n_1 n_2^\perp + n_1^\perp n_2)\, \partial_\xi \partial_t + n_1^\perp n_2^\perp\, \partial_t^2.$$

Choose $\varphi, \chi \in C_c^\infty((-1,1))$ with $\varphi'(0) = 1$, $\chi \geq 0$ and $\int \chi = 1$, let R be so small that $B(u_1, 2R)$ lies in the disc, and put $\Delta_\epsilon = \epsilon\, \varphi(\xi/w)\chi(t/R)$ with $w = (2M'\epsilon)^{1/2} \leq R$, where M' is chosen, by scaling, so that $|\partial_1 \partial_2 \Delta_\epsilon| \leq \epsilon M'/w^2 = \frac{1}{2}$. Then $\sup|\Delta_\epsilon| \leq \epsilon\, \|\varphi\|_\infty\, \|\chi\|_\infty$. Use (31) for c_0; on $\{\xi > 0\}$ integrate the ∂_t terms by parts in t and the ∂_ξ^2 term once in ξ. Since $|\partial_\xi \Delta_\epsilon| \leq C\epsilon/w$ on a set of area $O(wR)$, the terms $\int \partial_1 \partial_2 c_0\, \Delta_\epsilon$, $\int_{\xi>0} \partial_t J\, \partial_\xi \Delta_\epsilon$, $\int_{\xi>0} \partial_t^2 J\, \Delta_\epsilon$ and $\int_{\xi>0} \partial_\xi J\, \partial_\xi \Delta_\epsilon$ are $O(\epsilon R)$, and

$$\begin{aligned}(\nu_\epsilon - \lambda)(g - a) &= -n_1 n_2 \int J(0,t)\, \partial_\xi \Delta_\epsilon(0,t)\, \mathrm{d}t + O(\epsilon R) \\ &= -n_1 n_2 \frac{\epsilon}{w} \int J(0,t)\chi(t/R)\, \mathrm{d}t + O(\epsilon R).\end{aligned}$$

As $\int J(0,t)\chi(t/R)\,\mathrm{d}t = R\,(J(u_1) + O(R))$, for R small its modulus is at least $|n_1 n_2 J(u_1)| R\, \epsilon^{1/2}/(2(2M')^{1/2}) - O(\epsilon R)$.

Point sets. Let Q_n be the empirical measures of i.i.d. samples from ν_ϵ. By the Glivenko–Cantelli theorem for distribution functions on $\mathbb{R}^2$, $D^*(Q_n) \to D^*(\nu_\epsilon) = \sup|\Delta_\epsilon|$ almost surely; $g - a$ is bounded and continuous off a Lebesgue-null set, hence off a ν_ϵ-null set, so $(Q_n - \lambda)(g - a) \to (\nu_\epsilon - \lambda)(g - a)$; and $|(Q_n - \lambda)a| \leq V_{\mathrm{HK}}(a) D^*(Q_n)$ by (12). Choose n with $D^*(Q_n) \leq 2\sup|\Delta_\epsilon| \leq C\epsilon$ and $|(Q_n - \lambda)(g - a)|$ at least half its limit. In (a) this gives $|Q_n g - \lambda g| \geq \frac{\epsilon}{8}|b(u_0)| \log \frac{1}{\epsilon} \geq c\, D^*(Q_n) \log(1/D^*(Q_n))$ for ϵ small, because $t \log(1/t)$ increases near 0; in (b) it gives $|Q_n g - \lambda g| \geq c\, \epsilon^{1/2} \geq c' D^*(Q_n)^{1/2}$. Letting $\epsilon \to 0$ gives the point sets of the statement. □

The two lower bounds are the proof of Theorem D.10 read backwards. There each dyadic shell of the order-0 singularity costs a constant and the bump at scale $(D^*)^{1/2}$ closes the sum; here a discrepancy function of height ϵ that follows the angular profile of the singularity makes each shell between $\epsilon^{1/2}$ and 1 contribute ϵ, and positivity stops it at $\epsilon^{1/2}$. For an oblique edge a strip of width $\epsilon^{1/2}$ plays the role of the shells; for an axis-parallel edge, $n_1 n_2 = 0$, the leading term vanishes, as the finite variation of the jump requires. The field integrand of Proposition E.10(b) with $h(-x) \neq 0$ has the form of (a), with $b(u_0) = \pi h(-x)$ and ω a coordinate of the direction, and the field integrand of a flat torus with a non-rectangular lattice has, for suitable x, an oblique edge across which it jumps by a nonzero lattice vector (Proposition E.11); the logarithm of row 5 and the square root of row 6 of Table 1 are therefore sharp. A direct evaluation for $\omega = \cos$ and for a single oblique edge confirms the constants (Appendix G.8).

D.6 The floor and the designs

Proposition D.13 (the field of one node). *Let $h \in C(\mathbb{S}^d)$, $h \geq 0$, $Y_N = \{y_1, \dots, y_N\}$ a set of distinct nodes with weights $w_i > 0$, $f = \sigma F$ and $f_N^h = \sum_i w_i h(y_i) \frac{1}{2} d(\cdot, y_i)^2$. For $x \notin -Y_N$ let $e(x) = \operatorname{grad} f_N^h(x) - \operatorname{grad} f(x)$ be the field, and $R_i(x) = \sum_{j \neq i} w_j h(y_j) \operatorname{Log}_x(y_j) - \int h \operatorname{Log}_x \mathrm{d}\sigma$. Then*

$$\sup_{x \notin -Y_N} \|e(x)\| \;\geq\; \max_i \Big(\|R_i(-y_i)\| + \pi w_i h(y_i) \Big) \;\geq\; \pi \max_i w_i h(y_i),$$

which is $\frac{\pi}{N} \max_i h(y_i)$ for equal weights.

Proof. For $x \notin -Y_N$, $-e(x) = \sum_j w_j h(y_j) \operatorname{Log}_x(y_j) - \int h \operatorname{Log}_x \mathrm{d}\sigma = w_i h(y_i) \operatorname{Log}_x(y_i) + R_i(x)$, and only norms matter below. The remainder R_i is continuous at $x_0 = -y_i$: the other nodes differ

from y_i, so x_0 is not their antipode, and $x \mapsto \int h \operatorname{Log}_x \mathrm{d}\sigma$ is continuous by dominated convergence. For a unit vector $v \in T_{x_0}\mathbb{S}^d$ and $x_t = \cos t\, x_0 + \sin t\, v$, the point $y_i = -x_0$ lies at distance $\pi - t$ from x_t along the great circle through x_0 and v, so $\operatorname{Log}_{x_t}(y_i) = (\pi - t)(-\sin t\, x_0 + \cos t\, v) \to \pi v$ as $t \downarrow 0$, and $x_t \notin -Y_N$ for small $t > 0$. Hence $\|e(x_t)\| \to \|R_i(x_0) + \pi w_i h(y_i) v\|$, which for $v = R_i(x_0)/\|R_i(x_0)\|$ (any unit v if $R_i(x_0) = 0$) equals $\|R_i(x_0)\| + \pi w_i h(y_i)$. □

As $t \downarrow 0$ from all directions v, the limits of $-e(x_t)$ fill the circle of radius $\pi w_i h(y_i)$ about $R_i(x_0)$ (Figure 9).

Corollary D.14 (designs meet the floor up to N^ϵ). *Let $h \in C^\infty(\mathbb{S}^d)$, $h \geq 0$, and let $Q_N = \sum_i w_i \delta_{y_i}$ be positive rules with N distinct nodes and $\operatorname{wce}_{H^\tau}(Q_N) \leq c_\tau N^{-\tau/d}$ for some $\tau > d/2$; for equal weights, (Q_N) is a QMC design of order τ. Then, for N large,*

$$\pi \max_i w_i h(y_i) \;\leq\; \sup_{x \notin -\operatorname{supp} Q_N} \|e_N(x)\| \;\leq\; C\, N^{-\tau/(\tau + d/2)},$$

and $\sup_x |\Phi_N(x)| \leq C\, N^{-\min\{\tau, (d+1)\tau/(\tau+d/2)\}/d}$, times $(\log N)^{1/2}$ if $\tau = d/2 + 1$. For $d \geq 2$, positive rules that integrate every polynomial of degree at most t exactly, with $t \asymp N^{1/d}$, satisfy the hypothesis for every $\tau > d/2$; spherical t-designs with $t \asymp N^{1/d}$ are such rules with equal weights, and they exist (Bondarenko et al., 2013), see also Brauchart et al. (2014, Thm. 6). For these rules $\sup_x \|e_N(x)\| \leq C_\epsilon N^{-1+\epsilon}$ and $\sup_x |\Phi_N(x)| \leq C_\epsilon N^{-(d+1)/d+\epsilon}$ for every $\epsilon > 0$.

Proof. The lower bound is Proposition D.13. For the upper bounds apply Corollary D.7 with $D = \operatorname{wce}_{H^\tau}(Q_N) \leq c_\tau N^{-\tau/d} \leq 1$: $D^{d/(\tau+d/2)} \leq cN^{-\tau/(\tau+d/2)}$, and similarly for the value, with the factor of Corollary D.7 at $\tau = d/2 + 1$. For $d \geq 2$, a positive rule exact to degree t has $\operatorname{wce}_{H^\tau} \leq C_\tau t^{-\tau}$ for every $\tau > d/2$ (Brauchart & Hesse, 2007); for designs this is Brauchart et al. (2014, Thm. 1), and with $t \asymp N^{1/d}$ it makes (Q_N) a QMC design of every order, generic in the sense of Brauchart et al. (2014, Def. 5). Given ϵ, choose τ with $\tau/(\tau + d/2) \geq 1 - \epsilon$ and $(d+1)\tau/(d(\tau + d/2)) \geq (d+1)/d - \epsilon$. □

For node sets that maximize the sum of distances, QMC designs of order $(d+1)/2$ (Brauchart et al., 2014, Thms. 4, 16), the corollary gives the field bound $N^{-(d+1)/(2d+1)}$, which is $N^{-3/5}$ on $\mathbb{S}^2$, and Theorem D.11 shows that no bound through D_{L_2} alone gives more. The gap between $N^{-(d+1)/(2d+1)}$ and the floor N^{-1} is therefore a question about the order of equidistribution, not about the cap discrepancy.

E Geometry: where the field integrand is singular

By Proposition C.7, everything a certificate says about a minimizer passes through the field integrand G_x, and by Theorems D.6 and D.10 what it says is fixed by the order and the codimension of the singularities of G_x as the integration space sees them. This section locates them. Transports with integrable coordinate singularities cost nothing (Appendix E.1); Hadamard manifolds and small balls have no cut locus to meet (Appendices E.2 and E.3); on spheres with full support the whole singularity of the Fréchet loss is the chordal distance to the antipode (Appendix E.4); and a chart resolves it exactly when it blows up the antipode (Appendix E.5).

E.1 Transported integrands

Transports are rarely smooth up to the boundary of the cube: polar-type maps have coordinate singularities, inverse distribution functions blow up at 0 and 1. The following lemma shows that integrable singularities of the mixed derivatives are harmless.

Lemma E.1 (variation from integrable mixed derivatives). *Let $g\colon [0,1]^s \to \mathbb{R}$ be continuous on the closed cube. Suppose that for every nonempty $\alpha \subseteq \{1, \dots, s\}$ the restriction $g_\alpha := g(\cdot_\alpha, \mathbf{1}_{-\alpha})$ is*

$C^{|\alpha|}$ on the open face $(0,1)^\alpha$ and $\int_{(0,1)^\alpha} |\partial^\alpha g_\alpha| < \infty$. Then

$$V_{\mathrm{HK}}(g) \;\le\; \sum_{\emptyset \neq \alpha} \int_{(0,1)^\alpha} \left|\partial^\alpha g_\alpha(u_\alpha)\right| \mathrm{d}u_\alpha < \infty.$$

The same holds if each g_α is $C^{|\alpha|}$ on its open face except at finitely many points, and for E-valued g with V_{HK} as in Definition C.1 and $|\cdot|$ replaced by $\|\cdot\|_E$.

Proof. Fix a nonempty α and a partition of the closed face $[0,1]^\alpha$ into boxes R; we bound $\sum_R |\Delta(g_\alpha; R)|$.

Step 1 (boxes in the open face). If $R \subset (0,1)^\alpha$, the iterated fundamental theorem of calculus gives $\Delta(g_\alpha; R) = \int_R \partial^\alpha g_\alpha$, hence $|\Delta(g_\alpha; R)| \le \int_R |\partial^\alpha g_\alpha|$.

Step 2 (boxes touching the boundary). If R touches the boundary, let $R_\epsilon \subset R$ be the box obtained by moving each face of R that lies on the boundary inward by ϵ. Then $R_\epsilon \subset (0,1)^\alpha$, so Step 1 applies to it, and $\Delta(g_\alpha; R_\epsilon) \to \Delta(g_\alpha; R)$ as $\epsilon \to 0$ by continuity of g; therefore

$$|\Delta(g_\alpha; R)| \;=\; \lim_{\epsilon \to 0} |\Delta(g_\alpha; R_\epsilon)| \;\le\; \lim_{\epsilon \to 0} \int_{R_\epsilon} |\partial^\alpha g_\alpha| \;\le\; \int_R |\partial^\alpha g_\alpha|.$$

Step 2′ (exceptional points). If g_α fails to be $C^{|\alpha|}$ at finitely many points, split every box of the partition by the coordinate hyperplanes through the exceptional points it contains. Increments are additive under such splits, so $\sum_R |\Delta(g_\alpha; R)|$ does not decrease, and every new box meets the exceptional points only on its boundary; moving the faces through them inward, as in Step 2, bounds it by the integral over the box.

Step 3 (summation). Summing over the partition, $\sum_R |\Delta(g_\alpha; R)| \le \int_{(0,1)^\alpha} |\partial^\alpha g_\alpha|$; the supremum over partitions bounds the Vitali variation of g_α, and summing over α gives the claim. For E-valued g, apply the scalar statement to $\langle g, v\rangle$ with $\|v\| = 1$ and take the supremum over v. □

We make this concrete for the maps that parametrize the sphere $\mathbb{S}^2$ by the square: the cylindrical maps

$$T_z(u_1, u_2) = \big(r(u_1)\cos 2\pi u_2,\; r(u_1)\sin 2\pi u_2,\; z(u_1)\big), \qquad r = \sqrt{1 - z^2}, \tag{32}$$

where $z\colon [0,1] \to [z_-, z_+] \subseteq [-1,1]$ is a C^2 increasing bijection with $z' > 0$ on $[0,1]$. By Archimedes' theorem the x_3-marginal of σ_2 is uniform on $[-1,1]$, so $(T_z)_\# \lambda$ has density $h(y) = 2/z'(z^{-1}(y_3))$ with respect to σ_2, supported on the zone $\{z_- \le y_3 \le z_+\}$. Two cases matter below: $z(t) = 2t - 1$ is the Lambert cylindrical equal-area map, $(T_z)_\# \lambda = \sigma_2$; and $z(t) = 1 + \varkappa^{-1} \log\big(t + (1-t)e^{-2\varkappa}\big)$ pushes λ to the von Mises–Fisher law (Mardia & Jupp, 2000) with concentration $\varkappa$ about the north pole, $h(y) = \varkappa e^{\varkappa y_3} / \sinh \varkappa$ (we reserve κ for curvature bounds). Uniform measures on caps $\{y_3 \ge \cos r_0\}$ arise from $z(t) = \cos r_0 + t(1 - \cos r_0)$. (The orientation matters: the Hardy–Krause faces are anchored at $u = \mathbf{1}$, and the bound below contains a term proportional to the radius r_+ of the circle of latitude $\{y_3 = z(1)\}$, which vanishes when $z(1) = 1$ is the pole.)

For a C^2 map G on a closed zone $Z \subseteq \mathbb{S}^2$ with values in a Euclidean space E, let $\widetilde{G}(y) = G(y/\|y\|)$ be its radial extension and $\|G\|_{C^2(Z)} := \max\{\big\|\widetilde{G}\big\|_\infty, \big\|D\widetilde{G}\big\|_\infty, \big\|D^2\widetilde{G}\big\|_\infty\}$, the suprema taken over the cone over Z intersected with $\{\frac{1}{2} \le \|y\| \le 2\}$.

Proposition E.2 (variation of cylindrically transported integrands)**.** *Let T_z be as in (32) and let $G\colon Z \to E$ be C^2 on the zone $Z = T_z([0,1]^2)$. Then*

$$V_{\mathrm{HK}}(G \circ T_z) \;\le\; \big(2\pi(z_+ - z_-) + (2\pi + 1)\theta_z + 2\pi r_+\big)\, \|G\|_{C^2(Z)}\,,$$

where $\theta_z := \int_{z_-}^{z_+} (1 - \zeta^2)^{-1/2}\, \mathrm{d}\zeta \;\le\; \pi$ and $r_+ := \sqrt{1 - z_+^2}$. In particular $V_{\mathrm{HK}}(G \circ T_z) \le (7\pi + 2\pi^2)\, \|G\|_{C^2(Z)}$ for every increasing C^2 profile.

Proof. Write $g = G \circ T_z \colon [0,1]^2 \to E$, $\varphi = 2\pi u_2$, $r = r(u_1) = \sqrt{1 - z(u_1)^2}$, and $\widetilde{G}$ for the radial extension of G, so that $g(u) = \widetilde{G}(T_z(u))$ and the ambient chain rule applies. We check the hypotheses of Lemma E.1 in four steps and bound the three integrals that it produces.

Step 1 (continuity and smoothness). T_z is continuous on the closed square (with $r(u_1) = 0$ where $z(u_1) = \pm 1$), and G is continuous on Z, so g is continuous on the closed square. On the open square $z \in (z_-, z_+)$; if $z_- = -1$ or $z_+ = 1$ the corresponding endpoint is excluded, so $r > 0$ and T_z is C^2 there, hence so is g.

Step 2 (derivatives of the transport). With $e_\varphi = (-\sin\varphi, \cos\varphi, 0)$ and $e_r = (\cos\varphi, \sin\varphi, 0)$,

$$\partial_{u_2} T_z = 2\pi r\, e_\varphi, \qquad \partial_{u_1} T_z = r'\, e_r + z'\, e_3, \quad r' = -\frac{z z'}{r}, \qquad \partial_{u_1}\partial_{u_2} T_z = 2\pi r'\, e_\varphi.$$

Hence $\|\partial_{u_2} T_z\| = 2\pi r \le 2\pi$, $\|\partial_{u_1} T_z\|^2 = r'^2 + z'^2 = z'^2\big(\frac{z^2}{1-z^2} + 1\big) = \frac{z'^2}{r^2}$, so $\|\partial_{u_1} T_z\| = z'/r$, and $\|\partial_{u_1}\partial_{u_2} T_z\| = 2\pi |z| z'/r \le 2\pi z'/r$.

Step 3 (the mixed derivative). By the chain rule,

$$\partial_{u_1}\partial_{u_2} g \;=\; D^2\widetilde{G}(T_z)[\partial_{u_1} T_z, \partial_{u_2} T_z] + D\widetilde{G}(T_z)[\partial_{u_1}\partial_{u_2} T_z],$$

so

$$\|\partial_{u_1}\partial_{u_2} g\| \le \|G\|_{C^2(Z)} \left(\frac{z'}{r} \cdot 2\pi r + 2\pi \frac{z'}{r}\right) = 2\pi \|G\|_{C^2(Z)} \left(z' + \frac{z'}{r}\right).$$

Integrating over the open square, using $\int_0^1 z'(u_1)\,\mathrm{d}u_1 = z_+ - z_-$ for the first term and the substitution $\zeta = z(u_1)$ for the second,

$$\begin{aligned} \int_{(0,1)^2} \|\partial_{u_1}\partial_{u_2} g\|\,\mathrm{d}u \;&\le\; 2\pi \|G\|_{C^2(Z)} \left(z_+ - z_- + \int_{z_-}^{z_+} \frac{\mathrm{d}\zeta}{\sqrt{1-\zeta^2}}\right) \\ &=\; 2\pi \|G\|_{C^2(Z)} \left(z_+ - z_- + \theta_z\right), \end{aligned}$$

where θ_z is as in the statement.

Step 4 (the faces). On the face $u_2 = 1$, $g(u_1, 1) = \widetilde{G}(T_z(u_1, 1))$ and $\|\partial_{u_1} g(u_1, 1)\| \le \|G\|_{C^2(Z)}\, z'/r$, whose integral over $(0,1)$ is at most $\|G\|_{C^2(Z)}\, \theta_z$. On the face $u_1 = 1$, $g(1, u_2) = G(T_z(1, u_2))$, where $u_2 \mapsto T_z(1, u_2)$ traces the circle of latitude $\{y_3 = z_+\}$ of radius $r_+ = \sqrt{1 - z_+^2}$ (a single point if $z_+ = 1$); hence $\|\partial_{u_2} g(1, u_2)\| \le 2\pi r_+ \|G\|_{C^2(Z)}$ and the one-dimensional variation of this face is at most $2\pi r_+ \|G\|_{C^2(Z)}$. Summing the three contributions,

$$V_{\mathrm{HK}}(g) \le \|G\|_{C^2(Z)} \left(2\pi(z_+ - z_-) + 2\pi\theta_z + \theta_z + 2\pi r_+\right),$$

which is the statement; the final bound uses $z_+ - z_- \le 2$ and $\theta_z \le \pi$.

For a decreasing profile the same bound holds with $|z(1) - z(0)|$ in place of $z_+ - z_-$ and $r_+ = \sqrt{1 - z(1)^2}$. □

Thus a smooth integrand on the sphere, transported by any cylindrical map, has Hardy–Krause variation bounded by its C^2 norm, with a constant that depends only on the endpoints of the profile. The reason is an invariance, not only a cancellation in the derivative computation. If Φ is a coordinatewise increasing homeomorphism of the cube that fixes the corner $\mathbf{1}$, then $V_{\mathrm{HK}}(g \circ \Phi) = V_{\mathrm{HK}}(g)$: partitions into boxes correspond under Φ, the alternating vertex sums agree, and the faces anchored at $\mathbf{1}$ are preserved. With z_0 the affine profile with the same endpoints, $T_z = T_{z_0} \circ \Phi$ for $\Phi(u_1, u_2) = ((z(u_1) - z_-)/(z_+ - z_-), u_2)$, hence $V_{\mathrm{HK}}(G \circ T_z) = V_{\mathrm{HK}}(G \circ T_{z_0})$ exactly, and the speed of the latitude parametrization cannot enter. (The mixed-smoothness norms behind the randomized bounds (S3)–(S4) involve derivatives of z and do depend on the profile, as do the error distributions of scrambled nets.) On a general manifold the same conclusion holds for any transport T that is C^s on the closed cube (then $V_{\mathrm{HK}}(G \circ T) \le \sum_\alpha \|\partial^\alpha (G \circ T)\|_\infty$ by Lemma C.2(iv)) and, by Lemma E.1, for transports with integrable coordinate singularities. Transports built from inverse Gaussian distribution functions are unbounded and fall under Owen's boundary growth condition instead, see Remark E.5. The choice of transport is not innocuous. Basu & Owen (2016) show

that the symmetric map $u \mapsto \Phi^{-1}(u)/\left\|\Phi^{-1}(u)\right\|$ from the cube onto the sphere does not preserve bounded Hardy–Krause variation, while the spherical-coordinate maps of Fang and Wang that they analyse do; on $\mathbb{S}^2$ the latter is the Lambert equal-area map $z(t) = 2t - 1$ of (32), up to the orientation of t and the order of the coordinates. Proposition E.2 extends this positive case to arbitrary increasing C^2 latitude profiles, which transport caps and von Mises–Fisher laws, with an explicit constant, a C^2 hypothesis on the integrand, and the endpoint-only dependence just explained.

E.2 No cut locus: Hadamard manifolds

Let ν be a probability measure on $\mathcal{M}$ with $\int d(x,y)^2 \,\mathrm{d}\nu(y) < \infty$. The Fréchet functional and its sample version are

$$f(x) = \frac{1}{2}\int_{\mathcal{M}} d(x,y)^2 \,\mathrm{d}\nu(y), \qquad f_N(x) = \frac{1}{2N}\sum_{i=1}^{N} d(x,y_i)^2, \qquad y_i = T(u_i), \tag{33}$$

and, wherever Log_x is smooth on the support, $\operatorname{grad} f(x) = -\int \mathrm{Log}_x(y)\,\mathrm{d}\nu(y)$, $\operatorname{grad} f_N(x) = -\frac{1}{N}\sum_i \mathrm{Log}_x(y_i)$. Minimizers are Fréchet (Karcher) means (Karcher, 1977; Afsari, 2011; Pennec, 2006). The gradient integrand is $G_x(u) = \mathrm{Log}_x(T(u))$.

Theorem E.3 (QMC Fréchet means on Hadamard manifolds)**.** *Let $\mathcal{M}$ be a Hadamard manifold with sectional curvatures in $[-\kappa, 0]$, let $T\colon [0,1]^s \to \mathcal{M}$ be continuous with $T([0,1]^s) \subseteq \bar{B}(p, r_0)$ for some $p \in \mathcal{M}$, $r_0 > 0$, and let $\nu = T_{\#}\lambda$. Then:*

(i) f and f_N are 1-strongly geodesically convex on $\mathcal{M}$; their minimizers $x^\star$ and $\hat{x}_N$ are unique and lie in $\bar{B}(p, r_0)$;

(ii) on $\bar{B}(p, r_0)$, $\operatorname{Hess} f$ and $\operatorname{Hess} f_N$ are bounded by $L_{r_0} = 2r_0\sqrt{\kappa}\coth(2r_0\sqrt{\kappa})$ (with $L_{r_0} = 1$ if $\kappa = 0$), and the projected gradient iteration $x \leftarrow \Pi_{\bar{B}(p,r_0)}\big(\operatorname{Exp}_x(-\eta \operatorname{grad} f_N(x))\big)$ with $0 < \eta \le 1/(\zeta(\kappa, 2r_0)L_{r_0}^2)$ converges linearly to $\hat{x}_N$ from every starting point in the ball;

(iii) with $V_1 := \sup_{x \in \bar{B}(p,r_0)} V_{\mathrm{HK}}(G_x)$,

$$d(\hat{x}_N, x^\star) \;\le\; V_1\, D_N^*(P_N); \tag{34}$$

(iv) $V_1 < \infty$ whenever T is C^s on the closed cube, and more generally whenever G_x satisfies the hypotheses of Lemma E.1 with bounds uniform in $x \in \bar{B}(p, r_0)$.

Proof. (i) Strong convexity and location of the minimizers. On a Hadamard manifold $\frac{1}{2}d(\cdot, y)^2$ is smooth, and (10) with $\Delta = 0$ gives $\operatorname{Hess} \frac{1}{2}d(\cdot, y)^2 \succeq I$. Averaging over y (Fubini's theorem and dominated convergence on compact sets) yields

$$\operatorname{Hess} f \;\succeq\; I \qquad \text{and} \qquad \operatorname{Hess} f_N \;\succeq\; I \qquad \text{on } \mathcal{M},$$

that is, 1-strong geodesic convexity; strongly convex continuous functions on a complete Hadamard manifold have unique minimizers. Let Π be the projection onto the closed convex ball $\bar{B}(p, r_0)$. It is nonexpansive and fixes the ball, so $d(\Pi x, y) \le d(x, y)$ for every y in the ball, and therefore $f(\Pi x) \le f(x)$ for every $x \in \mathcal{M}$. By uniqueness, $\Pi x^\star = x^\star$, that is, $x^\star \in \bar{B}(p, r_0)$; the same argument applies to $\hat{x}_N$.

(ii) Smoothness on the ball. For $x, y \in \bar{B}(p, r_0)$ we have $d(x, y) \le 2r_0$, and (10) bounds the Hessian of $\frac{1}{2}d(\cdot, y)^2$ at x by $\sqrt{\kappa}\,\ell\coth(\sqrt{\kappa}\,\ell) \le L_{r_0}$ with $\ell = d(x, y)$, the function $\ell \mapsto \sqrt{\kappa}\ell\coth(\sqrt{\kappa}\ell)$ being nondecreasing; averaging over y gives the bounds for f and f_N. For a concrete linear-rate method that stays in the ball, apply Theorem C.18 to f_N with exact gradients, $\sigma_N = 0$, $\mu = 1$, and projection onto the ball, using $\eta \le 1/(\zeta(\kappa, 2r_0)L_{r_0}^2)$. This supplies the claimed linear convergence without assuming the unit-step iteration remains in the ball.

(iii) The rate. Since Log is smooth on $\mathcal{M} \times \mathcal{M}$ and $T([0,1]^s)$ is compact, Assumption C.1 holds for $\widetilde{F}(x, u) = \frac{1}{2}d(x, T(u))^2$, with $\operatorname{grad}_x \widetilde{F}(x, u) = -G_x(u)$. As $\hat{x}_N$ is a global minimizer of f_N, $\operatorname{grad} f_N(\hat{x}_N) = 0$. Theorem C.12(b) with $C = U = \mathcal{M}$ and $\mu = 1$, applied at $\bar{x} = \hat{x}_N \in \bar{B}(p, r_0)$, gives

$$d(\hat{x}_N, x^\star) \;\le\; V_{\mathrm{HK}}(G_{\hat{x}_N})\, D_N^*(P_N) \;\le\; V_1\, D_N^*(P_N),$$

which is (34); note that part (b) of that theorem needs only the variation at $\bar{x}$, not a bound over all of $\mathcal{M}$. Alternatively, since f_N is 1-strongly convex by (i), Theorem C.13(a) with $C = \mathcal{M}$ and $\mu_0 = 1$ gives $d(\hat{x}_N, x^\star) \le V_{\mathrm{HK}}(G_{x^\star}) D_N^*(P_N)$, with the variation at the population mean only.

(iv) Finiteness of V_1. If $T \in C^s([0,1]^s; \mathcal{M})$, then $(x,u) \mapsto \partial_u^\alpha G_x(u)$ is continuous on the compact set $\bar{B}(p, r_0) \times [0,1]^s$ for every α, so Lemma C.2(iv) bounds $V_{\mathrm{HK}}(G_x)$ uniformly in x; under the hypotheses of Lemma E.1 the same holds with the integrals of the mixed derivatives in place of their suprema. □

Example E.4 (the affine-invariant cone of positive definite matrices)**.** Let $\mathcal{M} = \mathcal{P}_n$ be the cone of symmetric positive definite $n \times n$ matrices with the affine-invariant metric $\langle A, B\rangle_P = \mathrm{tr}(P^{-1}AP^{-1}B)$ (Pennec et al., 2006; Bhatia, 2007). It is a Hadamard manifold with $\mathrm{Exp}_P(V) = P^{1/2}\exp(P^{-1/2}VP^{-1/2})P^{1/2}$, $\mathrm{Log}_P(Q) = P^{1/2}\log(P^{-1/2}QP^{-1/2})P^{1/2}$ and $d(P,Q) = \left\|\log(P^{-1/2}QP^{-1/2})\right\|_F$; its sectional curvatures are nonpositive and bounded below by $-\frac{1}{2}$ (Thanwerdas & Pennec, 2022), so $\kappa = \frac{1}{2}$. For a C^s map $\Psi\colon [0,1]^s \to \mathrm{Sym}(n)$ the transport $T(u) = P^{1/2}\exp(\Psi(u))P^{1/2}$ is C^s into $\bar{B}(P, \sup\|\Psi\|_F)$ and Theorem E.3 applies: the sample Karcher mean of N Halton or Sobol' points converges to the population Karcher mean at rate $\widetilde{O}(N^{-1})$. The Karcher mean of the sample is computed by the classical fixed-point iteration $x \leftarrow \mathrm{Exp}_x\left(\frac{1}{N}\sum_i \mathrm{Log}_x(y_i)\right)$, which is gradient descent with unit step on f_N; the linear-rate guarantee of Theorem E.3(ii) is for the projected iteration with the conservative step $\eta \le 1/(\zeta(\kappa, 2r_0)L_{r_0}^2)$ inherited from Theorem C.18; the unit step, which is the one used in practice and in Appendix G, is not covered by that guarantee but converged in every run there. For the bounded model (B) of Appendix G.3, the integrand $u \mapsto \mathrm{Log}_{x^\star}(T(u))$ is C^∞ on the closed cube, so Theorem C.13(b) with (S3) and Lemma B.4 gives $\mathbb{E}\, d(\hat{x}_N, x^\star)^2 = O(N^{-3}(\log N)^5)$ for the scrambled Sobol' nets of the implementation, one net per replicate. The recorded distance slope -1.64 over $N \le 2^{14}$ is a finite-range observation consistent with this root-mean-square order $N^{-3/2}$ up to the logarithm.

Remark E.5 (Gaussian transports). The log-Gaussian model $\Psi(u) = \tau\sum_j \Phi^{-1}(u_j)E_j$ (an orthonormal basis E_j of $\mathrm{Sym}(n)$, Φ the standard normal distribution function) is the natural "Gaussian on the tangent space" model and is used in Appendix G. It has unbounded support, so Theorem E.3 does not apply. We have $\|G_x(u)\| \le d(x,P) + \tau\left\|\Phi^{-1}(u)\right\|$. Suppose in addition that for every $\epsilon > 0$ the mixed derivatives satisfy $\|\partial^\alpha G_x(u)\| \le B_\epsilon \prod_j \min(u_j, 1-u_j)^{-\epsilon - \mathbf{1}\{j\in\alpha\}}$, uniformly for x in bounded sets. This is the bound suggested by the chain rule: $|(\Phi^{-1})'(u_j)| \le C\min(u_j, 1-u_j)^{-1}$, and derivatives of the matrix exponential and logarithm grow at most exponentially in $\|\Psi(u)\| \le \tau\sum_j |\Phi^{-1}(u_j)|$, which is $o(\min_j \min(u_j, 1-u_j)^{-\epsilon})$; we state it as a hypothesis rather than a lemma. Under it, G_x satisfies Owen's boundary growth condition with every $A_j > 0$, and the expected error of scrambled nets or randomized Halton points is $O(N^{-1+\epsilon})$ for every $\epsilon > 0$ (Owen, 2006a), uniformly in x on bounded sets. Under the same hypothesis, $\mathbb{E}\,\|\widehat{g}_N(x) - \mathrm{grad}\, f(x)\| = O(N^{-1+\epsilon})$ at each fixed x, in particular at $x^\star$, and hence, by Theorem C.13 (each sample objective on a Hadamard manifold is 1-strongly convex, so $d(\hat{x}_N, x^\star) \le \|\mathrm{grad}\, f_N(x^\star)\|$ for every draw), $\mathbb{E}\, d(\hat{x}_N, x^\star) = O(N^{-1+\epsilon})$ for one fixed scrambled net or randomized Halton set. No uniform-in-x statement is needed for this conclusion; the uniform bound (34) would require an expected supremum over x, which the boundary-growth theory does not supply and which we do not claim. The measured ratio of Karcher-mean error to gradient discrepancy at $x^\star$ in Appendix G.3 (between 0.95 and 1.0) is consistent with (24).

E.3 No cut locus: the Karcher ball

On a manifold with positive curvature, $y \mapsto \frac{1}{2}d(x,y)^2$ is smooth only away from the cut locus $\mathrm{Cut}(x)$, and support in a sufficiently small ball guarantees uniqueness of the Fréchet mean (Karcher, 1977; Afsari, 2011). The same sufficient radius that yields uniqueness and strong convexity keeps the field integrand away from the cut locus, and the readings of Appendix C apply inside it.

Theorem E.6 (QMC Fréchet means inside the Karcher radius)**.** *Let $\mathcal{M}$ be complete with sectional curvatures in $[-\kappa, \Delta]$, $\Delta > 0$, and positive injectivity radius, and let*

$$r_K := \tfrac{1}{2}\min\{\mathrm{inj}(\mathcal{M}),\ \pi/(2\sqrt{\Delta})\}.$$

Let $0 < r_0 < r < r_K$, $p \in \mathcal{M}$, $C = \bar{B}(p, r)$, and let $T\colon [0,1]^s \to \bar{B}(p, r_0)$ be continuous, $\nu = T_\#\lambda$. Then:

(i) C is geodesically convex, and for $x \in C$, $y \in \bar{B}(p, r_0)$ the map $(x, y) \mapsto \mathrm{Log}_x(y)$ is smooth;
(ii) f and f_N are μ_r-strongly geodesically convex on C, $\mu_r := 2r\sqrt{\Delta}\cot(2r\sqrt{\Delta}) > 0$, and have Hessians bounded by $L_r = 2r\sqrt{\kappa}\coth(2r\sqrt{\kappa})$ on C;
(iii) the Fréchet means $x^\star$ of ν and $\hat{x}_N$ of $\frac{1}{N}\sum_i \delta_{y_i}$ exist, are unique, and lie in $\bar{B}(p, r_0) \subset \operatorname{int} C$;
(iv) with $V_1 := \sup_{x \in C} V_{\mathrm{HK}}(G_x)$,

$$d(\hat{x}_N, x^\star) \;\le\; \frac{V_1}{\mu_r}\, D_N^*(P_N), \tag{35}$$

and $V_1 < \infty$ under the conditions of Theorem E.3(iv).

On $\mathbb{S}^d$ (with $\Delta = 1$, $\mathrm{inj} = \pi$) one has $r_K = \pi/4$; on $\mathbb{S}^2$, for $T = T_z$ the increasing cylindrical map (32) with $z(t) = \cos r_0 + t(1 - \cos r_0)$ onto the cap $\{y_3 \ge \cos r_0\}$, $r_0 < \pi/4$, Proposition E.2 (with $z_+ = 1$, $z_- = \cos r_0$, $r_+ = 0$, $\theta_z = r_0$) gives $V_1 \le (2\pi(1 - \cos r_0) + (2\pi + 1)r_0)\sup_{x \in C} \|\mathrm{Log}_x\|_{C^2(\mathrm{cap})} < \infty$.

Proof. (i) Convexity and smoothness. Since $r < \frac{1}{2}\min\{\mathrm{inj}(\mathcal{M}), \pi/\sqrt{\Delta}\}$, the ball C is strongly convex (Chavel, 2006, Thm. IX.6.1); see also Afsari (2011, eq. (1.2)). For $x \in C$ and $y \in \bar{B}(p, r_0)$,

$$d(x, y) \;\le\; r + r_0 \;<\; 2r_K \;\le\; \mathrm{inj}(\mathcal{M}),$$

so $y \notin \mathrm{Cut}(x)$ and $\mathrm{Log}_x(y)$ is smooth in (x, y).

(ii) Hessian bounds. For such x, y put $\ell = d(x, y) < 2r < \pi/(2\sqrt{\Delta})$. By (10) the eigenvalues of $\mathrm{Hess}_x \frac{1}{2}d(\cdot, y)^2$ lie between $\sqrt{\Delta}\ell\cot(\sqrt{\Delta}\ell)$ and $\sqrt{\kappa}\ell\coth(\sqrt{\kappa}\ell)$; since $t \mapsto t\cot t$ is decreasing on $(0, \pi/2)$ and $t \mapsto t\coth t$ is increasing,

$$\mu_r\, I \;\preceq\; \mathrm{Hess}_x \tfrac{1}{2}d(\cdot, y)^2 \;\preceq\; L_r\, I \qquad (x \in C,\; y \in \bar{B}(p, r_0)).$$

Integrating over y gives the same bounds for f and f_N on C, and integrating the lower bound along the geodesics of the convex set C gives μ_r-strong geodesic convexity.

(iii) Existence, uniqueness and location of the means. For a measure supported in $B(p, r')$ with $r' < \frac{1}{2}\min\{\mathrm{inj}(\mathcal{M}), \pi/\sqrt{\Delta}\}$, the Fréchet mean exists, is unique, and lies in $B(p, r')$ (Afsari, 2011, Thm. 2.1). Apply this with $r' = r$ to ν and to the empirical measure, both supported in $\bar{B}(p, r_0) \subset B(p, r)$; applying it once more with centre p and every radius $r_0' \in (r_0, r)$ and letting $r_0' \downarrow r_0$ shows $x^\star, \hat{x}_N \in \bar{B}(p, r_0) \subset \operatorname{int} C$.

(iv) The rate. By (iii), $\hat{x}_N$ is a global minimizer of f_N lying in the interior of C, hence $\mathrm{grad}\, f_N(\hat{x}_N) = 0$. Choose $r' \in (r, r_K)$ and $U := B(p, r')$; for $x \in U$ and $y \in \bar{B}(p, r_0)$, $d(x, y) < r' + r_0 < 2r_K \le \mathrm{inj}(\mathcal{M})$, so by (i) Assumption C.1 holds on U. Theorem C.12(b) with this U, $C = \bar{B}(p, r)$ and $\mu = \mu_r$ gives (35). The bound on V_1 for the sphere follows from Proposition E.2 applied to $G = \mathrm{Log}_x$, which is C^∞ on the cap for $x \in C$ because $-x$ is not in the cap. □

Remark E.7 (three thresholds in one radius). The Karcher radius of Theorem E.6 bundles three conditions with different thresholds: strong convexity of the Fréchet functional, through the Hessian comparison, which needs $2r\sqrt{\Delta} < \pi/2$; avoidance of the cut locus, which needs only $r + r_0 < \mathrm{inj}(\mathcal{M})$ and keeps $\mathrm{Log}_x \circ T$ continuous; and bounded variation of the transported field integrand, which needs in addition an integrable mixed derivative. It is not necessary for bounded variation. On $\mathbb{S}^2$, a cap of radius $r_0 = \pi/3 > \pi/4$ and a decision ball of radius $r = 5\pi/12$ satisfy $r + r_0 = 3\pi/4 < \pi$, so Log_x is smooth on the cap for every x in the ball and Proposition E.2 gives uniformly bounded variation for the cylindrical cap transport, while the Hessian lower bound $2r\cot(2r)$ is negative and the convexity argument no longer applies. Bounded variation fails only when the support meets the cut locus of a relevant decision point, and then the order and codimension of the singularity decide (Appendix D and Appendix E.5).

E.4 The sphere is electrostatics

With a full-support law on $\mathbb{S}^d$ the cut point $-x$ of every decision point carries data. Equation (30) shows the singularity of the field integrand; the next proposition shows where it comes from.

Proposition E.8 (Sobolev regularity of the Fréchet integrand on $\mathbb{S}^2$)**.** *For every $x \in \mathbb{S}^2$, the function $\varphi_x = \frac{1}{2}d(x,\cdot)^2$ belongs to $H^\tau(\mathbb{S}^2)$ for every $\tau < 2$, and each component of $y \mapsto \mathrm{Log}_x(y)$ (in a fixed basis of $T_x\mathbb{S}^2$) belongs to $H^\tau(\mathbb{S}^2)$ for every $\tau < 1$. Both are C^∞ on $\mathbb{S}^2 \setminus \{-x\}$. For every unit vector $v \in T_x\mathbb{S}^2$, the function $\langle \mathrm{Log}_x(\cdot), v\rangle$ does not belong to $H^1(\mathbb{S}^2)$, hence to no $H^\tau(\mathbb{S}^2)$ with $\tau \geq 1$.*

Proof. Step 1 (normal coordinates at the antipode). Let $x \in \mathbb{S}^2$ and let $z \in T_{-x}\mathbb{S}^2 \cong \mathbb{R}^2$ be normal coordinates centred at $-x$, $y = \mathrm{Exp}_{-x}(z)$, valid for $\|z\| < \pi$. Every $y \neq x$ is joined to $-x$ by a unique minimizing geodesic, and the great circle through $-x$ and y passes through x, so that

$$d(x,y) \;=\; \pi - d(-x,y) \;=\; \pi - \|z\| \qquad \text{exactly.}$$

Step 2 (the two integrands in these coordinates). From Step 1,

$$\varphi_x(y) \;=\; \tfrac{1}{2}(\pi - \|z\|)^2 \;=\; \tfrac{\pi^2}{2} - \pi\,\|z\| + \tfrac{1}{2}\,\|z\|^2 .$$

Since $T_x\mathbb{S}^2$ and $T_{-x}\mathbb{S}^2$ are the same plane $x^\perp \subset \mathbb{R}^3$, we may regard $z \in x^\perp$ and write $y = \mathrm{Exp}_{-x}(z) = -\cos(\|z\|)\,x + \sin(\|z\|)\,z/\|z\|$. The minimizing geodesic from x to y has length $\theta = \pi - \|z\|$ and runs in the plane spanned by x and z; writing $y = \cos\theta\, x + \sin\theta\, v$ with $v \in x^\perp$ a unit vector, and using $\cos\theta = -\cos\|z\|$, $\sin\theta = \sin\|z\|$, we find $v = z/\|z\|$, so that

$$\mathrm{Log}_x(y) \;=\; (\pi - \|z\|)\,\frac{z}{\|z\|} \;=\; \pi\,\frac{z}{\|z\|} - z.$$

Both functions are smooth on $\mathbb{S}^2 \setminus \{-x\}$ (equivalently, off $z = 0$; note that $\|z\|^2$ and z are smooth).

Step 3 (reduction to homogeneous functions). Multiplying by a cutoff $\chi \in C_c^\infty(\{\|z\| < \pi\})$ equal to 1 near 0, the Sobolev regularity of φ_x on $\mathbb{S}^2$ is that of $\chi\,\|z\|$ on $\mathbb{R}^2$, and that of the components of Log_x is that of $\chi\,\omega(z/\|z\|)$ with ω a coordinate function on the circle (the term z being smooth).

Step 4 (Fourier count). Let $\psi_m(z) = \|z\|^m\,\omega(z/\|z\|)$ with ω smooth on the circle and not the restriction of a homogeneous polynomial of degree m (so that ψ_m is not smooth at 0; for $m = 1$ the constant $\omega \equiv 1$ is admitted, since a homogeneous polynomial of degree 1 has no constant term, and for $m = 0$ constants are excluded). Its distributional Fourier transform is homogeneous of degree $-m-2$ and smooth away from the origin (Grafakos, 2014, Ch. 2), and multiplication by χ convolves it with a Schwartz function, so

$$|\widehat{\chi\psi_m}(\xi)| \;\leq\; C\,(1 + \|\xi\|)^{-m-2}.$$

Therefore

$$\int_{\mathbb{R}^2} (1 + \|\xi\|^2)^\tau\, |\widehat{\chi\psi_m}(\xi)|^2\,\mathrm{d}\xi \;<\; \infty \qquad \text{whenever} \quad 2\tau - 2m - 4 < -2, \text{ that is, } \tau < m + 1.$$

For $\psi_1 = \|z\|$ ($\omega \equiv 1$) this gives $\tau < 2$, and for $\psi_0 = \omega(z/\|z\|)$ it gives $\tau < 1$.

Step 5 (the gradient integrand is not in H^1). Let $v \in T_x\mathbb{S}^2$ be a unit vector and write $z = \varrho(\cos\vartheta, \sin\vartheta)$, $v = (\cos\vartheta_v, \sin\vartheta_v)$ in the coordinates of Step 2. By Step 2, $g(z) := \big\langle \mathrm{Log}_x(\mathrm{Exp}_{-x}(z)), v\big\rangle = \pi\cos(\vartheta - \vartheta_v) - \langle z, v\rangle$, whose classical gradient on the punctured disc satisfies $\|\nabla g\| \geq \pi|\sin(\vartheta - \vartheta_v)|/\varrho - 1$. On the set where $|\sin(\vartheta - \vartheta_v)| \geq \frac{1}{2}$, of angular measure $4\pi/3$, and $\varrho \leq \pi/4$, this is at least $\pi/(4\varrho)$, so

$$\int_{0<\|z\|<\pi/4} \|\nabla g\|^2\,\mathrm{d}z \;\geq\; \frac{4\pi}{3}\int_0^{\pi/4} \Big(\frac{\pi}{4\varrho}\Big)^2 \varrho\,\mathrm{d}\varrho \;=\; \infty.$$

The metric of $\mathbb{S}^2$ in normal coordinates is uniformly equivalent to the Euclidean one on $B(0,\pi/4)$, and a function in $H^1(\mathbb{S}^2)$ has a square-integrable weak gradient there that coincides with the classical gradient off $z = 0$. Hence $\langle \mathrm{Log}_x(\cdot), v\rangle \notin H^1(\mathbb{S}^2) \supseteq H^\tau(\mathbb{S}^2)$ for $\tau \geq 1$. □

Proposition E.9 (the Fréchet problem on $\mathbb{S}^d$ is electrostatics)**.** *Let $d \geq 1$ and write $\delta = d(x,y)$.*

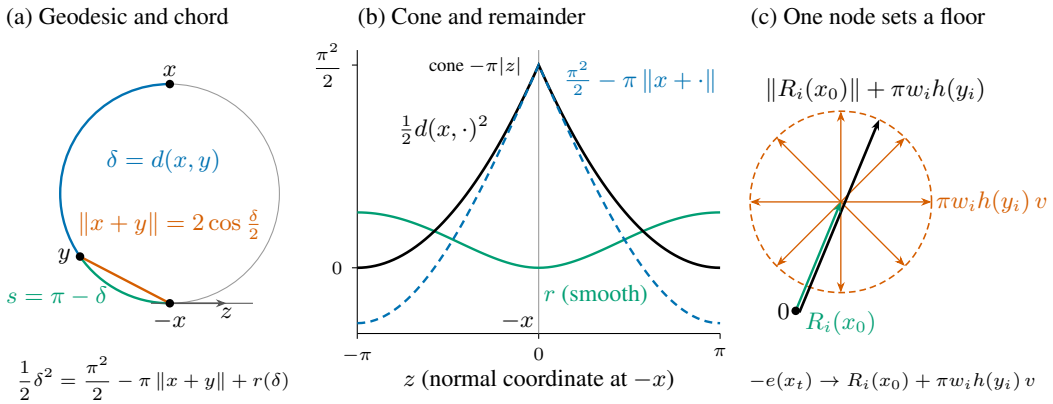

Figure 9: **The Fréchet problem on the sphere is electrostatics, and one node sets a floor.** (a) On the great circle through x and y, half the squared geodesic distance is $\pi^2/2$ minus π times the chordal distance $\|x+y\|$ from y to the antipode $-x$, plus a remainder r (Proposition E.9(i)). (b) Along this circle, in the normal coordinate z at $-x$, the loss has a cone at $-x$; the chordal term carries exactly this cone, and the remainder $r = \frac{1}{2}z^2 - \frac{\pi}{24}|z|^3 + \dots$ is C^2 there. So, up to smooth terms, the value error is $-\pi$ times the chordal potential of the charge at $-x$ and the field is π times its chordal field (Proposition E.9(iii)); the energy of the charge is C_d times its squared cap discrepancy (Lemma B.6). (c) Near the antipode $x_0 = -y_i$ of a node, $\mathrm{Log}_{x_t}(y_i) \to \pi v$ as $x_t \to x_0$ from direction v, so the limits of $-e(x_t)$ fill the circle of radius $\pi w_i h(y_i)$ around the continuous part $R_i(x_0)$, and $\sup\|e\| \geq \|R_i(x_0)\| + \pi w_i h(y_i) \geq \pi w_i h(y_i)$ (Proposition D.13).

(i) For all $x, y \in \mathbb{S}^d$,

$$\tfrac{1}{2}\delta^2 = \tfrac{\pi^2}{2} - \pi\|x+y\| + r(\delta), \qquad r(\delta) = \tfrac{1}{2}\delta^2 + 2\pi\cos\tfrac{\delta}{2} - \tfrac{\pi^2}{2}.$$

(ii) The function $R_x = r(d(x,\cdot))$ is C^2 on $\mathbb{S}^d$ and C^∞ off $\{-x\}$, and in normal coordinates z at $-x$, $R_x(\mathrm{Exp}_{-x} z) = \frac{1}{2}\|z\|^2 + \|z\|^3\varphi(\|z\|^2)$ with φ entire, $\varphi(0) = -\pi/24$. The vector function $S_x(y) = \mathrm{grad}_x\, r(d(x,y)) \in T_x\mathbb{S}^d$, extended by continuity, is C^1 on $\mathbb{S}^d$ and C^∞ off $\{-x\}$, and $S_x(\mathrm{Exp}_{-x} z) = z - \|z\|\, z\,\psi(\|z\|^2)$ with ψ entire, $\psi(0) = \pi/8$. Consequently $R_x \in H^\tau(\mathbb{S}^d)$ for $\tau < 3 + d/2$ and the components of S_x lie in $H^\tau(\mathbb{S}^d)$ for $\tau < 2 + d/2$, with norms independent of x.

(iii) Let ϱ be a finite signed measure with $\varrho(\mathbb{S}^d) = 0$, $U_\varrho(z) = \int\|z-y\|\,\mathrm{d}\varrho(y)$ its chordal potential and $\Phi_\varrho(x) = \frac{1}{2}\int d(x,y)^2\,\mathrm{d}\varrho(y)$. Then $\Phi_\varrho(x) = -\pi\,U_\varrho(-x) + \varrho R_x$. If $\varrho = Q - \sigma$ for a probability measure Q with finite support not containing $-x$, then Φ_ϱ is differentiable at x and

$$\mathrm{grad}\,\Phi_\varrho(x) = \pi\,\nabla U_\varrho(-x) + \varrho S_x, \qquad \nabla U_\varrho(z) = \int \frac{P_z(z-y)}{\|z-y\|}\,\mathrm{d}Q(y) \quad (z \notin \mathrm{supp}\,Q),$$

where P_z is the orthogonal projection onto $T_z\mathbb{S}^d = z^\perp$, and $T_x\mathbb{S}^d = T_{-x}\mathbb{S}^d$.

(iv) With $k_0 = \iint K_d\,\mathrm{d}\sigma\,\mathrm{d}\sigma$, $C_d\,\mathcal{D}(\varrho)^2 = -\varrho U_\varrho$, $\sup|U_\varrho| \leq C_d\sqrt{1-k_0}\,\mathcal{D}(\varrho)$, $|\varrho R_x| \leq c_r\,\mathcal{D}(\varrho)$ and $\|\varrho S_x\| \leq c_r'\,\mathcal{D}(\varrho)$, where c_r and c_r' are the seminorms in H_{K_d} (without the constant component) of R_x and of the unit-direction components of S_x; they do not depend on x.

Proof. (i) $\|x+y\|^2 = 2 + 2\langle x, y\rangle = 2 + 2\cos\delta = 4\cos^2(\delta/2)$ and $\cos(\delta/2) \geq 0$ for $\delta \in [0,\pi]$, so $\pi\|x+y\| = 2\pi\cos(\delta/2)$; the identity is the definition of r.

(ii) The function r is an even entire function of δ, hence an entire function of $\delta^2 = d(x,y)^2$, which is smooth in y off $\{-x\}$; so R_x is smooth off $\{-x\}$. Near $-x$ put $s = \pi - \delta = d(-x,y)$. Then $r = \frac{1}{2}s^2 + \left(2\pi\sin\frac{s}{2} - \pi s\right)$, and the bracket is an odd entire function of s whose expansion begins with $-\pi s^3/24$, that is $s^3\varphi(s^2)$ with φ entire and $\varphi(0) = -\pi/24$; with $s = \|z\|$ this is the stated form. For $y \notin \{x, -x\}$, $S_x(y) = r'(\delta)\,\mathrm{grad}_x\, d(x,y) = -r'(\delta)\,\mathrm{Log}_x(y)/\delta$. Since $r'(\delta) = \delta - \pi\sin(\delta/2)$ is odd, $r'(\delta)/\delta$ is an entire function of δ^2 and S_x is smooth near $y = x$. Near

$-x$, (30) gives $\mathrm{Log}_x(y)/\delta = z/\|z\|$, and $r'(\pi - s) = -s + \pi(1 - \cos\frac{s}{2}) = -s + s^2\psi(s^2)$ with ψ entire, $\psi(0) = \pi/8$, so $S_x = z - \|z\|\, z\,\psi(\|z\|^2)$. The terms $\|z\|^3\,\varphi(\|z\|^2)$ and $\|z\|\, z_k\psi(\|z\|^2) = \|z\|^2\,(z_k/\|z\|)\,\psi(\|z\|^2)$ have the form of Lemma D.4 with $m = 3$ and $m = 2$; by that lemma with $\alpha = m + d/2 - \tau > 0$ and $r \to 0$, and Lemma D.5, they lie in H^τ for $\tau < 3 + d/2$ and $\tau < 2 + d/2$. Rotations g satisfy $R_{gx}(gy) = R_x(y)$ and the Sobolev norms are rotation invariant, so the norms do not depend on x.

(iii) Integrate (i) against ϱ and use $\varrho(\mathbb{S}^d) = 0$ and $\|x + y\| = \|(-x) - y\|$. For the gradient, differentiate under the integral: for $y \neq \pm x$, $\mathrm{grad}_x \|x + y\| = P_x(x + y)/\|x + y\| = P_x(y)/\|x + y\|$, bounded by 1, and at $y = x$ both sides of (i) have gradient 0; dominated convergence handles σ and the atoms of Q are not at $-x$. Hence $\mathrm{grad}\,\Phi_\varrho(x) = -\pi\int P_x(y)/\|x + y\|\,\mathrm{d}\varrho(y) + \varrho S_x$. On the other hand $\nabla U_\varrho(-x) = \int P_{-x}(-x - y)/\|x + y\|\,\mathrm{d}\varrho(y) = -\int P_x(y)/\|x + y\|\,\mathrm{d}\varrho(y)$, since $P_{-x} = P_x$ and $P_x x = 0$. Finally U_σ is constant by rotation invariance, so $\nabla U_\varrho = \nabla U_Q$.

(iv) The first identity is (13). By Lemma B.6, $\|z - y\| = C_d(1 - K_d(z, y))$, so $U_\varrho(z) = -C_d\int K_d(z, y)\,\mathrm{d}\varrho(y) = -C_d\,\langle K_d(z, \cdot), \tilde\varrho\rangle_{H_{K_d}}$ with $\tilde\varrho = \int K_d(\cdot, y)\,\mathrm{d}\varrho(y)$ and $\|\tilde\varrho\|_{H_{K_d}} = \mathcal{D}(\varrho)$. Since $\varrho(\mathbb{S}^d) = 0$, $\tilde\varrho$ has no constant component, so $K_d(z, \cdot)$ may be replaced by $K_d(z, \cdot) - k_0$, whose squared norm is $K_d(z, z) - k_0 = 1 - k_0$. The bounds on $\varrho R_x = \langle R_x, \tilde\varrho\rangle_{H_{K_d}}$ and on $\langle\varrho S_x, v\rangle$ follow in the same way, because $H_{K_d} = H^{(d+1)/2}(\mathbb{S}^d)$ with equivalent norms and $(d + 1)/2 < 2 + d/2$ by (ii). □

On $\mathbb{S}^2$ the decomposition also gives the best constant of the value certificate in closed form. By Lemma B.6, $-\pi\|x + y\| = 4\pi K_2(-x, y) - 4\pi$, so up to constants $\frac{1}{2}d(x, \cdot)^2 = 4\pi K_2(-x, \cdot) + R_x$, and by the reproducing property its seminorm in H_{K_2} without the constant component is

$$\kappa_{\mathrm{val}}^2 = \left\|\tfrac{1}{2}d(x, \cdot)^2\right\|_{H_{K_2},0}^2 = 16\pi^2(1 - k_0) + c_r^2 - 8\pi\bar R, \qquad \bar R = \int R_x\,\mathrm{d}\sigma = \frac{4\pi}{3} - \frac{\pi^2}{4} - 1,$$

where $c_r = 1.4966$ is the seminorm of R_x, computed from Legendre coefficients that decay like ℓ^{-5} (`code/check_electro_constant.py`); the cross term is $8\pi\big(R_x(-x) - \bar R\big)$ and $R_x(-x) = r(\pi) = 0$. Hence, for every node set,

$$\sup_{x\in\mathbb{S}^2}\big|f_N(x) - f(x)\big| \le \kappa_{\mathrm{val}}\,D_{L_2}(Y_N), \qquad \kappa_{\mathrm{val}} = 6.062 \qquad (h \equiv 1), \tag{36}$$

and the constant is sharp to within 1%: a single node attains $6.006\,D_{L_2}$. The same seminorm computed directly from the Legendre coefficients of $\frac{1}{2}d(x, \cdot)^2$ converges only like the inverse of the truncation degree, because those coefficients decay like ℓ^{-3}; the decomposition isolates the slowly decaying part in closed form. On $2 \cdot 10^4$ random decision points, Fibonacci, scrambled-Lambert, i.i.d. and clustered node sets give ratios between 0.08 and 3.39.

The gradient part of the proposition moves the whole question about fields to the chordal field ∇U_ϱ of the charge at the antipode: an average of vectors of length at most 1 that point away from the nodes, jumping by $2/N$ at each node. How large it can be given the energy $C_d\mathcal{D}(\varrho)^2$ is Corollary D.7, Theorem D.11, and Proposition D.13. Figure 9 draws the decomposition and the floor.

E.5 Charts are blow-ups

A cylindrical chart (32) with $z(0) = -1$ and $z(1) = 1$ maps the edges $\{u_1 = 0\}$ and $\{u_1 = 1\}$ of the square to the poles $\mp e_3$ and is a diffeomorphism elsewhere, up to the seam $\{u_2 \in \{0, 1\}\}$. In polar terms it is the real blow-up of the two poles: each pole is replaced by a circle of directions, parametrized by u_2. A function on the sphere that is singular only in its direction at a pole becomes, after composition with the chart, a function of u_2 on that edge.

Proposition E.10 (a chart resolves the cut points it blows up)**.** *Let $T = T_z$ be a cylindrical chart (32) with an increasing C^2 profile, $z' > 0$ on $[0, 1]$, $z(0) = -1$, $z(1) = 1$, and let $h \in C^2(\mathbb{S}^2)$. For $x \in \mathbb{S}^2$ let $g_x(u) = h(T(u))\,\mathrm{Log}_x(T(u))$, defined for $T(u) \neq -x$.*

(a) If $x \in \{e_3, -e_3\}$, then g_x has a continuous extension to the closed square with $V_{\mathrm{HK}}(g_x) \le C\,\|h\|_{C^2}$, and for every point set with nodes off $T^{-1}(-x)$, $\|e_N(x)\| \le C\,\|h\|_{C^2}\,D_N^*$.

(b) If $-x = T(u_0)$ with u_0 in the open square off the seam, then $\|e_N(x)\| \le C\, D_N^*(1+\log(1/D_N^*))$ for every point set with nodes off u_0, uniformly for x in compact sets of such points; if moreover $h(-x) \neq 0$, then $V_{\mathrm{HK}}(g_x) = \infty$.

Here $e_N(x) = \operatorname{grad} f_N(x) - \operatorname{grad} f(x)$ for $f_N(x) = \frac{1}{N}\sum_i h(T(u_i))\frac{1}{2}d(x,T(u_i))^2$ and $f(x) = \int h(T(u))\frac{1}{2}d(x,T(u))^2\,\mathrm{d}u$, so that $e_N(x) = -(Q-\lambda)g_x$.

Proof. (a) Let $x = e_3$ and write $r = \sqrt{1-z^2}$, $\phi = 2\pi u_2$. For $u_1 > 0$, $T(u) = (r\cos\phi, r\sin\phi, z)$ and $\operatorname{Log}_{e_3} T(u) = \arccos(z)\,(\cos\phi, \sin\phi, 0)$, which extends continuously to $u_1 = 0$ by $\pi(\cos\phi, \sin\phi, 0)$: the direction at the blown-up antipode is the smooth function $u_2 \mapsto (\cos 2\pi u_2, \sin 2\pi u_2, 0)$. Put $A = h \circ T$ and $B(u) = \arccos(z(u_1))(\cos\phi, \sin\phi, 0)$, both continuous on the closed square and C^2 on the open square. From the proof of Proposition E.2, $|\partial_{u_1} A| \le \|h\|_{C^1} z'/r$, $|\partial_{u_2} A| \le 2\pi\,\|h\|_{C^1}$ and $|\partial_{u_1}\partial_{u_2} A| \le 2\pi\,\|h\|_{C^2}\,(z' + z'/r)$; directly, $|\partial_{u_1} B| = z'/r$, $|\partial_{u_2} B| \le 2\pi^2$ and $|\partial_{u_1}\partial_{u_2} B| = 2\pi z'/r$. By the product rule $|\partial_{u_1}\partial_{u_2}(AB)| \le C\,\|h\|_{C^2}\,(z' + z'/r)$, whose integral over the square is at most $C\,\|h\|_{C^2}\,(2+\pi)$, since $\int_0^1 z'/r\,\mathrm{d}u_1 = \pi$. On the face $\{u_2 = 1\}$ the u_1-derivative of AB is at most $C\,\|h\|_{C^1} z'/r$, integrable; on the face $\{u_1 = 1\}$, $T(1,u_2) = e_3$ and $AB = 0$. Lemma E.1 gives the variation bound, and Lemma C.2 the field bound. For $x = -e_3$, $\operatorname{Log}_{-e_3} T(u) = \arccos(-z)(\cos\phi, \sin\phi, 0)$, the blown-up antipode is the edge $\{u_1 = 1\}$, on which $g_x = h(e_3)\pi(\cos\phi, \sin\phi, 0)$ has variation at most $2\pi^2|h(e_3)|$ in u_2, and the rest is identical.

(b) Near u_0 the chart is a C^2 diffeomorphism onto a neighbourhood of $-x$, since $|\partial_{u_1} T \times \partial_{u_2} T| = 2\pi z' > 0$, so $Z(u) = \operatorname{Exp}_{-x}^{-1}(T(u))$ is a C^2 diffeomorphism near u_0 with $Z(u_0) = 0$, and by (30) $g_x = h(T)\left(\pi Z/\|Z\| - Z\right)$ there. Let ψ be a smooth cutoff on the sphere that equals 1 near $-x$ and vanishes outside a ball around $-x$ avoiding the poles and the seam, with T^{-1} of that ball inside a compact $U \subset (0,1)^2$. Then $g_x = \pi\,(\psi h)(T)\,Z/\|Z\| + a$, where $a = -(\psi h)(T)Z + \big((1-\psi)h\operatorname{Log}_x\big) \circ T$; the first term of a is C^2 with compact support in the open square and the second has bounded variation by Proposition E.2, because $(1-\psi)h\operatorname{Log}_x$ is C^2 on the sphere. Theorem D.10(a), applied to each component with ω the coordinate functions of $\mathbb{S}^1$, gives the bound, with constants depending continuously on x. If $h(-x) \neq 0$, the mixed derivative of $(\psi h)(T)Z/\|Z\|$ is of exact order $\|u-u_0\|^{-2}$ in a sector around u_0, its integral diverges logarithmically, and the variation is infinite; if $h(-x) = 0$ the factor $h(T(u)) = O(\|u-u_0\|)$ lowers the order of the singularity and the field integrand has bounded variation. □

Four remarks complete the statement. *Values.* In both cases the value integrand $u \mapsto h(T(u))\frac{1}{2}d(x,T(u))^2$ has bounded Hardy–Krause variation, so the value error is at most $C\,D_N^*$ (the value column of Table 1): at a pole it is $h(T(u))\frac{1}{2}\arccos(\pm z(u_1))^2$, whose mixed derivative is $O(z' + z'/r)$ as in the proof of (a); at an interior u_0 its singular part is the cone $-\pi(\psi h)(T)\,\|Z\|$, whose mixed derivative is $O(\|u-u_0\|^{-1})$, integrable, and Lemma E.1 applies with the exceptional point u_0. *Decreasing profiles*, such as the Lambert map $z = 1 - 2u_1$ of the experiments, reduce to increasing ones: with the reflection $S(y) = (y_1, y_2, -y_3)$, a decreasing chart is $S \circ T$ for an increasing one T, and $g_x \circ (S\circ T) = S\left[(h\circ S)\operatorname{Log}_{Sx}\right] \circ T$ because S is an isometry; the frame-free variation is invariant under the linear isometry S, and Sx is a pole of T exactly when x is a pole of $S \circ T$. *The seam*: if $-x$ lies on the meridian $\{\phi = 0\}$ but is not a pole, the singularity sits at the two boundary points $(u_1^*, 0)$ and $(u_1^*, 1)$ and part (b) does not apply as stated; a half-disc version of Theorem D.10(a) gives the same bound, and rotating the azimuth origin avoids the case. *Uniformity*: for x in a compact set K of points with $-x$ off the poles and the seam, the cutoff radius, the neighbourhood U, the bi-Lipschitz constants of Z and its C^2 norms can be chosen uniformly on K, which gives the uniform constant.

What the chart does to the minimizer. Where the sampled objective is strongly convex and smooth near the segment from $x^\star$ to $\hat{x}_N$, the one-point reading makes the minimizer inherit the rate of the field at $x^\star$. On the sphere with full-support nodes this holds for typical draws but not for every draw: a node within distance of order $h(y_i)/(N\lambda)$ of $-x^\star$ puts a concave kink of f_N next to $x^\star$, the sample objective is then neither smooth nor convex there, and the minimizer can move by order $1/N$ even when $e_N(x^\star) = 0$ (a symmetric configuration with three nodes at distance $5\cdot 10^{-4}$ from $-e_3$ makes $x^\star$ a strict local maximum of f_N). The guarantee that covers every draw is Corollary D.8,

through the uniform field bound near $x^\star$. The standard von Mises–Fisher chart of a law centred at $x^\star$ has its pole at $x^\star$ and resolves $-x^\star$; this is why its field at $x^\star$ has bounded variation and its sample Fréchet means converge fast in Appendix G.4, where scrambled nets are observed to keep a rate near $N^{-3/2}$; the field integrand of a resolved chart has integrable, not bounded, mixed derivatives near the blown-up edges, so this randomized rate is observed, not proved. A chart with its poles elsewhere, the Lambert chart of a law whose mean is off the poles for instance, leaves $-x^\star$ interior and pays the logarithm of Proposition E.10(b), and scrambled nets lose their fast rate there (Appendix G.6). The remedy is to rotate the chart so that its pole is at an estimate of $x^\star$ and carry the law by importance weights.

Proposition E.11 (flat tori). *Let $A \in \mathrm{GL}_2(\mathbb{R})$, $\Lambda = A\mathbb{Z}^2$, $\mathcal{M} = \mathbb{R}^2/\Lambda$ with the flat metric, and $T(u) = Au \bmod \Lambda$ for $u \in [0,1)^2$, which carries Lebesgue measure to the normalized Riemannian measure. For $F = \frac{1}{2}d^2$ there is C_A such that for every $x \in \mathcal{M}$ and every point set in $[0,1)^2$*

$$|\Phi_N(x)| \le C_A\, D_N^*, \qquad \|e_N(x)\| \le C_A\, (D_N^*)^{1/2},$$

the second when no node lies on $T^{-1}(\mathrm{Cut}(x))$. If A is diagonal, $\|e_N(x)\| \le C_A\, D_N^*$.

Proof. Lift x to $\tilde{x} \in A[0,1)^2$ and let δ_A be the diameter of $A[0,1]^2$. For $u \in [0,1]^2$, $d(x,T(u)) = \min_{\ell\in\Lambda} \|Au - \tilde{x} - \ell\|$, and since $\|Au - \tilde{x}\| \le \delta_A$ the minimum is attained in the finite set $\Lambda_0 = \{\ell \in \Lambda : \|\ell\| \le 2\delta_A\}$, which does not depend on x.

Values. $\frac{1}{2}d(x,T(u))^2 = \frac{1}{2}\|Au - \tilde{x}\|^2 - \psi(u)$ with $\psi(u) = \max_{\ell\in\Lambda_0}\big(\langle Au - \tilde{x}, \ell\rangle - \frac{1}{2}\|\ell\|^2\big)$. The quadratic has bounded mixed derivatives, so its variation is at most a constant C_A by (11). The function ψ is a maximum of affine functions, hence convex and Lipschitz with constant $L_A = 2\delta_A\|A\|$. For a convex C^2 function ϕ near $[0,1]^2$, $|\partial_1\partial_2\phi| \le (\partial_1^2\phi\,\partial_2^2\phi)^{1/2} \le \frac{1}{2}\Delta\phi$, so $\int_{[0,1]^2}|\partial_1\partial_2\phi| \le \frac{1}{2}\int_{\partial[0,1]^2}\partial_n\phi \le 2\,\mathrm{Lip}(\phi)$, and each face variation is at most $\mathrm{Lip}(\phi)$. The mollifications of ψ are convex, smooth and L_A-Lipschitz and converge to ψ uniformly, and every sum $\sum_R|\Delta(\cdot\,;R)|$ over a fixed partition passes to the limit; hence $V_{\mathrm{HK}}(\psi) \le 4L_A$, and (12) gives the value bound.

Fields. Let $V = \{v : \|v\| \le \|v - \ell\| \text{ for all } \ell \in \Lambda\}$ be the Voronoi cell of Λ, a convex polygon. Off $T^{-1}(\mathrm{Cut}(x))$ the minimizing $\ell(u)$ is unique and $\mathrm{Log}_x T(u) = Au - \tilde{x} - \ell(u)$ in $T_x\mathcal{M} = \mathbb{R}^2$. The set where $\ell(u) = \ell$ lies in $K_\ell = [0,1]^2 \cap A^{-1}(\tilde{x} + \ell + V)$, a convex polygon of perimeter at most 4, and different K_ℓ overlap only on $T^{-1}(\mathrm{Cut}(x))$, a Lebesgue null set that contains no node. The rule and Lebesgue measure therefore integrate $G_x \circ T = -\mathrm{Log}_x \circ T$ and $-(Au-\tilde{x}) + \sum_{\ell\in\Lambda_0}\ell\,\mathbf{1}_{K_\ell}$ alike, and Theorem D.10(b), applied to each component with constant amplitudes and at most $|\Lambda_0|$ polygons, gives $C_A(D_N^*)^{1/2}$. If A is diagonal, V is an axis-parallel rectangle, every K_ℓ is an axis-parallel box, whose indicator has Hardy–Krause variation at most 8, and (12) gives $C_A D_N^*$. □

The value integrand of the torus is the case $m = \beta = 1$, $c = 1$ of the law in which no logarithm appears: its mixed derivative is a finite measure, whose singular part lies on the cut locus with the jump of the gradient as density, and the variation is finite.

Remark E.12 (other cut loci, other manifolds). (a) *Codimension.* On $\mathbb{S}^d$ the cut locus of x is the point $-x$, of codimension d. On real projective spaces, on flat tori and on most compact manifolds it is a hypersurface, of codimension one, and the field integrand jumps across it. The dilation argument of Appendix D predicts the exponent $1/(\tau + \frac{1}{2})$ in H^τ for such jumps; Theorem D.10(b) proves the analogue $(D^*)^{1/2}$ for polygonal jumps in the square, and Proposition E.11 applies it to flat tori.

(b) *The chart decides alignment.* On the flat torus $\mathbb{R}^s/\mathbb{Z}^s$ with the identity chart, $\mathrm{Cut}(x)$ is a union of axis-parallel hyperplanes and Log_x has coordinates $\mathrm{saw}(u_j - x_j)$, each a function of one variable of bounded variation, so $V_{\mathrm{HK}} < \infty$ and the rate is D^*. For a quotient $\mathbb{R}^s/\Lambda$ by a non-rectangular lattice with its fundamental-cell chart the cut-locus hyperplanes are oblique and $V_{\mathrm{HK}} = \infty$; in two dimensions Proposition E.11 shows that the field costs the square root $(D_N^*)^{1/2}$, junctions and boundary crossings included, while the value certificate stays linear. This is the torus row of Table 1.

(c) *Smeariness.* For i.i.d. samples, the smeary central limit theorems of Hotz & Huckemann (2015); Eltzner & Huckemann (2019) show that Fréchet means can converge more slowly than $N^{-1/2}$. This concerns degeneracy of the Fréchet functional, not the regularity of the field integrand, and is a separate phenomenon from the one studied here.

F Designing the charge: Riemannian optimization of node sets

Low-discrepancy point sets on the cube come from number theory; on a sphere the natural constructions are variational. Spherical designs, minimal-energy configurations and QMC designs are all minimizers of an energy over $(\mathbb{S}^d)^N$ (Hardin & Saff, 2004; Brauchart & Grabner, 2015; Borodachov et al., 2019; Brauchart et al., 2014), and $(\mathbb{S}^d)^N$ is a Riemannian manifold, so constructing nodes on a curved space is itself a Riemannian optimization problem; Riemannian ascent on such energies is how spherical designs and minimal-energy configurations have long been computed (Gräf & Potts, 2011; Hardin & Saff, 2004). In the language of Proposition E.9, the Stolarsky energy $-\varrho U_\varrho = C_d D_{L_2}^2$ is the self-energy of the discrepancy charge, and maximizing the sum of distances moves the charges to lower it.

By (14), minimizing $D_{L_2}(X_N)$ over $X_N \in (\mathbb{S}^d)^N$ is the same as maximizing the sum of distances

$$\mathcal{E}(X_N) = \sum_{1\le i<j\le N} \|x_i - x_j\|, \qquad X_N = (x_1, \dots, x_N) \in (\mathbb{S}^d)^N \subset \mathbb{R}^{(d+1)N}. \tag{37}$$

$\mathcal{E}$ is smooth on the open set where the points are pairwise distinct. Equip $(\mathbb{S}^d)^N$ with the product metric; the tangent space at X_N is $\prod_i \{v_i \in \mathbb{R}^{d+1} : \langle v_i, x_i\rangle = 0\}$.

Proposition F.1 (gradient of the sum of distances)**.** *At a configuration of pairwise distinct points, the Riemannian gradient of $\mathcal{E}$ has i-th component*

$$(\operatorname{grad}\mathcal{E})_i \;=\; (I - x_i x_i^\top) \sum_{j\neq i} \frac{x_i - x_j}{\|x_i - x_j\|},$$

and X_N is a critical point if and only if, for every i, the sum of the unit vectors from the other points to x_i is parallel to x_i. Riemannian gradient ascent with the normalization retraction $R_x(v) = (x+v)/\|x+v\|$ and an Armijo step increases $\mathcal{E}$ monotonically, and every accumulation point with pairwise distinct points is critical.

Proof. The Euclidean gradient of $\|x_i - x_j\|$ with respect to x_i is $(x_i - x_j)/\|x_i - x_j\|$, and the Riemannian gradient on the product of spheres is its orthogonal projection onto the tangent space (Absil et al., 2008, Ch. 3); the characterization of critical points is $\operatorname{grad}\mathcal{E} = 0$. The convergence statement is the standard one for retraction-based line-search methods (Absil et al., 2008, Thm. 4.3.1), applied to $-\mathcal{E}$ on the open submanifold of distinct configurations. □

The landscape is nonconvex and has many critical points, but the discrepancy of every configuration produced, critical or not, is known exactly: by (14), $D_{L_2}(X_N)^2 = \frac{1}{4}\big(\frac{4}{3} - \frac{2}{N^2}\mathcal{E}(X_N)\big)$ on $\mathbb{S}^2$. The deficit $\delta(X_N) = \frac{4}{3} - \frac{1}{N^2}\sum_{i,j}\|x_i - x_j\|$ is $\Theta(N^{-1})$ for i.i.d. points and, by Beck (1984), $\Theta(N^{-3/2})$ for optimal configurations on $\mathbb{S}^2$; gradient ascent from a good initialization (for instance the Lambert image of a scrambled Sobol' net) reduces the deficit toward the optimal order, see Appendix G.

Spherical designs, the sets that integrate all polynomials up to a given degree exactly, are computed the same way by Gräf & Potts (2011), who minimize a smooth polynomial energy over $(\mathbb{S}^2)^N$ with Riemannian Newton and conjugate-gradient methods; and every discrete Riesz energy of Hardin & Saff (2004) defines an analogous problem on M^N for any compact manifold M. In each case the retraction is the normalization or the exponential map of M, and the machinery of Appendix B.1 applies.

What the energy certifies. By Corollaries D.7 and D.14, a node set that attains the optimal order $D_{L_2} = O(N^{-(d+1)/(2d)})$ carries a value certificate of that order and a field certificate of order $N^{-(d+1)/(2d+1)}$, which is $N^{-3/4}$ and $N^{-3/5}$ on $\mathbb{S}^2$, and Theorem D.11 shows that nothing better follows from D_{L_2} alone. The floor $\pi \max h/N$ of Proposition D.13 is reached by node sets

equidistributed to higher order, such as spherical designs, which minimize a different energy (Gräf & Potts, 2011). The optimized node sets of our experiments nonetheless sit at the floor (Appendix G.5): the energy certifies less than the ascent delivers.

G Numerical experiments

Each experiment tests a statement of Appendices C–E: the one-point reading on the positive definite cone (Appendices G.3 and G.8), the mean-square and uniform readings and the role of the transport on the Grassmannian (Appendices G.1 and G.2), the charts and the floor on the sphere (Appendices G.4–G.6), and the dipole scalings behind the sharpness of the law (Appendix G.8). Throughout, "MC" is i.i.d. sampling, "Halton" the Halton sequence with the origin dropped, "SSobol" the scrambled Sobol' net of `scipy` (direction numbers of Joe & Kuo, 2008; Matoušek's random linear scrambling followed by a digital shift, Matoušek, 1998; Owen, 2003, with the default 30 randomized bits); random samplers are averaged in root-mean-square over replicates unless a median is stated, and slopes are least-squares fits of $\log(\text{error})$ against $\log N$ over the six largest N; the Grassmann noise fits use all nine sizes and the plateau fits all four or five sizes. All code is in Python and is listed in Appendix G.10. Three conditions apply to every plot. The scrambling implemented in `scipy` is the affine matrix scrambling of Lemma B.4 with 30 random digits; in its ideal form every point is uniform and the variance equals that of nested uniform scrambling, so (S1)–(S3) apply whenever the integrand satisfies their hypotheses, and the implemented rule differs from the ideal one by a deterministic error bounded in Lemma B.5 and measured in Appendix G.9, where it changes no reported root-mean-square error by more than a relative $5 \cdot 10^{-5}$. Every scrambled-net exponent below is a measured, finite-range value; the range $N \le 2^{17}$ cannot separate a power N^{-1} from $N^{-1}(\log N)^s$. And the Gaussian transports are outside the bounded-variation hypotheses and probe the boundary-growth regime instead, so that the bounded models (the SPD model (B) and the bounded Grassmann transport) are the ones whose integrands satisfy the hypotheses of the theorems.

Table 5: **What each experiment tests.** Measured values are fitted slopes over the six largest sizes unless stated; * marks behaviour beyond the proved bound.

statement; setting	proved prediction	measured
one point (3); Karcher means on $\mathcal{P}_3$	error $= \left\|\mathrm{Hess}\, f(x^\star)^{-1} e_N(x^\star)\right\|$, to second order	deviation $\le 2.4 \cdot 10^{-6}$ per draw (SSobol)
mean square; $\mathrm{Gr}(8,2)$, bounded transport	field MSE N^{-3}	-2.92 (all nine sizes)
reuse against refresh; $\mathrm{Gr}(8,2)$	same field and rate, other norm	factor ≈ 3 at $N = 4096$
Proposition 4.2(a); $\mathbb{S}^2$, resolved chart	$V_{\mathrm{HK}} < \infty$, error $\le C D_N^*$	-1.76 (field), -1.55 (mean)*
Proposition 4.2(b); $\mathbb{S}^2$, unresolved chart	$V_{\mathrm{HK}} = \infty$, error $\le C D_N^* \log(1/D_N^*)$	-0.91 to -1.02; -1.01 (mean)
Corollary 3.5; rules exact to degree t	field max in [floor, $C N^{-1+\epsilon}$]	at the floor, ratio $1.05 \to 1.00$
Theorem 3.4; optimized nodes	field max in [floor, $C D_{L_2}^{4/5}$]	at the floor ($\pm 2\%$, $N \ge 64$)*
values; exact rules, optimized nodes	$N^{-3/2+\epsilon}$; $C D_{L_2} \approx N^{-3/4}$	-1.46 (five largest); -1.48*

G.1 Principal subspaces on the Grassmannian

Setup. This is Example C.6 with $n = s = 8$, $p = 2$, $\Sigma = Q\,\mathrm{diag}(4, 3, 1, 0.9, 0.8, 0.7, 0.6, 0.5)Q^\top$ for a fixed orthogonal Q, and Gaussian noise $a(u) = W\Phi^{-1}(u)$, $W = Q\,\mathrm{diag}(\sqrt{\lambda})$. The population minimizer is the span $X^\star$ of the two dominant eigenvectors of Σ, and the error of a point $X \in \mathrm{Gr}(8,2)$ is $\left\|XX^\top - X^\star X^{\star\top}\right\|_F / \sqrt{2}$, the root of the sum of squared sines of the principal angles. The integrand is unbounded (Example C.6), so this experiment is outside the strict hypotheses of Appendices C.1 and C.3: the integrand is unbounded (Assumption C.2 fails) and the smoothness constant of $F(\cdot, u)$ grows with $\|a(u)\|^2$ (Assumption C.3 fails), and it probes the boundary-growth regime of Appendix B.2. Theorem C.15 itself does apply: only the population objective needs to be smooth, the gradient formula (16) holds by dominated differentiation because $\|\mathrm{grad}_X F(X,u)\|_F \le \|a(u)\|^2$ is integrable, and the gradient integrand is square integrable (Assumption C.5 with the variance bound of (S2)); what fails is the mixed smoothness behind the N^{-3} rate of (S3). The

inverse distribution function is evaluated without clipping; Appendix G.9 records that no scrambled coordinate fed to it was 0. It is also a problem in which $\dim \mathcal{M} = 12$ and $s = 8$, so that the $(\log N)^s$ factors are not negligible over the range of N used. The samplers are i.i.d. uniform ("MC"), Halton with the origin dropped, and scrambled Sobol' (see the note on the scrambling at the beginning of this section). Appendix G.2 repeats the experiment with a bounded transport for which every hypothesis holds.

Sample average approximation. For $N = 2^m$, $m = 4, \dots, 14$, the SAA minimizer $\hat{X}_N$ is the dominant eigenspace of the sample covariance $S_N = \frac{1}{N}\sum_i a_i a_i^\top$; it was also computed by Riemannian gradient descent on f_N with the QR retraction, which agreed with the eigendecomposition to $6 \cdot 10^{-11}$ for every N. Figure 10 shows the covariance error $\|S_N - \Sigma\|_F$ (which bounds $\sup_X \|\operatorname{grad} f_N(X) - \operatorname{grad} f(X)\|$, since the gradient integrand is linear in $aa^\top$) and the subspace error, with RMS over 20 replicates for the random samplers. Fitted slopes over the six largest N: MC -0.50 for both quantities; Halton -0.84 for both; scrambled Sobol' -1.07 (covariance) and -1.11 (subspace). At $N = 2^{14}$ the subspace errors are $1.6 \cdot 10^{-2}$ (MC), $1.4 \cdot 10^{-3}$ (Halton) and $3.1 \cdot 10^{-4}$ (scrambled Sobol'): a factor 50 between i.i.d. sampling and the scrambled net at the same number of samples. The subspace error inherits the rate of the gradient error. This is the mechanism of Theorem C.12, quadratic growth converting a gradient error into a distance error, although the theorem does not literally apply: f is not geodesically convex on $\mathrm{Gr}(8, 2)$, and the growth constant near $X^\star$ is the local eigengap $\lambda_2 - \lambda_3 = 2$ rather than a global modulus.

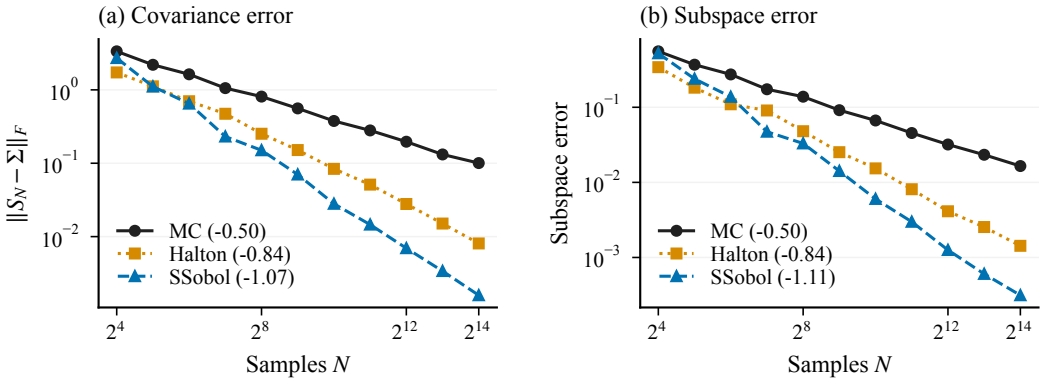

Figure 10: SAA on $\mathrm{Gr}(8, 2)$ with Gaussian noise in $s = 8$ dimensions. Left: covariance error $\|S_N - \Sigma\|_F$; right: subspace error of the SAA minimizer. RMS over 20 replicates; slopes over the six largest N.

Riemannian SGD with scrambled minibatches. Algorithm 2 was run with the QR retraction, constant step $\eta = 0.05$, minibatch sizes $N \in \{16, 64, 256, 1024\}$, a fixed total budget of 2^{20} samples ($K = 2^{20}/N$ iterations), 10 replicates, and either i.i.d. Gaussian minibatches or a freshly scrambled Sobol' net at every iteration. Figure 11(a) shows the error against the number of samples consumed. For each N both variants reach their noise floor after at most 12% of the budget. Theorem C.15 bounds the variance contribution to $\mathbb{E}\|\operatorname{grad} f\|^2$ by $L\eta\sigma_N^2$. If this contribution determines the attained plateau, local quadratic growth suggests that distance scales like $\sqrt{L\eta}\,\sigma_N$ up to the local growth constant; in particular its slope in N is $-a/2$ when $\sigma_N^2 \sim N^{-a}$. Panel (b) plots the plateau error (RMS over the last quarter of the iterations and over replicates) against N: slope -0.50 for MC and -1.13 for RQMC. Figure 12 measures $\sigma_N^2 = \mathbb{E}\|\widehat{g}_N(X^\star)\|^2$ directly at the optimum (200 replicates, $N = 2^4, \dots, 2^{12}$): slope -1.00 for MC and -2.19 for RQMC, so that the ratio of the two mean squared errors grows from 2 at $N = 32$ to 580 at $N = 4096$, and the predicted plateau slopes $-a/2$ (-0.50 and -1.10) match the measured ones to within 0.03. The plateau of RQMC-RSGD is therefore lower than that of RSGD by a factor that grows like $N^{0.6}$ over this range, at identical cost per iteration.

Two features of Figures 11 and 12 deserve comment. First, at the smallest minibatch $N = 16$ the scrambled net is *worse* than i.i.d. sampling (ratio 0.7 for the variance, 0.84 for the plateau). The

16-point net still stratifies every coordinate into 16 intervals; at this size its projection structure does not reduce the variance of the interaction terms of this integrand. This is consistent with the theory, in which no scrambled net is guaranteed to beat i.i.d. sampling at every N (the factor $\Gamma_{b,t,s} > 1$ in Lemma C.14(i)); the crossover occurs between $N = 16$ and $N = 64$, and the advantage grows monotonically afterwards. Second, the RQMC exponents (≈ -2.2 for the variance) lie between the N^{-2} of a deterministic $O(N^{-1})$ error and the N^{-3} of (S3), as expected for an integrand that grows at the boundary of the cube; by Lemma B.4 the linear scrambling has the same variance as nested scrambling, so the scrambling is not the cause. These are rates measured over $N \leq 2^{12}$, not asymptotic statements.

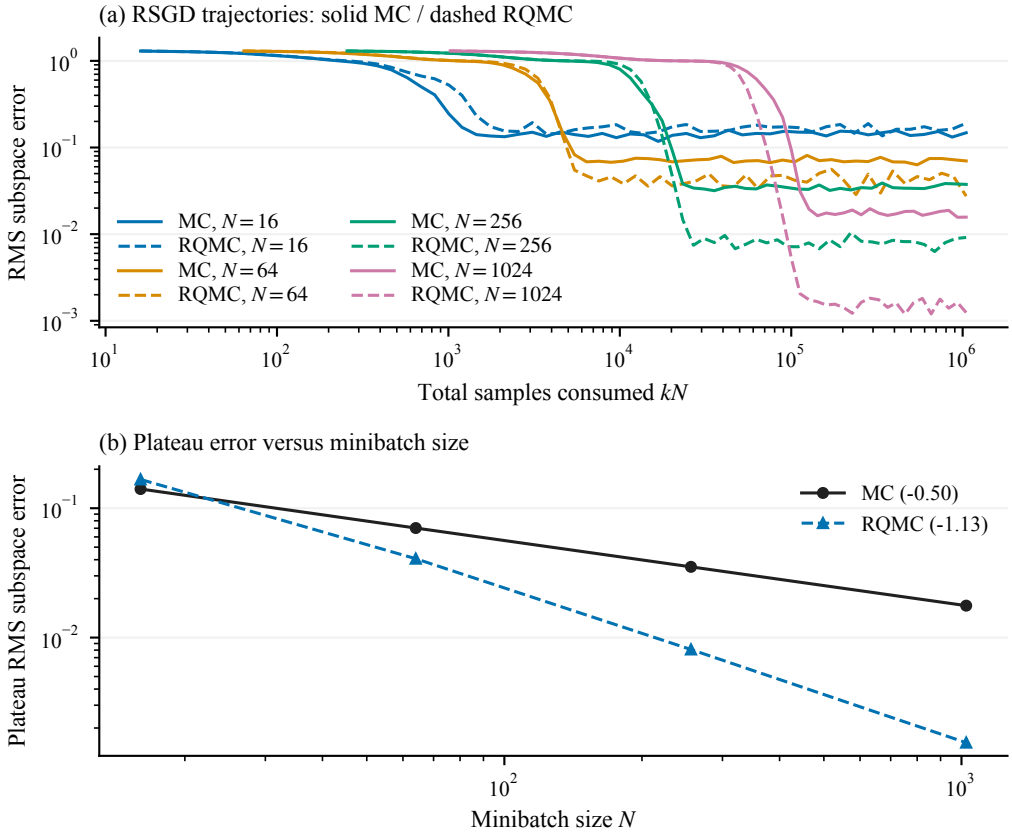

Figure 11: Riemannian SGD on $\mathrm{Gr}(8, 2)$ with i.i.d. (solid) and scrambled-Sobol' (dashed) minibatches of size N, constant step $\eta = 0.05$. (a) Subspace error versus samples consumed. (b) Plateau error versus N, with fitted slopes.

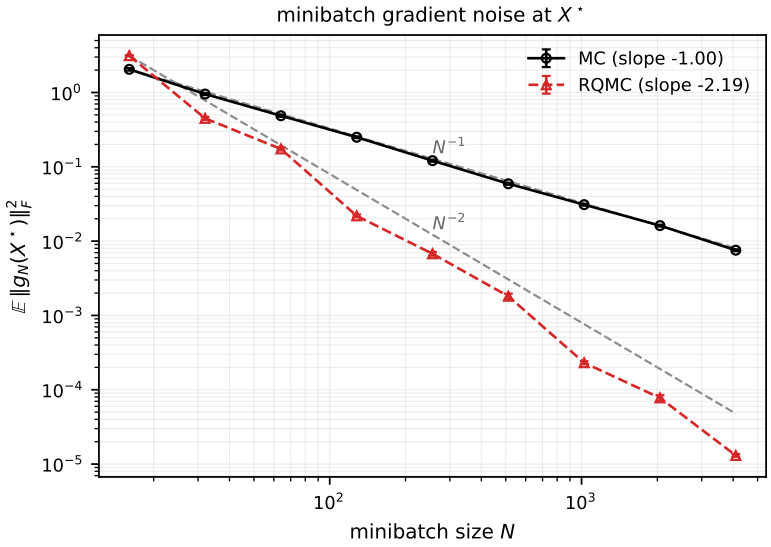

Figure 12: Mean squared error $\mathbb{E}\,\|\widehat{g}_N(X^\star)\|^2$ of the minibatch gradient at the optimum for i.i.d. and scrambled-Sobol' minibatches, 200 replicates per N.

G.2 The Grassmannian with a bounded transport: same objective, regular integrand

Setup. The Gaussian transport of Appendix G.1 is replaced by $a(u) = \sqrt{3}\,W(2u - \mathbf{1})$, $u \in [0,1]^8$, with the same Q, Σ, W, $X^\star$ and initial point. Since $\mathbb{E}[(2u - \mathbf{1})(2u - \mathbf{1})^\top] = \frac{1}{3}I$, the population covariance, objective and minimizer are unchanged. The data law is not: tails and higher moments differ, so the comparison controls the optimization problem rather than every property of the stochastic integrand; at $X^\star$, where the gradient integrand involves only the cross covariances between the target eigenspace and its complement, it is cleanest. The transport is bounded and C^∞ on the closed cube, the gradient integrand $-(I - XX^\top)a(u)a(u)^\top X$ is a polynomial in u, and Assumptions C.2 and C.3 hold on all of $\mathrm{Gr}(8,2)$; the mixed smoothness of (S3) holds at every X, so the $N^{-3}(\log N)^7$ variance bound applies to the ideal scrambling (Lemma B.4) and, up to an additive $\beta_{30} \le 3.6 \cdot 10^{-7}$ in root mean square, to the implemented one (Lemma B.5). The routes of the first revision are run by `code/exp_grassmann_bounded.py`: the gradient-noise measurement at $X^\star$ (200 minibatches per size, $N = 2^4, \dots, 2^{12}$); RSGD with fresh i.i.d. or scrambled minibatches ($\eta = 0.05$, $N \in \{16, 64, 256, 1024, 4096\}$, 2^{20} samples per run, 10 runs); and the reused-sample route, Riemannian gradient descent with the QR retraction on the sample objective of one point set to $\|\mathrm{grad}\, f_N\|_F < 10^{-10}$. `code/exp_grassmann_controls.py` adds 64 scramblings and 64 i.i.d. draws of the reused route ($N = 2^4, \dots, 2^{15}$), the matched comparison of Figure 3(c), and a smooth non-polynomial control.

Gradient noise. The mean squared error of the minibatch gradient at $X^\star$ has slope -1.00 for MC and -2.92 for scrambled Sobol' (bootstrap 95% interval $[-3.04, -2.84]$; sliding five-point windows between -2.5 and -3.3), against -2.19 for the Gaussian transport in Figure 12; the squared error of the deterministic Halton set has slope -1.74. At $N = 4096$ the MC/RQMC ratio of mean squared errors is $3.2 \cdot 10^4$, against 580 for the Gaussian transport. At $N = 16$ the scrambled net is worse than MC (ratio 0.6). A 16-point net in eight dimensions still stratifies each coordinate into 16 intervals; what it does not reduce at this size is the variance of the pairwise interaction terms of which this integrand consists.

Returned iterates. Theorems C.9 and C.15 bound the iterate an algorithm returns, so the primary comparison uses implementable outputs (Figure 3(c)). The fresh routes run RSGD for 2^{20} samples and return the last iterate; the reused routes solve the sample objective of one set of N points and return the final iterate of Riemannian gradient descent, which agrees with the sample eigenspace to

$3 \cdot 10^{-10}$ (24–370 iterations for MC sets, 30–73 for scrambled and Halton sets). At $N = 4096$ the population gradient norms of the returned iterates are $2.5 \cdot 10^{-2}$ (fresh MC), $8.7 \cdot 10^{-2}$ (reused MC), $2.5 \cdot 10^{-4}$ (fresh scrambled), $6.7 \cdot 10^{-4}$ (reused scrambled, root mean square over 64 scramblings) and $4.1 \cdot 10^{-3}$ (reused Halton): the choice between independent and scrambled sampling changes the accuracy by two orders of magnitude, the choice between fresh and reused sets by a factor of about three. Over the sizes tested the fresh routes decay like $N^{-0.52}$ (MC) and $N^{-1.40}$ (scrambled), the reused routes like $N^{-0.49}$ (MC) and $N^{-1.02}$ (Halton); the reused scrambled route has the heavy-tailed distribution described below and is summarized by its root mean square and median rather than a slope. The RSGD plateaus of the last quarter of each run have slopes -0.50 for MC and -1.47 for scrambled nets (bootstrap intervals over runs $[-0.51, -0.49]$ and $[-1.50, -1.45]$), matching $-a/2$ for the measured MSE exponent $a = 2.92$; the subspace-error plateaus have slopes -0.50 and -1.48 (Figure 13(a)).

Budgets and cost. Figure 15(a) plots accuracy against the number kN of component gradients. Its lines are *oracle-tuned lower envelopes*: for each route, the smallest root-mean-square population gradient norm over all checkpoints and all minibatch sizes up to that budget, a quantity that uses hindsight and the known population objective. Over budgets $\geq 2^{12}$ their slopes are -0.48 (MC-RSGD), -1.54 (scrambled RSGD) and -0.96 (Halton, reused). With a constant step, RSGD reaches the noise floor of a minibatch size after a number of iterations that does not depend on N (the transient occupies at most 57% of a run), so an envelope slope reflects the decay of the estimator's error in N; it is not a measurement of the worst-case exponents of Table 4. The markers are the implementable outputs of the reused routes, placed at N times their number of iterations. The count kN is the accounting of Section 2, not the cost of this implementation. The sample objective depends on the data only through $S_N = N^{-1} \sum_i a_i a_i^\top$, which the reused routes form once before iterating on 8×2 matrices, and a direct eigensolver computes the same minimizer; Table 6 separates the costs. The times are indicative, and the problem isolates integration behavior rather than benchmarking eigenspace solvers.

Table 6: Cost of the Grassmann routes, separated. Reused routes: one set of $N = 4096$ bounded samples, median time over 64 sets in one process. Fresh routes: 2^{20} samples in minibatches of $N = 256$, median time per run (runs on nine processes). The component-gradient count is the accounting of Section 2.

Route	samples	covariance assembly	iterations	wall time
reused set, Riemannian gradient descent	4096	once	47	2.9 ms
reused set, direct eigensolver	4096	once	–	0.04 ms
fresh MC minibatches (RSGD)	2^{20}	every step	4096	0.3 s
fresh scrambled minibatches (RSGD)	2^{20}	every step	4096	1.7 s
component gradients, accounting	reused: $4096 \times 47 \approx 1.9 \cdot 10^5$; fresh: $2^{20} \approx 1.0 \cdot 10^6$			

One fixed point set over 64 scramblings. The reused scrambled route is the regime of Theorem C.13(b), locally: f is not geodesically convex on $\mathrm{Gr}(8, 2)$ and the growth constant is the local eigengap. Figure 13(b) and Figure 14 record it over 64 independent scramblings. Its root-mean-square subspace error decays like $N^{-1.50}$ over the six largest N (bootstrap interval over scramblings $[-2.51, -1.31]$), the order $N^{-3/2}$ of the theorem, but its distribution is far from Gaussian. The median drops from $9.0 \cdot 10^{-8}$ at $N = 2^{12}$ to $5.2 \cdot 10^{-11}$ at 2^{13} and $3.9 \cdot 10^{-15}$ at 2^{14}, and the fraction of scramblings with subspace error below 10^{-10} rises from 0.28 at 2^{12} to 0.53, 0.75 and 0.86 at 2^{13}, 2^{14} and 2^{15}, while the root mean square stays near $7.5 \cdot 10^{-5}$ at 2^{13} and 2^{14}, carried by the remaining scramblings. A few excellent realizations therefore say little about the root-mean-square error, as the distinction between equal variances and equal error distributions predicts (Goda & Suzuki, 2026).

The mechanism is a cancellation in one block of the covariance error. Because $W = Q\,\mathrm{diag}(\sqrt{\lambda})$, the matrix $E = Q^\top (S_N - \Sigma) Q$ equals $3\,\mathrm{diag}(\sqrt{\lambda})\,(M_N - \frac{1}{3} I)\,\mathrm{diag}(\sqrt{\lambda})$, where M_N is the sample second-moment matrix of $2u - \mathbf{1}$; the cube coordinates index the eigendirections. At $X^\star$ the sample gradient sees only the cross block, $\|\mathrm{grad}\, f_N(X^\star)\|_F = \|E_{21}\|_F$ with E_{21} the rows 3–8 and columns 1–2 of E, and the dominant sample eigenspace moves by about $\|E_{21}\|_F / (\lambda_2 - \lambda_3)$ as long as the eigenvalue order is preserved (the sample eigengap stays within 0.005 of 2 for $N \geq 2^8$). The entries

of M_N are averages of products of the binary digits of two coordinates, each product weighted by $2^{-(k+l)}$ for digit positions k and l. A scrambled digital net averages such a product to zero unless the scrambled generator rows of the two positions coincide on the net, which is impossible when $k + l \le m - t$ for a two-dimensional projection with quality parameter t and depends on the scrambling otherwise. Each entry is therefore exact up to a few collision terms, and which entries carry them varies with the scrambling. For seed 0 at $N = 2^{13}$ the only large collision affects the moment of coordinates 4 and 6: the covariance error is $4.1 \cdot 10^{-4}$ but lies in the complementary diagonal block, the cross block is $1.6 \cdot 10^{-11}$, and the subspace error is $6.9 \cdot 10^{-12}$. For seed 1 the collision falls in the cross block, and the subspace error is $5.8 \cdot 10^{-5}$ (Table 7). The full covariance cancels less often than the cross block (fractions 0.16 against 0.48 at $N = 2^{13}$), and the unscrambled net keeps a covariance error of $2.3 \cdot 10^{-4}$ for $N \ge 2^{14}$. The cancellation is a property of the net rather than of the 30-bit truncation: under the ideal scrambling coupled to the same first 30 digits, the fraction of scramblings with a cross block below 10^{-10} changes from 0.48 to 0.47 at $N = 2^{13}$ and from 0.84 to 0.83 at 2^{15} (Appendix G.9).

Table 7: One fixed scrambled Sobol' net on the bounded Grassmann model: covariance error, cross block (which equals the sample gradient norm at $X^\star$) and subspace error of the sample minimizer, for scrambling seeds 0–3 and for the unscrambled net.

N	seed 0	seed 1	seed 2	seed 3	unscrambled
covariance error $\|S_N - \Sigma\|_F$					
2^{12}	$4.1 \cdot 10^{-4}$	$1.9 \cdot 10^{-4}$	$9.7 \cdot 10^{-7}$	$1.3 \cdot 10^{-5}$	$5.9 \cdot 10^{-4}$
2^{13}	$4.1 \cdot 10^{-4}$	$1.9 \cdot 10^{-4}$	$8.3 \cdot 10^{-7}$	$1.3 \cdot 10^{-5}$	$3.5 \cdot 10^{-4}$
2^{14}	$3.7 \cdot 10^{-12}$	$1.3 \cdot 10^{-11}$	$< 10^{-13}$	$< 10^{-13}$	$2.3 \cdot 10^{-4}$
cross block $\|E_{21}\|_F = \|\mathrm{grad}\, f_N(X^\star)\|_F$					
2^{12}	$1.8 \cdot 10^{-5}$	$1.3 \cdot 10^{-4}$	$6.9 \cdot 10^{-7}$	$6.2 \cdot 10^{-7}$	$3.6 \cdot 10^{-4}$
2^{13}	$1.6 \cdot 10^{-11}$	$1.3 \cdot 10^{-4}$	$5.9 \cdot 10^{-7}$	$< 10^{-13}$	$1.4 \cdot 10^{-4}$
2^{14}	$2.6 \cdot 10^{-12}$	$9.0 \cdot 10^{-12}$	$< 10^{-13}$	$< 10^{-13}$	$2.0 \cdot 10^{-5}$
subspace error of $\hat{X}_N$					
2^{12}	$5.2 \cdot 10^{-6}$	$5.8 \cdot 10^{-5}$	$3.0 \cdot 10^{-7}$	$3.1 \cdot 10^{-7}$	$1.5 \cdot 10^{-4}$
2^{13}	$6.9 \cdot 10^{-12}$	$5.8 \cdot 10^{-5}$	$2.8 \cdot 10^{-7}$	$< 10^{-13}$	$5.0 \cdot 10^{-5}$
2^{14}	$8.4 \cdot 10^{-13}$	$4.3 \cdot 10^{-12}$	$< 10^{-13}$	$< 10^{-13}$	$5.9 \cdot 10^{-6}$

A smooth non-polynomial control. The transport $a(u) = W\xi(u)$ with $\xi_j(u) = \sqrt{2}\sin(2\pi u_j)$ has $\mathbb{E}\xi = 0$ and $\mathbb{E}\xi\xi^\top = I$, hence the same covariance, objective and minimizer, and a smooth gradient integrand that is not a polynomial in u. Its gradient mean squared error at $X^\star$ has slopes -0.99 (MC), -2.59 (scrambled Sobol') and -1.91 (Halton, squared error), with an MC/RQMC ratio of $5.6 \cdot 10^3$ at $N = 4096$ and 0.5 at $N = 16$ (Figure 15(b)). Over 64 scramblings the reused route has root-mean-square subspace slope -1.51 (bootstrap interval $[-1.68, -1.41]$), and no scrambling reaches 10^{-10} (Figure 14(b)). The gain of the scrambled net thus does not depend on the polynomial cancellation, which changes only the distribution of the fixed-set error; the noncommutative SPD model of Appendix G.3 is a second smooth non-polynomial setting.

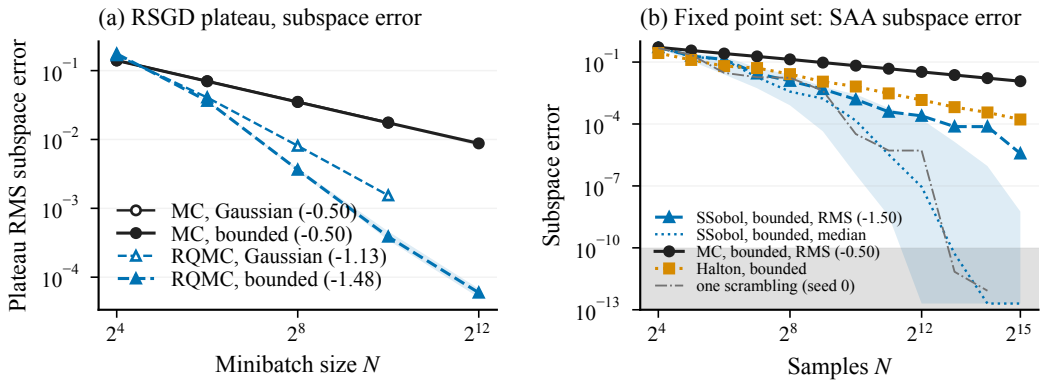

Figure 13: Bounded transport on $\mathrm{Gr}(8, 2)$. (a) RSGD plateau of the subspace error versus minibatch size, against the Gaussian transport of Appendix G.1 (hollow markers); shading is the 10th–90th percentile range of per-run values for the bounded model. (b) Subspace error of the sample minimizer for one fixed point set: root mean square and median over 64 scramblings with their 10th–90th percentile range, root mean square over 64 i.i.d. draws, the Halton set, and the single scrambling (seed 0) of the first revision. The grey band is below the solver tolerance.

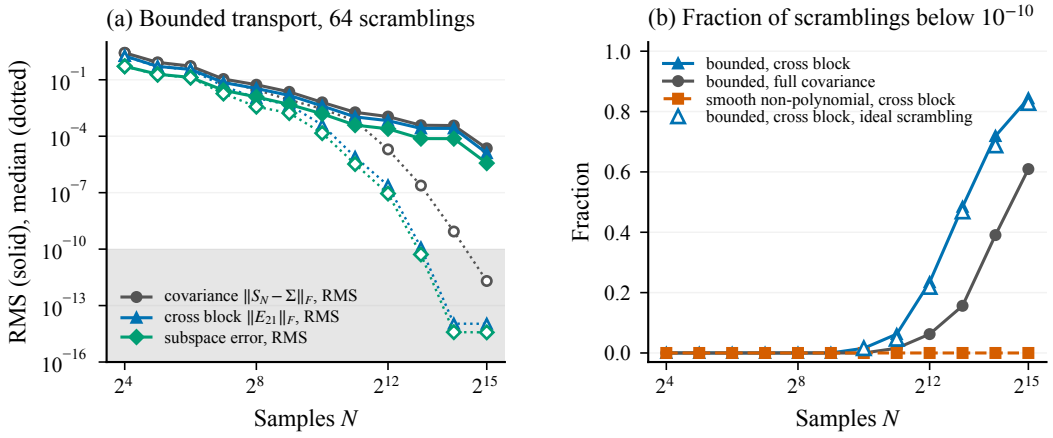

Figure 14: Where the fixed-net error goes. (a) Root mean square (solid) and median (dotted) over 64 scramblings of the covariance error, the cross block $\|E_{21}\|_F = \|\mathrm{grad}\, f_N(X^\star)\|_F$, and the subspace error. (b) Fraction of scramblings with value below 10^{-10}: cross block and full covariance for the bounded transport, cross block for the smooth non-polynomial transport, and cross block under the ideal scrambling coupled to the same first 30 digits (hollow markers).

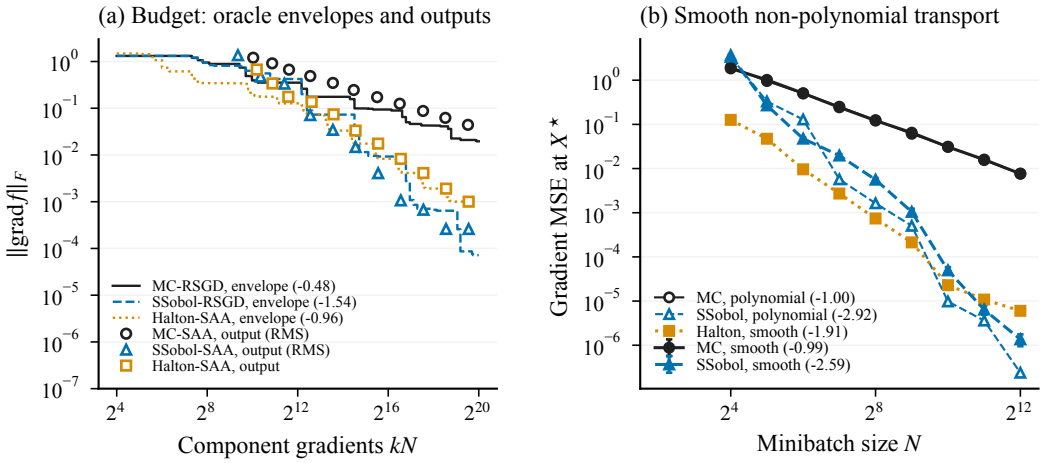

Figure 15: (a) Oracle-tuned lower envelopes over checkpoints and minibatch sizes (lines) and implementable outputs of the reused routes (markers, root mean square over 64 sets) against the component-gradient accounting kN; the envelopes use hindsight and the known population objective. (b) Gradient mean squared error at $X^\star$ for the smooth non-polynomial transport (filled) and the polynomial bounded transport (hollow).

G.3 Karcher means on the affine-invariant positive definite cone

Setup. $\mathcal{M} = \mathcal{P}_n$ with $n = 3$ and the affine-invariant metric of Example E.4, $s = 6 = \dim \mathrm{Sym}(3)$, $\{E_j\}$ the Frobenius-orthonormal basis of $\mathrm{Sym}(3)$, and $P = Q\,\mathrm{diag}(0.5, 1, 3)Q^\top$ for a fixed orthogonal Q. Two transports are used.

(S) *Symmetric log-Gaussian model:* $T(u) = P^{1/2}\exp\big(\tau\sum_j \Phi^{-1}(u_j)E_j\big)P^{1/2}$, $\tau = 0.4$. The law of $S(u) = \tau\sum_j \Phi^{-1}(u_j)E_j$ is invariant under $S \mapsto -S$, so the geodesic symmetry at P (an isometry) preserves ν, and by uniqueness of the Karcher mean on a Hadamard manifold $x^\star = P$ exactly. This model is outside Theorem E.3 (unbounded support, Remark E.5), and it is degenerate in an instructive way: at $x^\star = P$ the gradient integrand is $\mathrm{Log}_P(T(u)) = P^{1/2}S(u)P^{1/2}$, a *separable* function of u, so the gradient discrepancy at the mean only tests the one-dimensional projections of the point set.

(B) *Bounded non-commutative model:* with $w_j = 2u_j - 1$, $T(u) = P^{1/2}\exp\big(m_0 + \tau\sum_j w_jE_j + \beta\sum_j w_jw_{j+1}E_{j+3}\big)P^{1/2}$ (indices mod 6), $\tau = 0.5$, $\beta = 0.3$, and a fixed offset $m_0 = \sum_j c_jE_j$ with $c = (0.4, -0.2, 0.3, 0.5, 0, -0.4)$. The transport is C^∞ on the closed cube with values in a bounded ball, so the hypotheses of Theorem E.3 hold. Since $\mathrm{Log}_P(T(u)) = P^{1/2}S(u)P^{1/2}$ exactly, the mean equals P if and only if $\mathbb{E}S = 0$; the offset makes $\mathbb{E}S = m_0$, and because m_0 does not commute with the fluctuations the mean is neither P nor $P^{1/2}e^{m_0}P^{1/2}$ (the reference below is at distance 0.82 from the former and 0.018 from the latter). At $x^\star$ the integrand $u \mapsto \mathrm{Log}_{x^\star}(T(u))$ is the matrix logarithm of a product of non-commuting factors, a genuinely six-dimensional smooth function. The population mean is computed as the Karcher mean of three sample means from independently scrambled Sobol' nets of 2^{20} points; the three estimates lie within $7\cdot 10^{-9}$ of their mean, two orders of magnitude below the smallest error reported.

For each sampler (i.i.d. uniform, Halton with the origin dropped, scrambled Sobol') and $N = 2^m$, $m = 4, \dots, 14$, we record the gradient discrepancy $G_N = \|\mathrm{grad}\, f_N(x^\star)\|_{x^\star} = \big\|\frac{1}{N}\sum_i \mathrm{Log}_{x^\star}(y_i)\big\|$ (the population gradient vanishes at $x^\star$) and the error $d(\hat{x}_N, x^\star)$ of the sample Karcher mean, computed by the fixed-point iteration of Example E.4 to residual 10^{-13} (7–12 iterations in all cases). Random samplers use 20 replicates and report root-mean-square errors; slopes are least-squares fits over the six largest N.

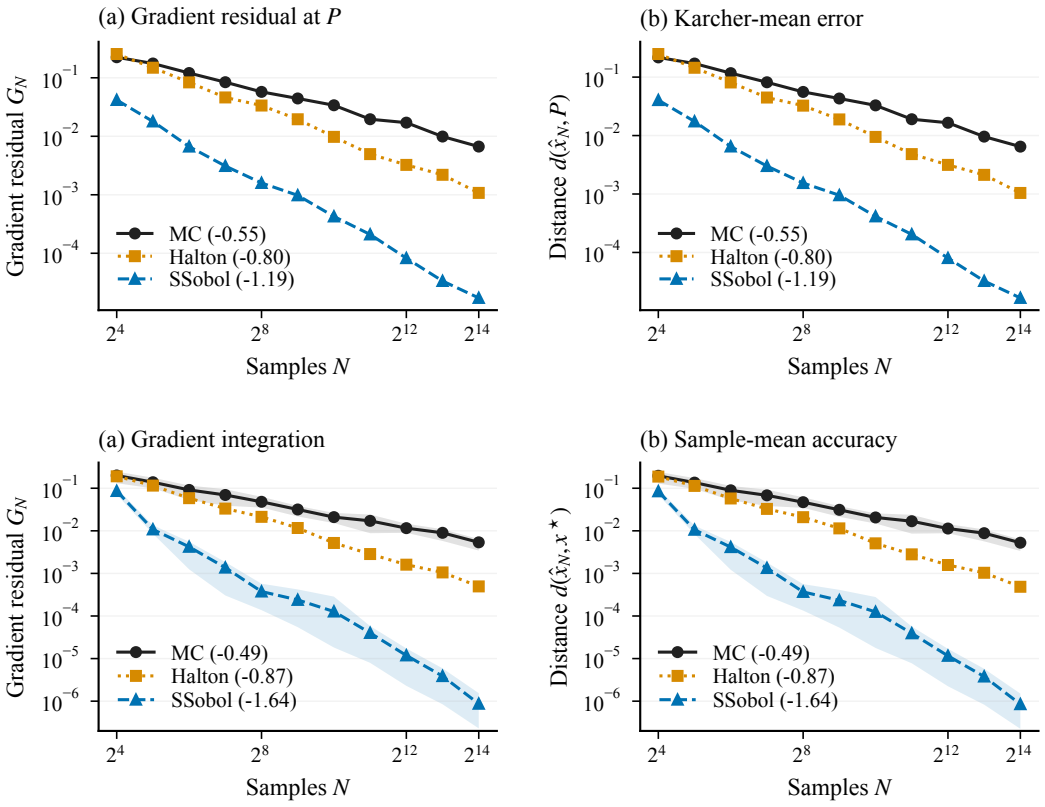

Figure 16: Karcher means on $\mathcal{P}_n$, $n = 3$. Top: symmetric log-Gaussian model (S). Bottom: bounded non-commutative model (B), inside the hypotheses of Theorem E.3. In the lower pair, shading shows the 10th–90th percentile range of individual replicate errors. Left: gradient discrepancy at the population mean; right: distance of the sample Karcher mean to the population mean. Root-mean-square over 20 replicates for the random samplers; fitted slopes over the six largest N in the legends.

Results. Figure 16 shows both models. In model (S), i.i.d. sampling follows $N^{-1/2}$ (the RMS of G_N is $\tau\sqrt{6/N}$ in closed form, $7.7 \cdot 10^{-3}$ at $N = 2^{14}$ against a measured $6.7 \cdot 10^{-3}$, a ratio 0.87 that lies within the scatter of a 20-replicate RMS), Halton decays like $N^{-0.80}$ and scrambled Sobol' like $N^{-1.1\pm0.1}$ over the range, with the local slope still steepening at $N = 2^{14}$; at $N = 2^{14}$ the Karcher-mean errors are $6.5 \cdot 10^{-3}$, $1.0 \cdot 10^{-3}$ and $1.7 \cdot 10^{-5}$ respectively. The shortfall from the N^{-1} and $N^{-3/2}$ rates of the bounded theory is the expected effect of the unbounded transformation Φ^{-1} (Remark E.5). The Karcher-mean error tracks the gradient discrepancy at $x^\star$ almost exactly (ratio between 0.95 and 1.0 in every replicate), consistent with the strong-convexity mechanism: with $\mu = 1$ and Hess $f_N \approx I$ near $x^\star$, $d(\hat{x}_N, x^\star) \approx \|\mathrm{grad}\, f_N(x^\star)\|$.

Model (B) is the test of Theorem E.3 proper. The slopes are -0.49 (MC; bootstrap interval over the 20 replicates $[-0.53, -0.45]$), -0.87 (Halton) and -1.64 (scrambled Sobol'; $[-1.78, -1.50]$). For this smooth bounded integrand and one fixed net, Theorem C.13(b) with (S3) bounds the root-mean-square error by order $N^{-3/2}(\log N)^{5/2}$, for the ideal scrambling (Lemma B.4) and, up to an additive β_{30}, for the implemented one (Lemma B.5); the paired comparison of Appendix G.9 changes the root-mean-square error at $N = 2^{14}$ by less than 10^{-11}. The theorem is an upper bound, not an asymptotic expansion, so the faster finite-range fit does not contradict it; the local slope is still moving at $N = 2^{14}$, and the fitted value does not establish an asymptotic exponent. At $N = 2^{14}$ the Karcher-mean errors are $5.3 \cdot 10^{-3}$, $4.9 \cdot 10^{-4}$ and $8.5 \cdot 10^{-7}$: a factor of about 6,200 between i.i.d. sampling and the scrambled net at equal sample size. The Halton slope is shallower than -1, as is usual for unscrambled Halton points at these sample sizes in six dimensions; the Koksma–Hlawka bound $N^{-1}(\log N)^6$ is not even monotone for $N < e^6$ and is quantitatively uninformative in this range, so the measured exponent, not a sharp finite-sample prediction from that bound, is what the experiment reports. The ratio of Karcher-mean error to gradient discrepancy at $x^\star$ is between 0.977 and 0.984 in every cell of the sweep, so the two panels of Figure 16 display the same empirical decay through the gradient-to-distance relationship, and the sample Karcher mean is as accurate as the QMC estimate of the gradient at the population mean.

G.4 The sphere: Stolarsky nodes, transported nets and the cut locus

Nodes from Riemannian optimization (Appendix F). For $N = 2^4, \ldots, 2^{12}$ we ran Riemannian gradient ascent on the sum of distances (37) over $(\mathbb{S}^2)^N$ (Proposition F.1; normalization retraction, Armijo backtracking, at most 1000 iterations), initialized at the Lambert image of a scrambled Sobol' net, and recorded the deficit $\delta(X_N) = \frac{4}{3} - \frac{1}{N^2}\sum_{i,j}\|x_i - x_j\|$, which is $C_2 D_{L_2}(X_N)^2$ by (14). Figure 17(a) compares i.i.d. uniform points (for which $\mathbb{E}\,\delta = \frac{4}{3N}$ exactly; the measured RMS tracks this within replicate noise, slope -0.96), Lambert images of scrambled Sobol' and Halton points, the spherical Fibonacci spiral (Marques et al., 2013), and the optimized sets. The structured constructions have finite-range behavior near the optimal deficit order $N^{-3/2}$ of Beck (1984) (slopes -1.50, -1.45, -1.50, -1.50, with sliding-window slopes constant to ±0.01), corresponding to observed cap-discrepancy behavior near $N^{-3/4}$. Riemannian ascent reduces the deficit to 0.54–0.70 of its initialization (mean 0.60) across the tested N (panels (b), (c) show $N = 256$) but does not change the exponent, and its constant coincides with that of the Fibonacci spiral to within 3.4%; for $N \geq 64$ the runs stop at the iteration cap with gradient norms of order 10^{-4}, so the reported deficits are upper bounds; the deficit changed by less than 1% over the second half of each run. The lesson of the reverse direction is thus quantitative: Stolarsky's principle turns discrepancy into a smooth objective, the ascent lowers the attained discrepancy of its initialization, and (14) evaluates that discrepancy exactly, as it does for any other node set. The observed deficit approaches the Fibonacci constant over the tested sizes; this does not certify global optimality or an asymptotic rate for the numerical sequence.

Fréchet means under a von Mises–Fisher law. Let ν be the von Mises–Fisher law (Mardia & Jupp, 2000) with concentration $\varkappa = 2$ about e_3, with density $h(y) = \varkappa e^{\varkappa y_3}/\sinh\varkappa$ with respect to σ_2; its Fréchet mean is e_3, and it is unique. Indeed, let x, x' lie on a meridian with x' closer to e_3, and let ς be the reflection across the plane bisecting x and x', with unit normal n pointing towards x', so that $\langle n, e_3\rangle > 0$. On the open hemisphere $H' = \{\langle y, n\rangle > 0\}$ one has $d(x, y) > d(x', y)$, while $\langle \varsigma y, e_3\rangle = \langle y, e_3\rangle - 2\langle y, n\rangle\langle n, e_3\rangle < \langle y, e_3\rangle$ gives $h(\varsigma y) < h(y)$; pairing $y \in H'$ with ςy,

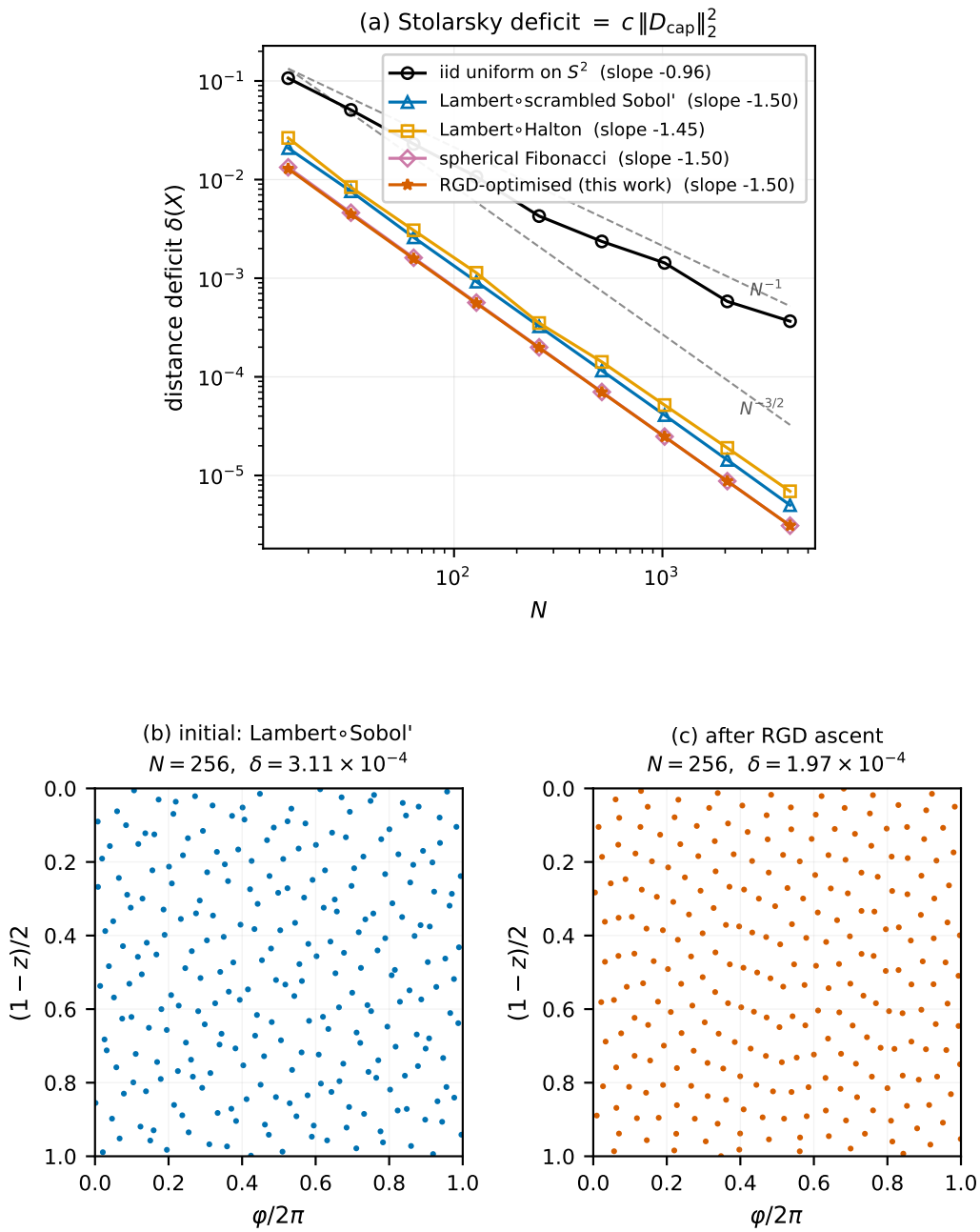

Figure 17: Riemannian optimization constructs QMC nodes on $\mathbb{S}^2$. (a) Distance deficit $\delta(X_N) \propto D_{L_2}(X_N)^2$ for i.i.d. points, Lambert-transported nets, the Fibonacci spiral and the sets obtained by Riemannian gradient ascent on the sum of distances; reference slopes -1 and $-3/2$. (b), (c) The $N = 256$ initialization and the optimized configuration in Lambert coordinates.

and using $d(x, \varsigma y) = d(x', y)$ and $d(x', \varsigma y) = d(x, y)$,

$$f(x) - f(x') \;=\; \frac{1}{2}\int_{H'} \big(d(x,y)^2 - d(x',y)^2\big)\big(h(y) - h(\varsigma y)\big)\,\mathrm{d}\sigma_2(y) \;>\; 0.$$

Thus $f(x_a)$ increases strictly in the polar angle a (quadrature gives $f(e_3) = 0.5684$ and $f(-e_3) = 2.5856$), and e_3 is the unique minimizer; the argument applies to every density that increases strictly with y_3. The support is the whole sphere, so the antipode of every x belongs to the support (the law has no point atoms) and Theorem E.6 does not apply; this is the setting of Appendices E.4 and E.5. Six estimators of the Fréchet mean were compared for $N = 2^4, \dots, 2^{14}$: transported cube samplers $y_i = T_z(u_i)$ with the inverse-distribution profile $z(t) = 1 + \varkappa^{-1}\log(t + (1-t)e^{-2\varkappa})$ of (32) and u_i i.i.d., Halton, or scrambled Sobol'; and intrinsic nodes y_i of uniform type (Lambert images of scrambled Sobol' nets, the Fibonacci spiral, and the Stolarsky-optimized sets for $N \le 2^{12}$) with normalized weights $w_i \propto h(y_i)$, that is, the weighted rule of (36). In every case the sample mean is computed by the weighted intrinsic fixed-point iteration $x \leftarrow \mathrm{Exp}_x(\sum_i w_i \,\mathrm{Log}_x(y_i))$ to residual 10^{-14}, and we record the gradient discrepancy $G_N = \big\|\sum_i w_i \,\mathrm{Log}_{e_3}(y_i)\big\|$ at the population mean and the error $d(\hat{x}_N, e_3)$.

Figure 18 shows the outcome. Transported i.i.d. sampling has slope -0.48; the transported scrambled net reaches $9.4 \cdot 10^{-6}$ at $N = 2^{14}$ against $9.4 \cdot 10^{-3}$ for i.i.d. sampling, with a local slope that steepens from -1.1 to -1.5 across the range; the weighted Lambert–Sobol' nodes behave almost identically (slope -1.39); the deterministic Fibonacci and Stolarsky nodes are the most accurate at small N (the optimized set attains $G_N = 6.6 \cdot 10^{-5}$ at $N = 32$ in its stored orientation, a value that the rotation test of Appendix G.5 identifies as an orientation coincidence) but, being single configurations, follow non-monotone curves whose fitted slopes (-1.08 and -1.13) should not be read as exponents. Two structural facts stand out. First, the Fréchet-mean error is 1.35 ± 0.08 times the gradient discrepancy for every estimator and every N, and $1/\lambda = 1.3506$ where $\lambda = 0.7404$ is the Hessian eigenvalue of f at e_3 computed by quadrature: the observed proportionality is consistent with a local Hessian linearization. It is not an equality assertion in Theorem C.12, whose small-region hypotheses are distinct from this full-support experiment. Second, the intrinsic weighted nodes, for which the Sobolev certificate yields the gradient bound $C_h D_{L_2}^{4/5} = O(N^{-3/5})$ of Corollary D.7 rather than a linear one, integrate the gradient at the population mean at a rate between N^{-1} and $N^{-3/2}$ in practice; Appendix G.5 examines the maximal errors over x that the proposition bounds.

The cut locus, measured. Proposition E.10 predicts that the transported integrand $u \mapsto \mathrm{Log}_x(T_z(u))$ has finite Hardy–Krause variation when $x = e_3$, because the chart blows up the antipode $-e_3$ into the whole edge $u_1 = 0$, on which the direction is a smooth function of u_2, whereas for x off the pole the antipode is the image of an interior point of the square at which the integrand has a point singularity of order 0, infinite variation, and error at most $D^*(1 + \log(1/D^*))$ for every point set. Figure 19 measures the QMC error of $\frac{1}{N}\sum_i \mathrm{Log}_x(T_z(u_i))$ against a high-precision reference at $x = e_3$ and at $x_a = (\sin 1, 0, \cos 1)$ (antipode at $u \approx (0.028, 0.5)$). Monte Carlo is indifferent to the position of the cut locus (-0.48 and -0.52). The scrambled net is not: at e_3 its local slope steepens to -1.5, while at x_a the slope is -1.00 with sliding-window values between -0.94 and -1.06 across the whole range, a stable fitted slope over this finite range. For this interior point singularity the measured scrambled-net error is close to N^{-1} over the tested sizes, the order, up to a logarithm, that Proposition E.10(b) proves for every realization of the net, although the variation is infinite; the smooth-integrand rate $N^{-3/2}$ is lost. Appendix G.6 repeats the comparison with the chart as the only variable. The codimension-two singularity is mild, as anticipated, and far from the $N^{-1/2-1/(4s-2)+\epsilon}$ regime of hypersurface discontinuities (He & Wang, 2015). The favorable gradient estimate at the population mean is consistent with the fast local behavior of the transported estimator in Figure 18. It does not establish a uniform gradient bound over the sphere or a global convergence guarantee.

G.5 The optimized nodes: orientation, maximal errors and cost

Figure 18 compares single configurations of the deterministic node sets in one orientation relative to the target density. The cap discrepancy is invariant under rotations $R \in SO(3)$, but the weighted Fréchet estimator is not, because the density h is fixed while the nodes move; a good value can therefore reflect a favourable orientation rather than integration quality. The script

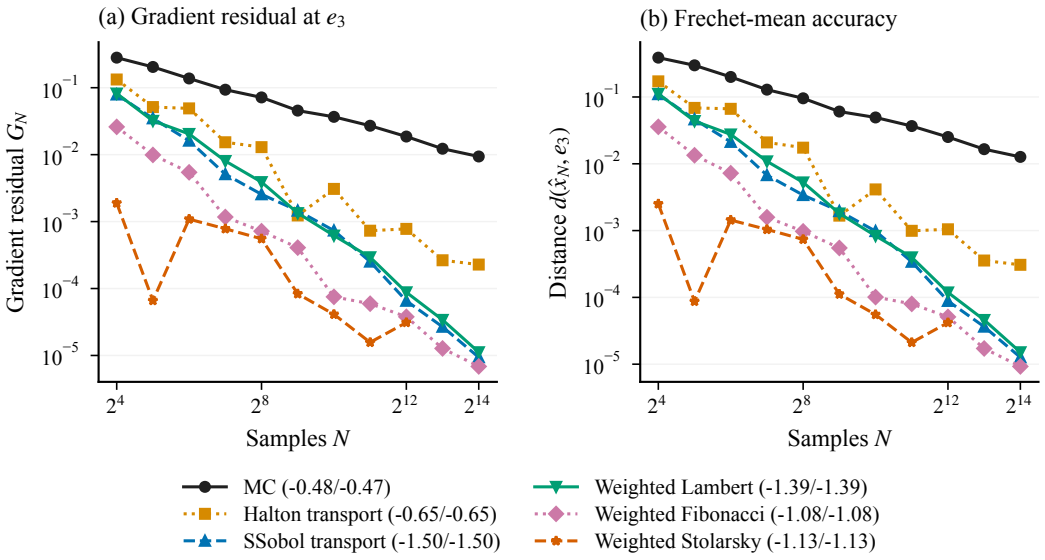

Figure 18: Fréchet mean of the von Mises–Fisher law ($\varkappa = 2$) on $\mathbb{S}^2$. (a) Gradient discrepancy at the population mean e_3; (b) error of the sample Fréchet mean. Transported samplers (i.i.d., Halton, scrambled Sobol' through the inverse-distribution map) and weighted intrinsic nodes (Lambert–Sobol', Fibonacci, Stolarsky-optimized). RMS over 20 replicates for the random estimators. Legend pairs give fitted gradient/distance slopes over the six largest N; slopes for single deterministic configurations describe the plotted range only.

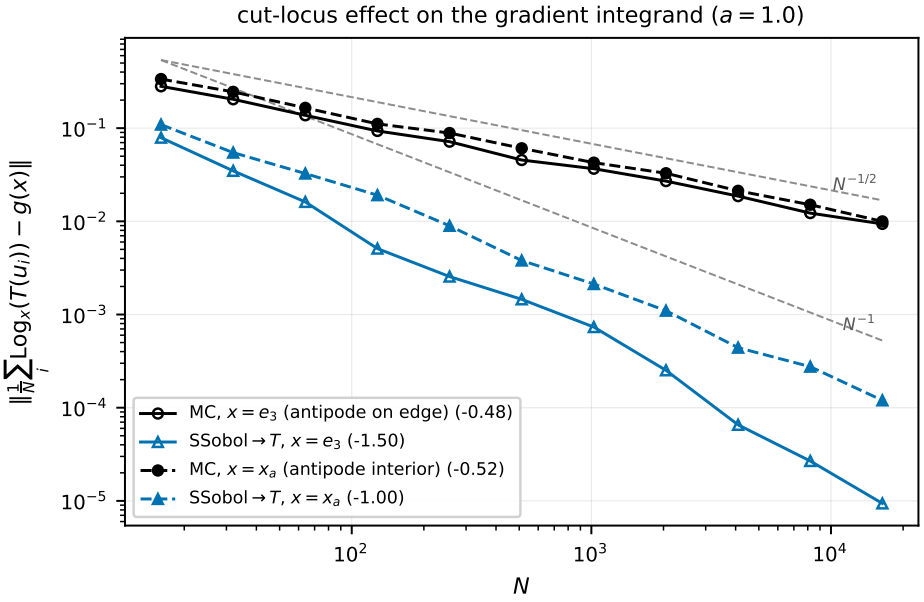

Figure 19: Error of the transported QMC estimate of the gradient integrand $\int \mathrm{Log}_x(y)\,\mathrm{d}\nu(y)$ at the pole $x = e_3$ (antipode on the edge $u_1 = 0$) and at x_a (antipode at an interior point of the square), for i.i.d. and scrambled Sobol' points; RMS over 20 replicates.

`code/exp_sphere_rotation.py` separates the two, and `code/exp_sphere_maxima.py` examines the maximal errors over x that Corollary D.7 bounds.

Rotation test. For each of the three intrinsic families (the Lambert image of the seed-0 scrambled Sobol' net, which is the ascent's initialization; the Fibonacci spiral; the optimized sets) and $N = 2^4, \dots, 2^{12}$, we apply 50 Haar-distributed rotations, recompute the weights $w_i \propto h(Ry_i)$, and record the gradient residual at e_3 and the Fréchet-mean error. The deficit of every rotated set agrees with the stored one to $4 \cdot 10^{-16}$, as it must. Figure 20(a,b) shows the root-mean-square over rotations with 10th–90th percentile bands and the stored orientation as a hollow marker. Three facts emerge. First, the stored value $G_N = 6.6 \cdot 10^{-5}$ of the optimized set at $N = 32$ is an orientation effect: over the rotations its gradient residual has RMS $2.7 \cdot 10^{-3}$ and its mean error $3.5 \cdot 10^{-3}$ (range $2.7 \cdot 10^{-4}$ to $8.2 \cdot 10^{-3}$), and the stored orientation beats every one of the 50 rotations; at $N = 16$, 512 and 2048 the stored orientation also lies in the best quarter of its distribution and at $N = 1024$ in the best third, whereas at $N = 128$, 256 and 4096 it lies in the worst third. Second, the same 50 rotations are applied to all three families, so ratios of their root-mean-square errors are paired, and a bootstrap over rotations gives their uncertainty. The optimized sets have the smaller point estimate at every N. Relative to Fibonacci the ratio is 0.45–0.73 for $N \le 128$, with 95% intervals that exclude 1 (upper ends 0.57–0.92), and 0.83–0.97 for $N \ge 256$, with intervals that contain 1 (for instance $[0.66, 1.02]$ at $N = 256$ and $[0.77, 1.24]$ at $N = 1024$). Relative to their own initialization the ratio is 0.06–0.42, with intervals excluding 1 at every N. The optimized and Fibonacci deficits agree to within 3.4% and are 0.54–0.70 of the initialization's. A smaller discrepancy lowers an upper bound; the observed ordering agrees with it without being implied by it, and two optimal-order families with almost the same D_{L_2} differ by up to a factor two at small N. The rotation-averaged errors decay like $N^{-0.99}$ (optimized), $N^{-1.05}$ (Fibonacci) and $N^{-1.32}$ (initialization) over the six largest N, all faster than the upper rate $N^{-3/5}$ of Corollary D.8. Third, the Fibonacci spiral's stored orientation, with its polar axis along the mean direction, is among the worst for it (quantile 0.84–1.00 of its rotation distribution for $N \le 512$), so Figure 18 understated Fibonacci.

Sampled maxima and the antipodal floor. Corollary D.7 bounds $\sup_x |f_N^h(x) - f(x)|$ by $C_h D_{L_2}$ and $\sup_{x \notin -Y_N} \left\| \operatorname{grad} f_N^h(x) - \operatorname{grad} f(x) \right\|$ by $C_h D_{L_2}^{4/5}$ for the unnormalized weights $h(y_i)/N$, with a constant C_h that is not evaluated. A maximum over finitely many decision points $\mathcal{X}$ is a lower bound for the supremum, so the experiment compares decay rates and looks for the location of the largest errors; it does not certify either bound. Reference values come from a 3000×2048 Gauss–Legendre/midpoint rule in (z, φ): the axial symmetry of the law makes $\operatorname{grad} f(x)$ a scalar function of the polar angle times the meridian direction (the residual orthogonal to it is $4 \cdot 10^{-16}$), the rule agrees with a 1500×1024 rule to $1.1 \cdot 10^{-6}$ for the gradient and $1.6 \cdot 10^{-9}$ for the value, and a cubic spline in the polar angle reproduces direct quadrature at off-grid angles to $4 \cdot 10^{-9}$. The search combines three sets of decision points: a 91×48 grid (4274 points), grids of 181×96 and 361×192 points, and, for every node, the points $\operatorname{Exp}_{-y_i}(r v_k)$ at distances $r \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-6}\}$ from its antipode in 8 tangent directions v_k. Exact antipodes are never evaluated (points with $\langle x, y_i \rangle \le -1 + 10^{-15}$ are excluded, which removes none of the searched points), and the logarithm is computed as $\operatorname{atan2}(\|y - \langle x, y \rangle x\|, \langle x, y \rangle)$ times the unit tangent, which is accurate near the antipode.

Refining the grid from 4274 to 68,930 points raises the gradient maximum by at most 18%, and the near-antipode search raises it further, to a limit that it reaches by $r = 10^{-4}$: the largest directional limit $\max_i(\|R_i(-y_i)\| + \pi h(y_i)/N)$ of Proposition D.13, which is itself a lower bound for the supremum and agrees with the sampled maximum to within 2% for every family and N. For the optimized sets the sampled maximum exceeds the floor $\pi \max_i h(y_i)/N$ by at most 2% for $N \ge 256$; at $N = 4096$ it exceeds it by 19% for Fibonacci and by 123% for the initialization. The sampled maxima decay like $N^{-1.00}$, $N^{-0.99}$ and $N^{-0.95}$ for the optimized, Fibonacci and initialization sets. Its ratio to $D_{L_2}^{4/5}$ falls from about 8 at $N = 16$ to 0.9–1.6 at $N = 4096$. For these node sets, then, the largest gradient error sits at the antipodes of the nodes with the largest weights and is of the order N^{-1} of the floor, between the rigorous lower bound of Proposition D.13 and the upper rate $N^{-3/5}$ of Corollary D.7, at the rate that Corollary D.14 gives for node sets equidistributed to high order. The sampled value maximum decays like $N^{-1.48}$, $N^{-1.47}$ and $N^{-1.31}$ for the optimized, Fibonacci and initialization sets, and its ratio to D_{L_2} falls from 2.5–3.2 to 0.05–0.35. The value integrand has a cone at the antipode, a singularity of order $m = 1$, and Corollary D.14 gives values of order $N^{-3/2+\epsilon}$

for designs of high order, the rate observed for the optimized and Fibonacci sets. Over eight rotations per set the grid maxima of the optimized and Fibonacci sets change by less than a factor 1.2, and the normalized weights give the same maxima to three digits.

Cost. The ascent spends $n_{\mathcal{E}} N(N-1)/2 + (n_{\mathrm{it}}+1)N(N-1)$ pairwise-distance evaluations ($n_{\mathcal{E}}$ energy evaluations, n_{it} iterations, both stored): $4.6 \cdot 10^5$ at $N = 16$, $4.1 \cdot 10^7$ at $N = 128$ and $4.2 \cdot 10^{10}$ at $N = 4096$ (276 s in the stored run), against N operations for the Fibonacci spiral. Each reuse costs the Fréchet solve, 16–24 iterations of N logarithm maps, essentially the same for all families. Reaching a rotation-RMS accuracy of 10^{-2} requires 16 optimized, 32 Fibonacci or 256 initialization nodes; 10^{-3} requires 128, 256 and 512; 10^{-4} requires 2048 nodes for both the optimized and the Fibonacci sets and is not reached by the initialization within $N \le 4096$. The ascent lowers the attained discrepancy of its initialization; Stolarsky's identity evaluates the same quantity for the free spiral and for every other node set. Against the spiral the construction is not faster: the saving per reuse is a few thousand logarithm maps against a construction cost of 10^7–10^{10} pairwise distances, and the two families reach the finest accuracy at the same N. Its value is a controlled comparison between node sets of known and nearly equal discrepancy.

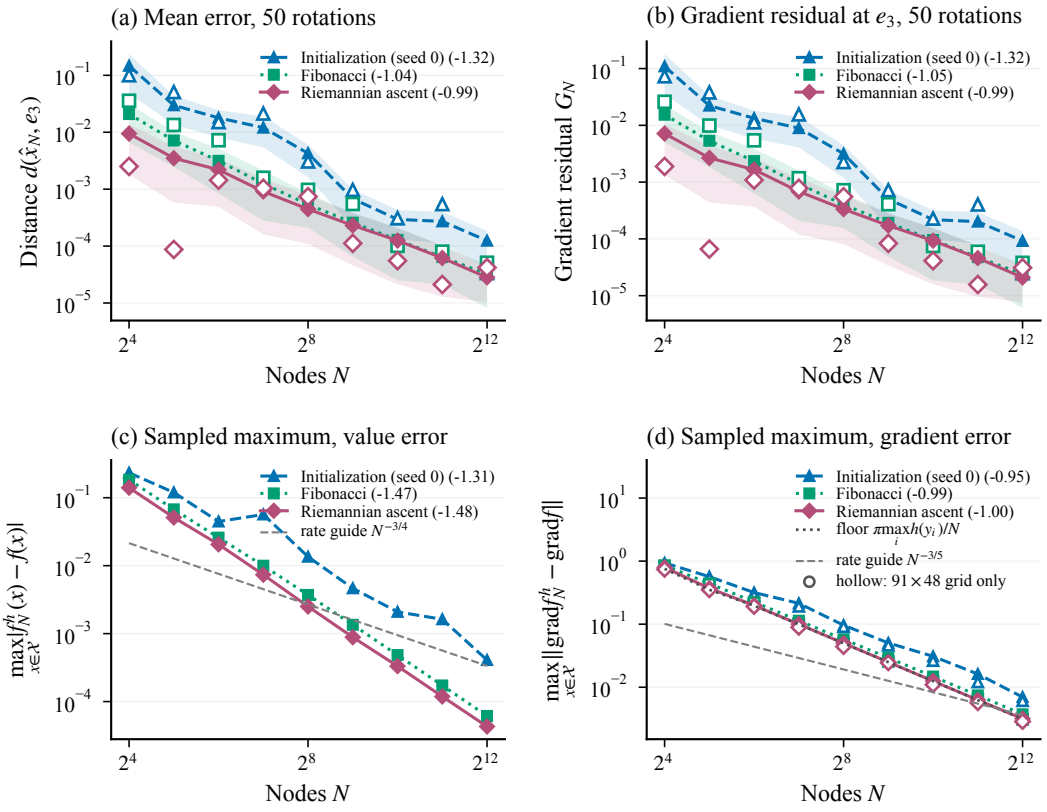

Figure 20: Orientation and maximal errors for the intrinsic node families under the von Mises–Fisher law ($\varkappa = 2$). (a) Fréchet-mean error and (b) gradient residual at e_3 over 50 random rotations of each node set: RMS (lines), 10th–90th percentile range (shading) and stored orientation (hollow markers); fitted slopes over the six largest N. (c) Sampled maximum over decision points of the value error and (d) of the gradient error, for the stored orientations: refined grid and near-antipode search (lines), the 4274-point grid of the first revision (hollow markers in (d)), rate guides $N^{-3/4}$ and $N^{-3/5}$ with arbitrary constants (dashed), and in (d) the floor $\pi \max_i h(y_i)/N$ of Proposition D.13 (dotted). A sampled maximum is a lower bound for the supremum.

G.6 THE CHART DECIDES: RESOLVED AND UNRESOLVED CUT POINTS

The script `code/exp_chart_resolution.py` tests Proposition E.10 and the one-point reading with the chart as the only variable. The Lambert chart $T_L(u) = (\sqrt{1-z^2}\cos 2\pi u_2, \sqrt{1-z^2}\sin 2\pi u_2, z)$, $z = 1-2u_1$, pushes the uniform law of the square to σ; it blows up e_3 into the edge $u_1 = 0$ and $-e_3$ into the edge $u_1 = 1$.

(A) Field test. For the uniform law every point is a critical point of the constant Fréchet functional and the field $e_N(x) = \frac{1}{N}\sum_i \mathrm{Log}_x(T_L(u_i))$ has target 0 at every x. We evaluate it at $x_a = (\sin a, 0, \cos a)$ for $a \in \{0, 0.5, 1, \pi/2\}$: at $a = 0$ the chart blows up $-x_a = -e_3$; for $a > 0$ the cut point $-x_a$ is the interior point $u_0 = ((1+\cos a)/2,\ 1/2)$. Samplers: 64 independent scramblings (seeds 0–63) and 64 i.i.d. sets per size, and the unscrambled Sobol' net shifted by half a cell so that no node lies on an edge; $N = 2^4, \ldots, 2^{17}$.

(B) Minimizer test. The von Mises–Fisher law with concentration $\varkappa$ and mean $x^\star = (\sin 1, 0, \cos 1)$ has Fréchet mean $x^\star$ and Hessian $\mathrm{Hess}\, f(x^\star) = \lambda I$, with $\lambda = \int_0^\pi \frac{1}{2}(1 + t\cot t)\frac{\varkappa e^{\varkappa\cos t}}{2\sinh\varkappa}\sin t\, \mathrm{d}t$, which is 0.2780 for $\varkappa = 0.5$ and 0.7404 for $\varkappa = 2$ (the latter agrees with the finite-difference value of Appendix G.4 to $2\cdot 10^{-6}$). Three charts carry the law: the Lambert chart about e_3 with importance weights $h(y) = \varkappa e^{\varkappa\langle x^\star, y\rangle}/\sinh\varkappa$ (the cut point $-x^\star$ is interior, at $u_0 \approx (0.77, 0.5)$); the same chart rotated by the rotation R with $Re_3 = x^\star$, with the same weights (the cut point is the edge $u_1 = 1$); and the von Mises–Fisher inverse-distribution chart about $x^\star$, unweighted (the cut point is the edge $u_1 = 0$). For each of 64 scramblings (or i.i.d. sets) per size we compute the field $e_N(x^\star) = \sum_i w_i \mathrm{Log}_{x^\star}(y_i)/\sum_i w_i$ and the weighted sample Fréchet mean by the fixed-point iteration of Appendix G.4 to a step below 10^{-14}.

Table 8: The chart experiment. Slopes are fitted over $N = 2^{12}, \ldots, 2^{17}$: of the RMS field for (A), and of the median Fréchet-mean error for (B). Ratio: median of $d(\hat{x}_N, x^\star)/\|e_N(x^\star)\|$ at $N = 2^{17}$, against $1/\lambda$.

(A) uniform law, Lambert chart about e_3, field at x_a

	$a = 0$ (pole)	$a = 0.5$	$a = 1$	$a = \pi/2$
SSobol, RMS slope	−1.76	−0.91	−0.99	−1.02
MC, RMS slope	−0.47	−0.46	−0.48	−0.50
unscrambled Sobol', slope	−2.69	−1.32	−0.85	−1.89

(B) von Mises–Fisher law with mean $x^\star$ at polar angle 1

$\varkappa$	chart	SSobol slope	MC slope	SSobol median error at 2^{17}	ratio $(1/\lambda)$
0.5	fixed (unresolved)	−1.01	−0.48	$2.8\cdot 10^{-5}$	3.579 (3.597)
0.5	rotated (resolved)	−1.55	−0.45	$4.5\cdot 10^{-7}$	3.591 (3.597)
0.5	vMF chart (resolved)	−1.54	−0.44	$3.9\cdot 10^{-7}$	3.584 (3.597)
2	fixed (unresolved)	−1.11	−0.52	$1.4\cdot 10^{-6}$	1.350 (1.351)
2	rotated (resolved)	−1.52	−0.45	$4.1\cdot 10^{-7}$	1.350 (1.351)
2	vMF chart (resolved)	−1.51	−0.43	$3.8\cdot 10^{-7}$	1.350 (1.351)

Results. Table 8 and Figure 4(a,b) summarize them. Monte Carlo is indifferent to the chart. Scrambled nets keep a rate near $N^{-3/2}$ or better exactly where the chart resolves the cut point: at the pole in (A), and in the rotated and von Mises–Fisher charts in (B); where the cut point is interior their rate drops to N^{-1}, the order, up to a logarithm, that Proposition E.10(b) proves as an upper bound for every realization. Neither rate is a theorem for scrambled nets: Proposition E.10(a) proves the order D_N^* in resolved charts, whose field integrand has mixed derivatives that are integrable but not square integrable near the blown-up edge ($\partial_1\partial_2$ of the direction is $2\pi z'/r$), outside (S3), and no variance lower bound is known in the unresolved chart. The unscrambled net is one deterministic sequence; its fitted slopes fluctuate between -0.85 and -2.69 because the errors of symmetric integrands partly cancel, and the slowest, -0.85 at $a = 1$, is the local slope of $N^{-1}\log^2 N$ over this range. In (B) the minimizer follows the field: the median ratio of error to field is within 0.5% of $1/\lambda$ in all six configurations, consistent with Remark C.8 on the draws where its hypotheses hold. Its upper tail is heavier in the resolving charts, where the chart places some nodes very close to $-x^\star$: a node at

distance δ from $-x^\star$ puts a concave kink of f_N at distance δ from $x^\star$ (Corollary D.8), the sample objective is not C^2 along the segment from $x^\star$ to $\hat{x}_N$, and the exact formula (20) does not apply to that draw; we therefore report medians. The singular part of the field in the fixed chart carries the factor $h(-x^\star)$, which is 0.58 for $\varkappa = 0.5$ and 0.075 for $\varkappa = 2$; for $\varkappa = 2$ the N^{-1} term is still comparable to the $N^{-3/2}$ term at $N = 2^{17}$, and the fixed chart shows the transitional slope -1.11 with a median error 3.4 and 3.6 times those of the rotated and the von Mises–Fisher chart at the largest size.

G.7 Rules exact to high degree sit at the floor

The script `code/exp_sphere_designs.py` tests Corollary D.14 with rules whose equidistribution is known to every order. The product of the $\lceil (t+1)/2 \rceil$-point Gauss–Legendre rule in z and the $(t+1)$-point trapezoid rule in the azimuth integrates every spherical polynomial of degree at most t exactly (the measured error on $P_\ell(\langle a, \cdot \rangle)$, $\ell \le t$, is below $2 \cdot 10^{-15}$), with $N \approx t^2/2$ nodes and positive weights w_i. Positive rules that are exact to degree t with $N \asymp t^2$ nodes have $\mathrm{wce}_{H^\tau} = O(t^{-\tau})$ for every $\tau > 1$ (Brauchart & Hesse, 2007), so Corollary D.14 applies to them as to QMC designs of every order, with the floor $\pi \max_i w_i h(y_i)$ of Proposition D.13 for the weighted rule. We rotate each rule by one fixed Haar rotation (seed 2027), weight it by the von Mises–Fisher density ($\varkappa = 2$) of Appendix G.5, and search the maximal field and value errors exactly as there.

Table 9: Rotated Gauss–Legendre $\times$ trapezoid rules of degree t: maximal field error over x (grid, near-antipode search and directional limits), the floor $\pi \max_i w_i h(y_i)$, their ratio, and the maximal value error. Fitted slopes over the five largest N: -0.97 for the field maximum and for the floor, -1.46 for the value maximum.

t	N	max field error	floor	ratio	max value error
7	32	$4.89 \cdot 10^{-1}$	$4.66 \cdot 10^{-1}$	1.049	$8.28 \cdot 10^{-2}$
11	72	$2.41 \cdot 10^{-1}$	$2.36 \cdot 10^{-1}$	1.023	$2.77 \cdot 10^{-2}$
15	128	$1.42 \cdot 10^{-1}$	$1.40 \cdot 10^{-1}$	1.014	$1.24 \cdot 10^{-2}$
23	288	$6.57 \cdot 10^{-2}$	$6.52 \cdot 10^{-2}$	1.007	$3.85 \cdot 10^{-3}$
31	512	$3.76 \cdot 10^{-2}$	$3.74 \cdot 10^{-2}$	1.005	$1.66 \cdot 10^{-3}$
47	1152	$1.70 \cdot 10^{-2}$	$1.70 \cdot 10^{-2}$	1.002	$5.05 \cdot 10^{-4}$
63	2048	$9.65 \cdot 10^{-3}$	$9.65 \cdot 10^{-3}$	1.000	$2.15 \cdot 10^{-4}$

The maximal field error is the field of the heaviest single node: the ratio to the floor falls from 1.05 to 1.000, so the largest error is attained next to the antipode of the node with the largest weighted mass, where the field of the other nodes nearly cancels. The value maxima decay like $N^{-1.46}$, close to the rate $N^{-3/2+\epsilon}$ of Corollary D.14. The corollary allows the maximum anywhere between the floor and $C_\epsilon N^{-1+\epsilon}$; these rules sit at the floor. The optimized and Fibonacci sets of Appendix G.5 behave similarly, within 2% of their floors for $N \ge 64$ and 1.15–1.20 times them; they are not known to be designs of high order, and the law explains why their behaviour is possible, not why it occurs.

G.8 Checks of the one-point law, the dipole scalings and the value constant

The one-point law, replicate by replicate. The script `code/check_spd_one_point.py` tests Remark C.8 on the bounded non-commutative model (B) of Appendix G.3. Every sample Karcher functional there is 1-strongly convex and smooth, so (20) holds for every draw. We map $x^\star$ to the identity by the isometry $y \mapsto x^{\star -1/2} y x^{\star -1/2}$, compute $\mathrm{Hess}\, f(x^\star)$ by second differences of the Fréchet functional in normal coordinates (where it equals the Riemannian Hessian) with step 10^{-3} on a 2^{18}-point scrambled Sobol' rule, whose field at $x^\star$ is $1.5 \cdot 10^{-8}$, and predict the error of each replicate by $d_{\mathrm{pred}} = \left\| \mathrm{Hess}\, f(x^\star)^{-1} e_N(x^\star) \right\|$. The Hessian has eigenvalues 1.0000, 1.0229, 1.0233, 1.0236, 1.0237 and 1.0249.

The ratio $d / \left\| e_N(x^\star) \right\|$ alone varies across draws between the reciprocal extreme eigenvalues of the averaged sample Hessian; the prediction resolves that variation to the second-order remainder. The same law holds on the sphere, where the Hessian at the mean is λI and the median ratio is within 0.5% of $1/\lambda$ in all six configurations of Table 8.

Table 10: The one-point law on $\mathcal{P}_3$: measured Karcher-mean error against the prediction from the field at $x^\star$, over 20 replicates per cell. The deviation shrinks in proportion to the error, as a second-order remainder must.

sampler	N	median d/d_{pred}	max $\lvert d/d_{\mathrm{pred}}-1\rvert$	range of $d/\left\lVert e_N(x^\star)\right\rVert$
MC	2^6	1.00097	$5.1\cdot10^{-3}$	$[0.9756, 0.9862]$
MC	2^{10}	0.99992	$1.4\cdot10^{-3}$	$[0.9757, 0.9988]$
MC	2^{14}	1.00001	$2.8\cdot10^{-4}$	$[0.9763, 0.9970]$
SSobol	2^6	1.00005	$2.2\cdot10^{-3}$	$[0.9741, 0.9885]$
SSobol	2^{10}	0.99999	$1.6\cdot10^{-4}$	$[0.9758, 0.9852]$
SSobol	2^{14}	1.00000	$2.4\cdot10^{-6}$	$[0.9759, 0.9774]$

Dipoles on $\mathbb{S}^2$ and $\mathbb{S}^3$. The script `code/check_scaling_Sd.py` evaluates, by product quadrature in normal coordinates, the cap discrepancy $\mathcal{D}(\varrho_r)$ of the dipole ϱ_r with density $\frac{1}{2}q(z/r)$ relative to σ in normal coordinates at $-x_0$, where $q(w)=w_1\eta(\|w\|)$ and $\eta(t)=\exp(-1/(1-t^2))$, through the signed Stolarsky identity (13), and the field $J_r=\int \mathrm{Log}_{x_0}\,\mathrm{d}\varrho_r$ it induces at x_0, for $r\in\{0.4, 0.2, 0.1, 0.05, 0.025\}$; this is the cap-discrepancy case $\ell=1$ of Theorem D.11. The fitted exponents are $\mathcal{D}\sim r^{2.495}$ and $\langle J_r,e_1\rangle\sim r^{1.973}$ on $\mathbb{S}^2$, and $\mathcal{D}\sim r^{3.491}$ and $\langle J_r,e_1\rangle\sim r^{2.967}$ on $\mathbb{S}^3$, against the predicted $r^{d+1/2}$ and r^d; the ratio $\langle J_r,e_1\rangle/\mathcal{D}(\varrho_r)^{2d/(2d+1)}$ converges, to 2.61 on $\mathbb{S}^2$ and 3.53 on $\mathbb{S}^3$. The field exponent approaches d from below because of the factor $\pi-r\|w\|$ in (30).

Multipoles on $\mathbb{S}^1$. The script `code/check_multipole.py` computes exactly, as Fourier sums, the discrepancy wce_{H^τ} of the signed measure with density $q_\ell((\theta-\pi)/r)$ on $\mathbb{S}^1$, where q_ℓ is the ℓ-th derivative of $\psi(w)=e^{-w^2/2}$: its moments below order ℓ vanish and $\hat{q}_\ell(\xi)=(i\xi)^\ell\hat{\psi}(\xi)$. The proof of Theorem D.11 predicts the order $r^{\min\{\tau+1/2,\,\ell+1\}}$, with a logarithm at equality. Table 11 gives the exponents fitted over $r=2^{-7},\dots,2^{-12}$. The dipole ($\ell=1$) saturates at r^2 once $\tau>3/2$, where it excludes only field exponents above $\frac{1}{2}$, while $\ell=3$ follows $r^{\tau+1/2}$ up to $\tau=3$, which is what makes the field exponent $1/(\tau+\frac{1}{2})$ sharp there. Odd ℓ move the field at x_0 by order r and even ℓ the value by order r^2, by closed-form integrals in the script (fitted exponents 0.999 and 1.000, and 1.999 and 2.000).

Table 11: Fitted exponent of wce_{H^τ} in r for profiles with ℓ vanishing moments on $\mathbb{S}^1$, with the predicted $\min\{\tau+\frac{1}{2},\ell+1\}$ in parentheses; at equality the prediction carries a logarithm, which lowers the fitted value.

ℓ	$\tau=0.8$	$\tau=1.0$	$\tau=1.5$	$\tau=2.0$	$\tau=2.5$	$\tau=3.0$
1	1.30 (1.3)	1.50 (1.5)	1.91 (2.0)	2.00 (2.0)	2.00 (2.0)	2.00 (2.0)
2	1.30 (1.3)	1.50 (1.5)	2.00 (2.0)	2.50 (2.5)	2.91 (3.0)	3.00 (3.0)
3	1.30 (1.3)	1.50 (1.5)	2.00 (2.0)	2.50 (2.5)	3.00 (3.0)	3.50 (3.5)
4	1.30 (1.3)	1.50 (1.5)	2.00 (2.0)	2.50 (2.5)	3.00 (3.0)	3.50 (3.5)

Lower bounds in the square. The script `code/check_square_sharpness.py` evaluates the two constructions of Theorem D.12, first in clean cases in which every remainder of the proof vanishes. For $s=\cos\theta$ around $u_0=(\frac{1}{2},\frac{1}{2})$, whose mixed derivative is $F(\theta)\rho^{-2}$ with $F=\sin\theta(3\cos^2\theta-1)$, and $G=F/\|F\|_{L^2}^2$, the positivity constant of the proof is $M=3.23$, the error of ν_ϵ is $\epsilon\log(R/r)$ exactly with $R=0.2$, and the measured density stays above 0.80. For $\epsilon=2^{-16},\dots,2^{-42}$ the ratio of the error to $D^*\log(1/D^*)$ rises from 0.50 to 0.79 towards the limit $1/(2\max|G|)=0.982$, at which the ratio to D^* grows per unit of $\log(1/D^*)$. For the indicator of a half-plane with normal at angle $\pi/6$ the error is $-n_1n_2\epsilon R/w$, a constant 0.029 times $(D^*)^{1/2}$, and for an axis-parallel half-plane it vanishes; in these clean cases the slopes are consequences of the exact formulas. Direct two-dimensional quadrature of $\int g\,\partial_1\partial_2\Delta_\epsilon$, without integration by parts, reproduces the formulas to a relative 10^{-5} for the point and 10^{-11} for the edge. Two generic cases test the remainder estimates. With a nonlinear Z, a variable b and three angular modes of ω ($M=12.5$), the error minus $\epsilon\,b(u_0)\log(R/r)$ stays between -0.053ϵ and -0.050ϵ for $\epsilon=2^{-14},\dots,2^{-38}$, the forms with and without integration by parts agree to $2\cdot10^{-14}$, and the density stays above 0.86. With a variable

jump J, a smooth part $c_0 \neq 0$ and a normal at angle 0.5, the error minus the main term of the proof is $-0.109\,\epsilon w R$, which shows that the remainder is even $O(\epsilon w R)$, the error is twice the stated lower bound, and the density stays above 0.70.

The value constant. The script `code/check_electro_constant.py` computes $c_r = 1.4966$ from the Legendre coefficients of R_x and of Stolarsky's kernel (the coefficients of K_2 decay like $0.124\,\ell^{-3}$ and those of R_x like ℓ^{-5}, as the orders $3/2$ and 3 of their singularities require) and the closed form $\kappa_{\mathrm{val}} = 6.0619$ of (36). The direct Legendre series for the same seminorm gives 6.0538 at degree 400, still converging like the inverse degree, and the triangle inequality $\|U\|_\infty \leq C_2\sqrt{1-k_0}\,D_{L_2}$ gives the weaker $4\pi/\sqrt{3} + c_r = 8.752$. A single node attains the ratio 6.0057. On $2 \cdot 10^4$ random decision points, the ratio $\sup_x |f_N - f|/D_{L_2}$ is 0.37 and 0.08 for Fibonacci sets of 64 and 512 nodes, 0.40 for a Lambert image of a scrambled net with 256 nodes, 3.39 and 3.25 for 64 and 512 i.i.d. nodes, and 3.31 for 200 nodes clustered towards the poles.

The three regimes of the law, and its borderline. The script `code/check_borderline.py` evaluates exactly, in Fourier space, the value error of probability densities on $\mathbb{S}^1$ of the form $1 + \sum_{j \leq L} c_j \psi_0((\theta - \pi)2^j)$, with $\psi_0(s) = e^{-8s^2} - \frac{1}{2}e^{-2s^2}$ even and of mean zero and $c_j 4^{-j}$ constant, against the cone $\frac{1}{2}\theta^2$ of the value integrand at the antipode ($m = 1$, $d = 1$). Every dyadic scale contributes equally to the error, and equally to $D^2 = \mathrm{wce}^2_{H^\tau}$ exactly when $\tau = d/2 + 1 = 3/2$. For $L = 2, \ldots, 14$: at $\tau = 1.3$, where the cone lies in H^τ, the ratio error$/D$ stays below 1.83; at $\tau = 1.8$, where the law gives the exponent $2/2.3$, the ratio error$/D^{2/2.3}$ stays below 1.77; at the borderline $\tau = 1.5$ the ratio error$/D$ grows from 2.66 to 4.68 while error$/(D\sqrt{\log(1/D)})$ stays within 4% of 1. This is the borderline construction of Theorem D.11 for $d = 1$ (the profile ψ_0 has vanishing moments of orders 0 and 1), and $D(\log(1/D))^{1/2}$ is exactly the borderline order of Theorem D.6. The same script checks the mechanism of the upper bound. By FFT on 2^{22} points it computes the $H^{3/2}(\mathbb{S}^1)$ norm of the cut cone $g_r(z) = \beta_0(|z|/r)\psi_{\mathrm{out}}(|z|)|z|$, where ψ_{out} equals 1 on $[0, \pi/8]$ and 0 on $[\pi/4, \infty)$, and of its dyadic pieces φ_j. For $r = 2^{-6}, \ldots, 2^{-12}$, $\|g_r\|^2$ grows by 0.140 per halving of r, the value $2\log 2/\pi^2$ predicted by the Fourier coefficients $-1/(\pi\ell^2)$ of a single cone, so $\|g_r\|$ grows like $(\log(1/r))^{1/2}$, while the triangle-inequality bound $\sum_j \|\varphi_j\|$ grows by 2.95 per halving. Each piece has squared norm about 8.9, and the signed normalized Gram entries of neighbouring pieces are -0.511, 0.017 and 0.0017 at lags 1, 2 and 3: adjacent shells nearly cancel, and the rest decays geometrically, as in the proof of Lemma D.4.

Flat tori: the chart decides alignment. The script `code/check_torus.py` tests Proposition E.11 on the square lattice $\mathbb{Z}^2$ and on the hexagonal lattice with basis $(1, 0)$, $(\frac{1}{2}, \frac{\sqrt{3}}{2})$, with the same point sets in the fundamental-cell chart: the first N points of the unscrambled Sobol' sequence, 16 scrambled Sobol' nets and 16 i.i.d. sets, $N = 2^6, \ldots, 2^{16}$. The population objective is constant in x, so the field is $-\frac{1}{N}\sum_i \mathrm{Log}_x T(u_i)$ and the value error is the deviation of $\frac{1}{N}\sum_i \frac{1}{2}d(x, T(u_i))^2$ from $\frac{1}{12}$ and $\frac{5}{72}$, half the mean squared norm on the two Voronoi cells (checked by quadrature to 10^{-6}). Over 64 random decision points, the maximal field error of the Sobol' points decays with slope -0.99 on the square lattice, whose cut locus is axis-parallel, and -0.67 on the hexagonal lattice, whose cut locus is oblique; for scrambled nets (RMS over scramblings) the slopes are -1.02 and -0.75, for i.i.d. points -0.52 and -0.49. The maximal value errors decay with slopes -1.95 and -1.34 (Sobol') and -2.06 and -1.25 (scrambled). Changing only the lattice, with the same points, costs the field its linear rate while the value keeps at least the order D_N^*, as Proposition E.11 states; on the oblique lattice the measured field rate lies between the proved $(D_N^*)^{1/2}$ and D_N^*. (A dyadic grid of decision points would align the square lattice's cut locus with the strata of the nets and hide its jump terms, which is why the decision points are random.)

G.9 Finite-precision scrambling and endpoints

SciPy's scrambled Sobol' points carry 30 random binary digits and no random tail (Lemma B.5). A complete one-dimensional net with 8 random digits and 256 points has mean $0.498046875 = \frac{1}{2} - 2^{-9}$, the grid expectation of Lemma B.5(iii); at 30 digits the corresponding bias for the linear integrand is $-2^{-31} \approx -4.7 \cdot 10^{-10}$. The script `code/exp_finite_precision.py` measures what this changes for the estimators of the paper. For each implemented point set it draws the tail digits 31–52

of the ideal scrambling that shares its first 30 digits, as in the proof of Lemma B.5(i), and evaluates both rules on the same draw.

Paired comparison with the ideal scrambling. For the bounded Grassmann gradient at $X^\star$ (the 200 nets per size of Figure 3(b)), the mean squared error changes by at most a relative $3.9 \cdot 10^{-9}$ and no paired gradient changes by more than $3.9 \cdot 10^{-9}$, against root-mean-square errors between 1.8 and $4.9 \cdot 10^{-4}$ and the worst-case bound $\epsilon_{30} = 3.6 \cdot 10^{-7}$ of Lemma B.5 with $L_u \le 12 \|W\|_2^2 \sqrt{8}$. For the fixed-set quantities of Appendix G.2 (64 scramblings, $N = 2^{10}, \dots, 2^{15}$), no cross-block, covariance or subspace value changes by more than $9.3 \cdot 10^{-10}$, and the fractions of scramblings with a cross block below 10^{-10} are 0.02, 0.06, 0.23, 0.48, 0.72, 0.84 for the implemented rule against 0.02, 0.05, 0.22, 0.47, 0.69, 0.83 for the ideal one. For the bounded SPD model (20 scramblings per size), the root-mean-square gradient residual at the reference mean changes by at most a relative $4.4 \cdot 10^{-5}$, paired residuals differ by at most $1.4 \cdot 10^{-9}$, and at $N = 2^{14}$ the root-mean-square Karcher-mean error is $8.4749 \cdot 10^{-7}$ for both rules, with paired means at most $1.1 \cdot 10^{-9}$ apart.

Endpoints. The Gaussian transports evaluate Φ^{-1} without clipping, and $\Phi^{-1}(0) = -\infty$; with 30 random digits a scrambled coordinate equals 0 with probability 2^{-30}. The script regenerates from its seed every scrambled net that fed Φ^{-1} in the stored runs (the Gaussian Grassmann sample averages, gradient-noise measurement and RSGD runs, and the symmetric SPD model): $3.6 \cdot 10^8$ coordinates, for which about 0.33 zeros would be expected. None is 0; the smallest is 2^{-30}, where $\Phi^{-1} \approx -6.0$, and every stored value of these runs is finite. The same audit finds no node on the edge $u_1 = 0$ in the von Mises–Fisher experiments, where the implementation would return the zero vector for $\mathrm{Log}_{e_3}(-e_3)$, the blown-up antipode of Proposition E.10. On the stored runs, the implemented Gaussian estimators are therefore the finite-grid rules of Lemma B.5; for these unbounded transports that rule is a diagnostic of the boundary-growth regime, not the subject of a theorem.

G.10 Implementation details

The core experiments (Grassmannian, positive definite cone, sphere) are implemented in Python 3.13 with NumPy 2.4 and SciPy 1.17 (the result files record CPython 3.13.14, NumPy 2.4.6 and SciPy 1.17.1 on Windows); point sets come from `scipy.stats.qmc` (Halton: `Halton(d, scramble=False)`, first point dropped; Sobol': `Sobol(d, scramble=True, seed)` with direction numbers of Joe & Kuo (2008) and the linear scrambling plus digital shift of Matoušek (1998), at the default precision of 30 randomized bits), Gaussian variables from the inverse normal distribution function, and all runs are seeded and reproducible from the scripts `code/exp_grassmann.py`, `code/exp_spd.py`, `code/exp_spd_asym.py` and `code/exp_sphere.py`; raw results, including every replicate value, are stored as JSON files in `code/results/`. The experiments `code/exp_grassmann_bounded.py` and `code/exp_sphere_rotation.py` import the problem definitions of these scripts and were run on macOS (Apple silicon) with CPython 3.13.9 from an Anaconda installation, SciPy 1.16.3 from conda, and NumPy 1.26.4 built from source by pip (meson; wheel tag `cp313-cp313-macosx_26_0_arm64`); the official 1.26.4 wheels cover Python 3.9–3.12 only. The control, finite-precision and uncertainty scripts named in the subsections above, together with `analysis_uncertainty.py` for the resampling intervals, ran in the same environment and store with each result file a record written by the executing process: interpreter path and version, versions and installation provenance of NumPy and SciPy, platform, processor count and thread settings (`code/revision_env.py`).

Grassmannian. Q is the orthogonal factor of the QR decomposition of a Gaussian 8×8 matrix (seed 1), signs fixed so that the triangular factor has positive diagonal. Riemannian gradient descent on f_N uses the gradient $-(I - XX^\top)S_N X$, step $1/\lambda_{\max}(S_N)$, the QR retraction $X \mapsto \mathrm{qf}(X + \eta\,\cdot)$ and stops at $\|\mathrm{grad}\, f_N\|_F < 10^{-10}$ (41–48 iterations). In the RSGD runs the scrambled net at iteration k of replicate r for minibatch size 2^m uses seed $(16m + r)10^7 + k$, so that nets are independent across iterations, replicates and N; the plateau statistic uses every iteration of the last quarter of each run, and the first iteration at which the error is within a factor 2 of the final plateau is at most 12% of the run length in every configuration. The gradient-noise measurement uses 200 independent nets per N. Total runtime about eight minutes on seven processes.

Affine-invariant cone. Q is the orthogonal factor of a Gaussian 3×3 matrix (seed 3). Matrix exponentials, logarithms and square roots are computed by eigendecomposition, batched over the sample index. The Karcher iteration starts at the Euclidean mean and stops when $\left\| \frac{1}{N} \sum_i \mathrm{Log}_x(y_i) \right\|_x < 10^{-13}$. In the symmetric model the identity $\mathrm{Log}_P(T(u)) = P^{1/2} S(u) P^{1/2}$ was used as a check of the matrix code (agreement to 10^{-15}). In the non-commutative model (B) the reference mean is the Karcher mean of three sample means computed from scrambled Sobol' nets of 2^{20} points (seeds $100, 101, 102$); as a check, the same model without the offset m_0 has $\mathbb{E}S = 0$ and therefore, by the identity $\mathrm{Log}_P(T(u)) = P^{1/2} S(u) P^{1/2}$, mean exactly P, which the reference computation reproduces to 10^{-9}. Runtime under two minutes per model.

Sphere. The Stolarsky ascent uses the Armijo parameter 10^{-4}, initial step $2/N$ (doubled after each accepted step), and stops at 1000 iterations or when the relative increase of $\mathcal{E}$ is below 10^{-14}; the deficit is evaluated from the Gram matrix and agrees with direct pairwise differences to 10^{-12}. The Fibonacci spiral is $z_i = 1 - (2i+1)/N$, $\varphi_i = 2\pi i/\phi$ with ϕ the golden ratio. The von Mises–Fisher transport was checked against the exact mean resultant $\coth \varkappa - 1/\varkappa$ to 10^{-9} on 2^{20} points. Reference values $f(x_a)$, the Hessian eigenvalue $\lambda = f''(0)$ at e_3 and the reference gradients $g(x) = \int \mathrm{Log}_x(y)\, \mathrm{d}\nu(y)$ were computed by a Gauss–Legendre rule in z with 12000 nodes and a midpoint rule in φ with 8192 nodes, with cross-checks against a second rule and a one-dimensional reduction agreeing to $4 \cdot 10^{-11}$; the reference gradient at x_a is accurate to $7 \cdot 10^{-11}$, six orders of magnitude below the smallest plotted error. Weighted sample Fréchet means use $w_i = h(y_i)/\sum_j h(y_j)$ and the intrinsic iteration started at the normalized weighted Euclidean mean. Runtime: about six minutes for the ascent runs, under two minutes for the Fréchet-mean sweep.

Bounded Grassmannian. Transport $a(u) = \sqrt{3} W(2u - \mathbf{1})$ with the problem constants imported from `exp_grassmann.py`. Gradient noise: 200 minibatches per size with seeds $5 \cdot 10^7 + 1000m + r$ (MC) and $6 \cdot 10^7 + 1000m + r$ (scrambled Sobol'), $N = 2^4, \ldots, 2^{12}$, slopes over all nine sizes with a 2000-resample bootstrap. RSGD: as in the Gaussian run, with the net at iteration k of replicate r seeded by $10^9 + (16m + r) \cdot 10^7 + k$ and MC generators by $7 \cdot 10^7 + 1000r + m$; the population gradient norm and the subspace error are recorded on the same iteration grid, and the plateau uses the last quarter of each run. SAA: Halton with the origin dropped and one scrambled Sobol' net with seed 0, $N = 2^4, \ldots, 2^{14}$, Riemannian gradient descent from the shared X_0 with step $1/\lambda_{\max}(S_N)$ to $\|\mathrm{grad}\, f_N\|_F < 10^{-10}$; the exactness probe uses seeds 0–3 and $N = 2^{11}, \ldots, 2^{15}$. Oracle-tuned envelopes are running minima over the pooled pairs $(kN,\ \mathrm{RMS}\ \|\mathrm{grad}\, f(X_k)\|_F)$ of one route, with slopes fitted over budgets $\geq 2^{12}$. Runtime 68 s on nine processes.

Grassmannian controls, finite precision and uncertainty. Fixed point sets: scrambled Sobol' nets with seeds 0–63, i.i.d. sets from `default_rng([9e7, m, r])`, Halton and the unscrambled net, $N = 2^4, \ldots, 2^{15}$; the reused routes use the solver above on $N \leq 2^{14}$, with times as the minimum of five repetitions in one process. Smooth control: seeds $9.2 \cdot 10^7 + 1000m + r$ (scrambled) and `default_rng([9.1e7, m, r])` (MC), 200 sets per size. Finite precision: the tail digits of each point set come from an independent generator seeded by the set's seed, and the endpoint audit regenerates every stored scrambled net from its seed. Resampling intervals are percentile intervals over independent units only (replicates, runs, or the 50 shared rotations), with 20,000 resamples for ratios and 5000 for slopes (seed 20260924). Runtimes: 25 s, 173 s and 410 s on up to nine processes for the controls, the finite-precision checks and the sphere maxima.

Sphere: rotations, sampled maxima and cost. Rotations: 50 Haar-distributed elements of $SO(3)$ obtained from QR factorizations of Gaussian matrices with sign correction (`default_rng(777)`); the deficit of each rotated set is recomputed as a check. Sampled maxima: grid of 91 polar angles $a_i = i\pi/90$ times 48 azimuths (4274 points), refined grids and near-antipode points as in Appendix G.5; reference values of $\mathrm{grad}\, f$ and f along the meridian from the (z, φ) Gauss–Legendre$(3000) \times$midpoint(2048) rule of the Fréchet script, with a 1500×1024 rule as the accuracy check; eight rotations per set for the rotated maxima. Weighted Fréchet means as in the Fréchet script (tolerance 10^{-14}). The cost accounting reads the stored ascent statistics. Runtime 166 s in a single process.

Chart experiment, designs and checks. The scripts `code/exp_chart_resolution.py`, `code/exp_sphere_designs.py`, `code/check_spd_one_point.py`,

`code/check_scaling_Sd.py`, `code/check_electro_constant.py`, `code/check_borderline.py` and `code/check_torus.py` import the problem definitions of the scripts above and ran on macOS (Apple silicon) with CPython 3.13, NumPy 1.26.4 and SciPy 1.16.3, in a single process. The chart experiment uses scrambled nets with seeds 0–63 for each size and i.i.d. sets from `default_rng`$(424242 + 1000m + r)$ for size 2^m and replicate r; the weighted Fréchet means use the fixed-point iteration of the sphere scripts with tolerance 10^{-14} and at most 200 iterations; runtime about ten minutes. The design experiment `code/exp_sphere_designs.py` reuses the reference values and the search of `code/exp_sphere_maxima.py`; runtime 35 s. The one-point check uses the 20 replicates of each sampler of `code/exp_spd_asym.py` and a scrambled Sobol' reference rule of 2^{18} points (seed 100) for the Hessian, 22 s. The dipole check uses product Gauss–Legendre rules with 40×96 nodes on the disc and $20 \times 20 \times 40$ nodes on the ball, 24 s. The constant check uses Legendre coefficients up to degree 400 from a 4000-node Gauss–Legendre rule and $2 \cdot 10^4$ random decision points (seed 0). The multiscale part of the borderline check is exact in Fourier space and takes under a second, and its shell part uses FFTs on 2^{22} points and takes about 30 s; the torus check uses the SciPy Sobol' generator (unscrambled, and scrambled with seeds 0–15), i.i.d. sets from `default_rng`(20260926) and 64 decision points from `default_rng`(7), about two minutes. `tools/verify_rising_sea.py` recomputes every number quoted from these results, and `tools/render_law_figures.py` redraws Figures 2 and 4.

H Related work and open directions

H.1 Additional related work

The use of QMC points in sample average approximation (SAA) of Euclidean stochastic programs is classical: Koivu (2005) and Pennanen & Koivu (2005) prove epi-convergence and observe variance reduction, and Homem-de Mello (2008) derives rates of convergence of SAA under non-i.i.d. sampling, including randomized QMC. In machine learning, Buchholz et al. (2018) use randomized QMC in stochastic variational inference and prove convergence of the resulting SGD, and Liu & Owen (2021) combine RQMC with quasi-Newton steps. None of these concern a manifold-valued decision variable, and none address vector-valued integrands in a moving tangent space. On the Riemannian side, stochastic gradient methods go back to Bonnabel (2013); nonasymptotic rates under geodesic convexity are due to Zhang & Sra (2016), variance reduction to Zhang et al. (2016); Sato et al. (2019), and iterate averaging to Tripuraneni et al. (2018); Boumal et al. (2019) and Bento et al. (2017) give the deterministic complexity of Riemannian gradient descent. Fréchet means from i.i.d. samples converge at rate $N^{-1/2}$ by the central limit theorems of Bhattacharya & Patrangenaru (2003; 2005); the "smeary" central limit theorems of Hotz & Huckemann (2015); Eltzner & Huckemann (2019) exhibit slower rates in regimes with degenerate local Fréchet geometry; this is a separate phenomenon from the chart-dependent singularities of Proposition E.10. QMC on the sphere is developed by Brauchart & Dick (2012); Aistleitner et al. (2012); Brauchart et al. (2014); Koksma–Hlawka inequalities on manifolds and metric measure spaces by Brandolini et al. (2013; 2019); energy-minimizing point configurations by Hardin & Saff (2004); Borodachov et al. (2019); Brauchart & Grabner (2015), and their computation by Riemannian methods by Gräf & Potts (2011); deterministic low-discrepancy grids on $SO(3)$ are constructed through the Hopf fibration by Yershova et al. (2010). For Fréchet means in Hadamard spaces, non-asymptotic $N^{-1/2}$ bounds for i.i.d. samples go back to Sturm (2003). QMC *integration* on manifolds other than spheres has been studied for its own sake: Breger et al. (2017) construct QMC point sets on Grassmannians and prove worst-case-error bounds for kernel-based rules, Ehler et al. (2019) obtain optimal integration rates up to a logarithmic factor on closed manifolds by reweighting random points, and Basu & Owen (2017) extend scrambled nets to products of spaces such as spheres and triangles; these works supply nodes and error bounds that Appendix E can use as input, but do not concern an optimization problem. Closer in spirit, Lu et al. (2022) analyse example-selection schemes for SGD, including low-discrepancy orderings, and Liu (2023) uses QMC in Langevin sampling; Leövey & Römisch (2021) and Melnikov & Milz (2025) apply randomized QMC to two-stage and risk-averse stochastic programs, always in Euclidean decision spaces. QMC minibatches have also been used inside a manifold-valued optimization: Luu et al. (2025) optimize a Gaussian variational objective on the Bures–Wasserstein manifold and, as an empirical baseline for their control-variate estimator, run their stochastic scheme with quasi-Monte Carlo minibatches (their Appendix B.5 and Figure 8), finding that around fifty scrambled-Sobol'

samples per iteration match the variance-reduced estimator. These strands supply the quadrature and optimization ingredients used here, and the last one shows the forward direction at work on a particular manifold. Against this background, the contribution is an organizing object and an analysis at the cut locus. The object is the field of the discrepancy charge. Its three readings (Proposition C.7) are known arguments, Koksma–Hlawka in the direction of the error, a variance decomposition and the influence expansion of M-estimators (Bhattacharya & Patrangenaru, 2003; 2005); organized around the field they turn quadrature bounds into the optimization statements of Appendix C and identify the minimizer error with the quadrature error of one vector-valued integrand. The analysis is the scaling law of Theorems D.6 and D.10 and Proposition E.11 in the settings proved there, with its sharpness on spheres and in the square (Theorems D.11 and D.12) and the floor (Proposition D.13); its exactness (Proposition D.2) is Sion's theorem. Applied to the geometry it yields the electrostatic form of the Fréchet problem on spheres (Proposition E.9) and the chart as a blow-up (Proposition E.10). On the sphere, Erdős–Turán inequalities bound the cap discrepancy through one-sided polynomial approximation of cap indicators (Grabner, 1991; Li & Vaaler, 1999) and give discrepancy bounds for spherical designs (Grabner & Tichy, 1993), a codimension-one instance of the sandwich; energy minimizers are linked to numerical integration by Damelin & Grabner (2003), positive quadrature on scattered nodes is constructed by Mhaskar et al. (2001), and the worst-case error of positive rules exact to degree t is bounded by Brauchart & Hesse (2007). Koksma–Hlawka inequalities for piecewise smooth functions (Brandolini et al., 2013; 2019) and one-sided (Beurling–Selberg) approximation (Vaaler, 1985) are the closest analytic antecedents of the sandwich; the classical bound of the isotropic by the star discrepancy (Kuipers & Niederreiter, 1974; Niederreiter, 1992) is the oblique-segment case of the law. Stolarsky's principle and its generalizations relate discrepancies to energies (Stolarsky, 1973; Brauchart & Dick, 2013; Bilyk et al., 2018); here the same potential that defines the energy also carries the Fréchet value and gradient errors. The closest neighbours on the optimization side are Buchholz et al. (2018), who prove convergence guarantees for RQMC-based stochastic optimization in Euclidean variational inference, Melnikov & Milz (2025) for epi-convergence of RQMC sample averages in Euclidean space, and Basu & Owen (2016) for the effect of transformations on Hardy–Krause variation, including positive results for the Fang–Wang spherical coordinates, which on $\mathbb{S}^2$ give the equal-area map; relative to it, Proposition E.2 supplies explicit constants for arbitrary increasing latitude profiles, and Proposition E.10 explains which decision points such a chart serves.

H.2 Open directions

1. *Cut loci of codimension one.* On real projective spaces, on flat tori with oblique lattices and on most compact manifolds the cut locus is a hypersurface. The dilation argument, run in a tubular neighbourhood instead of a ball, predicts the exponent $1/(\tau + \frac{1}{2})$ for a jump in H^τ and the matching dipole a layer of width r along the hypersurface; Theorem D.10(b) is the two-dimensional Hardy–Krause case, and Proposition E.11 applies it to flat tori.

2. *Borderline factors.* The law is sharp wherever it is proved, borderline factors included: $(\log(1/D))^{1/2}$ in Sobolev norms, where the dyadic shells add in quadrature, and $\log(1/D)$ in Hardy–Krause variation, an L^1 norm of the mixed derivative, in which they add linearly (Lemma D.4 and Theorems D.11 and D.12). For certificates built on L^p norms of derivatives one expects the power $1/p$ of the logarithm.

3. *Randomized rates at a point singularity.* The upper bound $N^{-1}\log^2 N$ for scrambled nets at an unresolved cut point follows from the deterministic certificate of every realization (Proposition E.10); the observed rate N^{-1} calls for a matching variance lower bound, and the observed smooth-integrand rate in resolved charts for a variance bound that tolerates the integrable boundary singularities of polar charts.

4. *Which equidistribution the floor needs.* Corollary D.14 reaches the floor with designs of every order. The optimized and Fibonacci sets reach it in our experiments without a proof that they are designs of high order; a certificate for them would need a QMC design order above $3/2$ (Brauchart et al., 2014), and an energy whose minimizers have it.

5. *Adaptive charts.* The prescription to blow up the antipode of the minimizer suggests re-centring the chart along an optimization path, with the importance weights updated accordingly; the cost of the moving chart against the gained rate is open.

6. *Lower bounds for the budgets.* Corollary C.11 gives $\widetilde{O}(\varepsilon^{-3})$ evaluations for ε-stationarity with a chosen-point oracle. Whether $2 + 1/\alpha$ is the optimal exponent for integrands of mixed smoothness α, as 4 is for unbiased oracles (Arjevani et al., 2023), would combine Roth-type lower bounds on discrepancy with oracle lower bounds for nonconvex optimization.
7. *High-dimensional integration spaces.* The constants $C_s(\log N)^s$ are the usual curse of QMC. Weighted function spaces and tractability theory (Dick et al., 2013) describe when the dependence on s is benign, and the weighted variation of the field integrand is the right object.
8. *Variance reduction.* Control variates and stratification are compatible: the full-gradient evaluation of Riemannian SVRG (Zhang et al., 2016; Sato et al., 2019) can be a large deterministic net while the inner steps use small scrambled nets, and growing the net size along the iterations gives a sample-size schedule with a built-in certificate.

I THE EPIGRAPH AND THE METHOD

The sentence below the abstract comes from the last part of Riemann's habilitation lecture of 1854, in Clifford's translation (Riemann, 1873). Riemann separates there the questions about the infinitely great, which he calls useless for the interpretation of nature, from those about the infinitely small, on which our knowledge of causal relations depends, and he goes on to suggest that the empirical notions on which metric determinations rest may cease to be valid in the infinitely small. The sentence describes the method of this paper, and the three places where it does so are worth making explicit.

The rate is decided in the infinitely small. Every exponent of Table 1 relates two global quantities, a discrepancy D and an integration error, but it is decided in a neighbourhood of the singular set whose size tends to zero with D. The proof of Theorem D.6 follows the singular integrand into this neighbourhood: it cuts the integrand at the scale r, pays for the cut part shell by shell over the dyadic annuli $2^j r \le \|z\| < 2^{j+1} r$, each a dilation of one profile (Lemmas D.3 and D.4), and pays for the discarded ball through the node mass it contains, which positivity bounds through a bump. Balancing the two costs puts the scale at $r = D^{1/(c+\beta)}$ (Figure 6). The smaller the discrepancy, the deeper the analysis must go, and the exponent $(m + c)/(c + \beta)$ retains only three infinitesimal features of the problem: the order m of the singularity, the codimension c of the set that carries it, and the dilation exponent β of the certificate norm.

The exactness is necessary. The lower bounds follow the same phenomenon to the same scale from the other side. The multipoles of Theorem D.11 are concentrated within distance r of the singular point, and the discrepancy functions of Theorem D.12 reach down to the scale $\epsilon^{1/2}$ that positivity allows (Figures 7 and 8); both show that no bound in terms of the discrepancy alone does better, and at the borderline $m = \beta$ the logarithms count the dyadic shells between the cut scale and the unit scale. What a certificate can say about the minimizer is therefore exactly what survives this analysis in the infinitely small, and the floor of Proposition D.13 is the same analysis applied to a single node, whose field reverses its direction as the decision point crosses the node's antipode.

Causal relations and blow-ups. Riemann's causal relations have a precise counterpart here, the chain from the point set to the minimizer. The discrepancy charge generates a potential and a field, and to first order the minimizer error is the field at the minimizer through an inverse Hessian (Proposition 2.2); this link is an identity, not a bound. At the cut locus the link passes through the local structure of the loss: on the sphere, $\frac{1}{2}d(x, \cdot)^2$ has a cone at the antipode $-x$, and its slope in normal coordinates there is the field integrand $-\operatorname{Log}_x$ (Figures 1 and 9). There a notion does cease to be valid in the infinitely small, the direction from x to a data point, which jumps at $-x$. A chart that blows up the cut point is the geometric form of following the phenomenon into the infinitely small: it magnifies the antipode into the circle of directions at it, on which the jumping direction becomes a smooth coordinate, and the field regains the rate of the star discrepancy (Proposition E.10).