Robust k-means quantization in metric spaces

Alexey Kroshnin (WIAS)

Statistics in Metric Spaces, 11–13 October 2023, ENSAE

Outline

1. Intro to quantization
2. Robust k-means
3. Lower bounds
4. Ideas of the proofs

k-means quantization/clustering

Let P be a probability measure on $\mathbb{R}^d$, $X \sim P$ random vector
k-means clustering/quantization problem:

$$D(Q) = D(Q; P) \coloneqq \mathbb{E} \min_{a \in Q} |X - a|^2 \to \min_{Q \subset \mathbb{R}^d : |Q| = k}$$

- Clustering: clusters are Voronoi cells
  $V_i(Q) \coloneqq \left\{ x \in \mathbb{R}^d : |x - a_i| = \min_j |x - a_j| \right\}$, $Q = (a_1, \ldots, a_k)$
- Quantization: $Q = (a_1, \ldots, a_k)$ is a "codebook",
  $i(x) \coloneqq \operatorname*{argmin}_{j = 1, \ldots, k} |x - a_j|$ is a "code"

Some history:

- Steinhaus (1957): division of a body in $\mathbb{R}^d$
- Lloyd (1957): algorithm for signal quantization
- MacQueen (1967): "k-means" name

Example: color quantization

original

$k = 3$

$k = 5$

$k = 10$

k-means in metric spaces

Let P be a probability measure on a metric space $(\mathcal{X}, d)$, $X \sim P$

$$D(Q) = D(Q; P) := \mathbb{E} \min_{a \in Q} d^2(X, a) \to \min_{Q \subset \mathcal{X} : |Q| = k}$$

Existence of solution: $\mathbb{E}\, d^2(X, x_0) < \infty$ for some $x_0 \in \mathcal{X}$; there is a weak topology τ_w on $\mathcal{X}$ s.t. any closed ball $B_r(x)$ is compact in τ_w

Examples: separable reflexive Banach spaces, Wasserstein spaces on $\mathbb{R}^d$, Riemannian manifolds

Statistical setting

Given an i.i.d. sample $\boldsymbol{X} = (X_1, \dots, X_n) \sim P$ we want to construct an empirical quantizer $\widehat{Q} \subset \mathcal{X}$, $|\widehat{Q}| = k$

Measure of quality: excess distortion $D(\widehat{Q}) - D(Q^*)$, where Q^* is an optimal quantizer

Our goal is to get $\widehat{Q}$ with good PAC bounds:

$$\mathbb{P}\left\{D(\widehat{Q}) - D(Q^*) > \varepsilon(n, \delta)\right\} \leq \delta$$

ERM consistency

Risk minimization problem $\Longrightarrow$ empirically optimal quantizer:

$$\widehat{Q}_n := \underset{Q \subset \mathcal{X} : |Q| = k}{\operatorname{argmin}} \sum_{i=1}^{n} \min_{a \in Q} d^2(X_i, a)$$

Strong consistency (Pollard, 1981): let $X_1, X_2, \dots \sim P$ be an i.i.d. sequence in $\mathbb{R}^d$ and $\mathbb{E}|X|^2 < \infty$; then

$$D(\widehat{Q}_n) - D(Q^*) \xrightarrow{\text{a.s.}} 0 \text{ as } n \to \infty$$

Under additional assumptions $\sqrt{n}(\widehat{Q}_n - Q^*)$ is asymptotically normal (Pollard, 1982)

Q: What about non-asymptotic rates of convergence?

ERM rates: bounded support in Hilbert space

Let $\mathcal{X}$ be a separable Hilbert space. Assume $\|X\| \le T$ a.s.

Non-asymptotic bounds on the excess distortion w.p. at least $1-\delta$:

- Linder, Lugosi, and Zeger (1994): $\mathcal{X} = \mathbb{R}^d$,

$$D(\widehat{Q}_n) - D(Q^*) \lesssim T^2\left(\sqrt{\frac{kd\log n}{n}} + \sqrt{\frac{\log(1/\delta)}{n}}\right)$$

- Biau, Devroye, and Lugosi (2008):

$$D(\widehat{Q}_n) - D(Q^*) \lesssim T^2\left(\sqrt{\frac{k^2}{n}} + \sqrt{\frac{\log(1/\delta)}{n}}\right)$$

- Fefferman, Mitter, and Narayanan (2016):

$$D(\widehat{Q}_n) - D(Q^*) \lesssim T^2\left(\sqrt{\frac{k(\log n)^4}{n}} + \sqrt{\frac{\log(1/\delta)}{n}}\right)$$

- Appert and Catoni (2021):

$$D(\widehat{Q}_n) - D(Q^*) \lesssim T^2\left(\sqrt{\frac{k\log^2(n/k)\log k}{n}} + \sqrt{\frac{\log(1/\delta)}{n}}\right)$$

ERM rates: light tails

Cadre and Paris (2012): if $\|X\|$ is sub-exponential, then with probability at least $1 - \delta - O\left(e^{-rn^{1.5}}\right)$

$$D(\widehat{Q}_n) - D(Q^*) \lesssim R^2(P) \frac{k \log \frac{k}{\delta}}{\sqrt{n}}$$

Questions

1. Heavy-tailed distribution: what if P has only two moments?
2. Outliers: what if the sample is contaminated?
3. (Sub-)optimality of ERM: can we do better?

Goals

Construct a quantizer $\widehat{Q}$ that

1. handles general metric space
2. is robust to heavy-tailed distributions/outliers
3. has a sub-Gaussian rate even for heavy tails

Counterexample

Take $\mathcal{X} = \mathbb{R}$, $k = 2$

Define the distributions P_n: $P_n(\{0\}) = 1 - \frac{1}{n}$, $P_n(\{\sqrt{n}\}) = \frac{1}{n}$.
Then $\mathbb{E}_{X \sim P_n} |X|^2 = 1$ and $D(Q^*; P_n) = 0$.

Let $X_1, \ldots, X_n \sim P_n$. Then with constant probability
$X_1 = \cdots = X_n = 0$, hence $\widehat{Q} = \{0\}$, $D(\widehat{Q}; P_n) = 1$.

Problem: there is too small cluster

Minimal cluster assumption

Voronoi cells (clusters) of $Q = \{a_1, \dots, a_k\} \subset \mathcal{X}$:

$$V_i(Q) \coloneqq \big\{x \in \mathcal{X} : d(x, a_i) \le d(x, a_j),\, 1 \le j < i,\\ d(x, a_i) < d(x, a_j),\, i < j \le k\big\}$$

Recall: let $|\operatorname{supp} P| \ge k$ and $\mathbb{E}\, d^2(X, x_0) < \infty$, then there is $0 < p_{\min} \le \min_i P(V_i(Q^*))$.

Suppose we are given a lower bound $p_{\min} > 0$ such that $np_{\min} \gg 1 \implies$ no "invisible" cluster. Our empirical quantizer and bounds will depend on $p_{\min}$.

Approaches to robust M-estimators

- Robust loss: ℓ_1, Huber loss, …
- Consensus: RANSAC, median of means, …
- Truncation: trimmed mean, …

Approaches to robust M-estimators: k-means

- Robust loss: k-medians, information k-means (Appert and Catoni, 2021)
- Consensus: MoM (Klochkov, Kroshnin, and Zhivotovskiy, 2021)
- **Truncation:** trimmed k-means (Cuesta-Albertos, Gordaliza, and Matrán, 1997)

Trimmed constrained k-means

Trimming operator:

$$T_\eta(\ell; P) \coloneqq \inf\left\{\int \ell\rho \,\mathrm{d}P : \rho \geq 0,\ \int \rho \,\mathrm{d}P = 1 - \eta\right\}, \quad 0 \leq \eta \leq 1$$

Given a confidence level $\delta \in (0, 1)$ and a lower bound $p_{\min} > 0$ on the mass of clusters, define a quantizer

$$\boxed{\widehat{Q}_{tr} \coloneqq \operatorname*{argmin}_{\substack{Q \subset \mathcal{X} : |Q| = k \\ P_n(V_j(Q)) \geq p_{\min}/2}} T_\eta(d^2(\cdot, Q); P_n)}$$

with $\eta \coloneqq 6\frac{\log(2/\delta)}{n}$.

Rate of convergence: finite-dimensional space

Let $\mathcal{X} = \mathbb{R}^d$. If $np_{\min} \gtrsim \log(1/\delta)$, then with probability at least $1 - \delta$

$$D(\widehat{Q}_{tr}) - D(Q^*) \lesssim D(Q^*)\Bigg((\log k)\sqrt{\frac{d + \log k}{np_{\min}}} + \sqrt{\frac{\log(1/\delta)}{np_{\min}}} \\ + (\log k)^2 \frac{d + \log k}{np_{\min}}\Bigg)$$

The same bound holds if

$$\log \mathcal{N}(B_R(x), t) \lesssim d \log \frac{R}{t}, \quad x \in \mathcal{X},\ 0 < t \le R,$$

where $\mathcal{N}(B_R(x), t)$ is the covering number of the ball $B_R(x) \subset \mathcal{X}$

Cf. bounded case: $\frac{D(Q^*)}{\sqrt{p_{\min}}}$ instead of $T^2\sqrt{k}$

Rate of convergence: finite-dimensional space

Trimmed k-means:

$$D(\widehat{Q}_{tr}) - D(Q^*) \lesssim D(Q^*)\Bigg((\log k)\sqrt{\frac{d+\log k}{np_{\min}}} + \sqrt{\frac{\log(1/\delta)}{np_{\min}}} \\ + (\log k)^2 \frac{d+\log k}{np_{\min}}\Bigg)$$

MoM k-means:

$$D(\widehat{Q}_{tr}) - D(Q^*) \lesssim \mathbb{E}\, d^2(x_0, X)\left(\sqrt{\frac{d\log k}{np_{\min}}} + \sqrt{\frac{\log(1/\delta)}{np_{\min}}}\right)$$

Rate of convergence: Hilbert space

Let $\mathcal{X}$ be a Hilbert space. Then with probability at least $1 - \delta$

$$D(\widehat{Q}_{tr}) - D(Q^*) \lesssim D(Q^*) \left(\frac{(\log n)^2}{\sqrt{np_{\min}}} + \sqrt{\frac{\log(1/\delta)}{np_{\min}}} + \frac{(\log n)^4}{np_{\min}} \right)$$

MoM k-means:

$$D(\widehat{Q}_{tr}) - D(Q^*) \lesssim \mathbb{E}\, d^2(x_0, X) \left((\log n)\sqrt{\frac{\log k}{np_{\min}}} + \sqrt{\frac{\log(1/\delta)}{np_{\min}}} \right)$$

Idea: Johnson–Lindenstrauss lemma

Rate of convergence: functional spaces

Let with some $\gamma \geq 0$, $A \geq 1$

$$\log \mathcal{N}(B_R(x), t) \leq A \left(\frac{R}{t}\right)^{\gamma} \log \frac{R}{t}, \quad x \in \mathcal{X},\ 0 < t \leq R$$

Examples: Sobolev space, Hölder space, Wasserstein space with a majorant

Then with probability at least $1 - \delta$

$$D(\widehat{Q}_{tr}) - D(Q^*) \lesssim_{\gamma} \begin{cases} D(Q^*) \left((\log k) \sqrt{\frac{A + \log k}{n p_{\min}}} + \sqrt{\frac{\log(1/\delta)}{n p_{\min}}} + \dots \right), & \gamma < 2 \\ D(Q^*) \left(\sqrt{\frac{A (\log n)^3}{n p_{\min}}} + \sqrt{\frac{\log(1/\delta)}{n p_{\min}}} + \dots \right), & \gamma = 2 \\ D(Q^*) \left(\frac{(\log n)^{1-\gamma/4}}{\sqrt{k p_{\min}}} \left(\frac{Ak}{n}\right)^{1/\gamma} + \sqrt{\frac{\log(1/\delta)}{n p_{\min}}} + \dots \right), & \gamma > 2 \end{cases}$$

Outliers

Suppose that instead of $\boldsymbol{X}$ we observe an (adversarially) contaminated sample $\boldsymbol{X}'$. If we are given an upper bound $n_o \geq |\boldsymbol{X}' \setminus \boldsymbol{X}|$, then we set

$$\eta \coloneqq 2\frac{n_o}{n} + 6\frac{\log(1/\delta)}{n}$$

$np_{\min} \gtrsim n_o \implies$ no "wiped out" cluster, the bounds hold with

$$\log(1/\delta) \mapsto \frac{n_o}{n} + \log(1/\delta)$$

Lower bounds: bounded case

Antos (2005): for any $d, k, n \in \mathbb{N}$, $k \lesssim n$, and empirical quantizer $\widehat{Q}$ there is a distribution P on $B_1(0) \subset \mathbb{R}^d$ such that

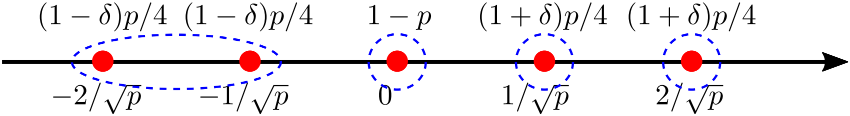
$$\mathbb{E}\, D(\widehat{Q}) - D(Q^*) \gtrsim k^{-2/d} \sqrt{\frac{k}{n}} \gtrsim D(Q^*) \sqrt{\frac{k}{n}}$$

No contradiction with upper bounds: $p_{\min} \leq \frac{1}{k}$!

Lower bounds: $p_{\min}$

Let $\mathcal{X} = \mathbb{R}$, $k = 4$. For any $\widehat{Q}$ there is a distribution P on $\mathbb{R}$ such that with probability at least $\frac{1}{4}$

$$D(\widehat{Q}) - D(Q^*) \gtrsim \frac{D(Q^*)}{\sqrt{np_{\min}}}.$$



Ingredients of the proof

- With high probability $\widehat{Q}_{tr}$ belongs to a nice class
- Bound on a squared loss for a functional class with finite L_∞-diameter

Class of quantizers

Due to the minimal cluster assumption, with high probability

$$\sum_{i=1}^{k} d^2(a_i, Q^*),\ \sum_{i=1}^{k} d^2(a_i^*, \widehat{Q}_{tr}) \lesssim \frac{T_\eta(Q^*; P_n)}{p_{\min}} \lesssim \frac{D(Q^*)}{p_{\min}}$$

where $Q^* = (a_1^*, \dots, a_k^*)$, $\widehat{Q}_{tr} = (a_1, \dots, a_k)$. Therefore, $\widehat{Q}_{tr} \in \mathcal{Q}_k$,

$$\mathcal{Q}_k \coloneqq \left\{ Q \subset \mathcal{X} : |Q| = k,\ Q \subset \bigcup_{s=1}^{k} B_{R_s}(a^*_{\pi(s)}),\ Q^* \subset \bigcup_{s=1}^{k} B_{R_s}(a_{\pi'(s)}) \right\},$$

where π, π' are permutations and

$$R_s \coloneqq C\sqrt{\frac{D(Q^*)}{s p_{\min}}}, \quad s = 1, \dots, k$$

Master bound

Let $\mathcal{F}$ be a functional class such that for some $M > 0$

$$|f - g| \leq M \quad \forall f, g \in \mathcal{F}$$

Suppose

$$\mathcal{E}_n(\mathcal{F}) \coloneqq \sup_{\boldsymbol{X}_n} \inf_{\beta > 0} \left(\beta + \frac{1}{\sqrt{n}} \int_{\beta}^{\infty} \sqrt{\log \mathcal{N}_\infty(\mathcal{F}, t, P_n)}\, \mathrm{d}t \right) < \infty$$

Then with probability at least $1 - \delta$ for all $f \in \mathcal{F}$

$$Pf^2 - Pf_*^2 \leq T_\eta(f^2; P_n) - T_\eta(f_*^2; P_n) + \sqrt{Pf_*^2} \left(\mathcal{E}_n(\mathcal{F}) + M \sqrt{\frac{\log 1/\delta}{n}} \right) + \mathcal{E}_n^2(\mathcal{F}) + M^2 \frac{\log 1/\delta}{n}$$

where $f_* \coloneqq \operatorname{argmin}_{f \in \mathcal{F}} Pf^2$

Combining ingredients

Consider

$$\mathcal{F}_k \coloneqq \{d(\cdot, Q) : Q \in \mathcal{Q}_k\}$$

Then

$$\|f - g\|_{L_\infty(P_n)} \lesssim \sqrt{\frac{D(Q^*)}{p_{\min}}}$$

$$\int_\beta^\infty \sqrt{\log \mathcal{N}_\infty(\mathcal{F}_k, t, P_n)}\, \mathrm{d}t \lesssim \sqrt{\frac{D(Q^*)}{p_{\min}}}(\log k)^{3/2} + \int_\beta^\infty \sqrt{\sum_{s=1}^k \log \mathcal{N}(B_{R_s}, t)\, \mathrm{d}t}$$

The master bound yields the result after estimating the Dudley integral $\mathcal{E}_n(\mathcal{F}_k)$

References I

- Antos, A. (2005). "Improved minimax bounds on the test and training distortion of empirically designed vector quantizers". In: *IEEE Transactions on Information Theory* 51(11), pp. 4022–4032.
- Appert, G. and O. Catoni (2021). "New bounds for k-means and information k-means". In: *arXiv preprint arXiv:2101.05728*.
- Biau, G., L. Devroye, and G. Lugosi (2008). "On the performance of clustering in Hilbert spaces". In: *IEEE Transactions on Information Theory* 54(2), pp. 781–790.
- Cadre, B. and Q. Paris (2012). "On Hölder fields clustering". In: *TEST* 21(2), pp. 301–316.
- Cuesta-Albertos, J. A., A. Gordaliza, and C. Matrán (1997). "Trimmed k-means: An attempt to robustify quantizers". In: *The Annals of Statistics* 25(2), pp. 553–576.
- Fefferman, C., S. Mitter, and H. Narayanan (2016). "Testing the manifold hypothesis". In: *Journal of the American Mathematical Society* 29(4), pp. 983–1049.

References II

Klochkov, Y., A. Kroshnin, and N. Zhivotovskiy (2021). "Robust k-means clustering for distributions with two moments". In: *The Annals of Statistics* 49(4), pp. 2206–2230.

Linder, T., G. Lugosi, and K. Zeger (1994). "Rates of convergence in the source coding theorem, in empirical quantizer design, and in universal lossy source coding". In: *IEEE Transactions on Information Theory* 40(6), pp. 1728–1740.

Hess-Schrader-Uhlenbrock inequality for the heat semigroup on differential forms over Dirichlet spaces tamed by distributional curvature lower bounds

Kazuhiro Kuwae (Fukuoka Univ.)

29 Sep. Statistics in Metric Spaces

CREST ENSAE Room 2033

Plan of talk:

(1) What is Hess-Schrader-Uhlenbrock inequality?

(2) Theory of Dirichlet spaces

(3) What is tamed Dirichlet space?

(4) Precise definition of tamed Dirichlet space

(5) Results (6) Recent known results

(7) Vector space calculus (8) Sketch of Proof

1 What is Hess-Schrader-Uhlenbrock inequality?

Hess-Schrader-Uhlenbrock ('77,'80), **Simon** ('79):

(M, g): a cpt R-mfd $\partial M = \emptyset$, $\mathrm{Ric} \geq K$

$\Longrightarrow |P_t^{\mathrm{HK}}\omega| \leq e^{-Kt}P_t|\omega|$, $\omega \in \Gamma(T^*M)$.

$P_t^{\mathrm{HK}} = e^{t\Delta^{\mathrm{HK}}}$: L^2-semigroup of de Rham-Hodge-Kodaira Laplacian $\Delta^{\mathrm{HK}} = -(\mathrm{dd}_* + \mathrm{d}_*\mathrm{d})$.

Ouhabaz ('99), **Shigekawa** ('97,'00): cpt (M, g) cvx ∂M.

Hsu ('02): cpt (M, g) ∂M, **Güneysu** ('17),

Driver-Thalmaier (01), **Elworthy-Le Jan-Li** ('99): non-cpt

2 Theory of Dirichlet spaces

$(\mathbb{D}, H^{1,2}(\mathbb{R}^d))$: **classical Dirichlet integral:**

$$\begin{aligned} \mathbb{D}(f,g) &= \int_{\mathbb{R}^d} \langle \nabla f(x), \nabla g(x) \rangle \mathrm{d}x, \qquad && f,g \in H^{1,2}(\mathbb{R}^d) \\ &= \int_{\mathbb{R}^d} (-\Delta f(x)) g(x) \mathrm{d}x \qquad && f,g \in C_c^2(\mathbb{R}^d). \end{aligned}$$

$\mathrm{X} = (\Omega, B_t, \mathsf{P}_x)$: **Brownian motion on** $\mathbb{R}^d$**:**

$$\mathsf{P}_x(B_t \in A) = \int_A p_t(x,y)\mathrm{d}y = \int_A \frac{1}{(2\pi t)^{d/2}} e^{-\frac{|x-y|^2}{2t}} \mathrm{d}y$$

$$= T_t 1_A(x) := \text{``}e^{t\Delta/2} 1_A(x)\text{''}$$

$$: L^2\text{-semigroup ass. to } \left(\frac{1}{2}\mathbb{D}, H^{1,2}(\mathbb{R}^d)\right).$$

$(\mathcal{E}, D(\mathcal{E}))$: Dirichlet form on $L^2(M;\mathfrak{m})$ iff

(i) non-negative symmetric bilinear form on $L^2(M;\mathfrak{m})$, whose domain $D(\mathcal{E})$ is desely defined in $L^2(M;\mathfrak{m})$

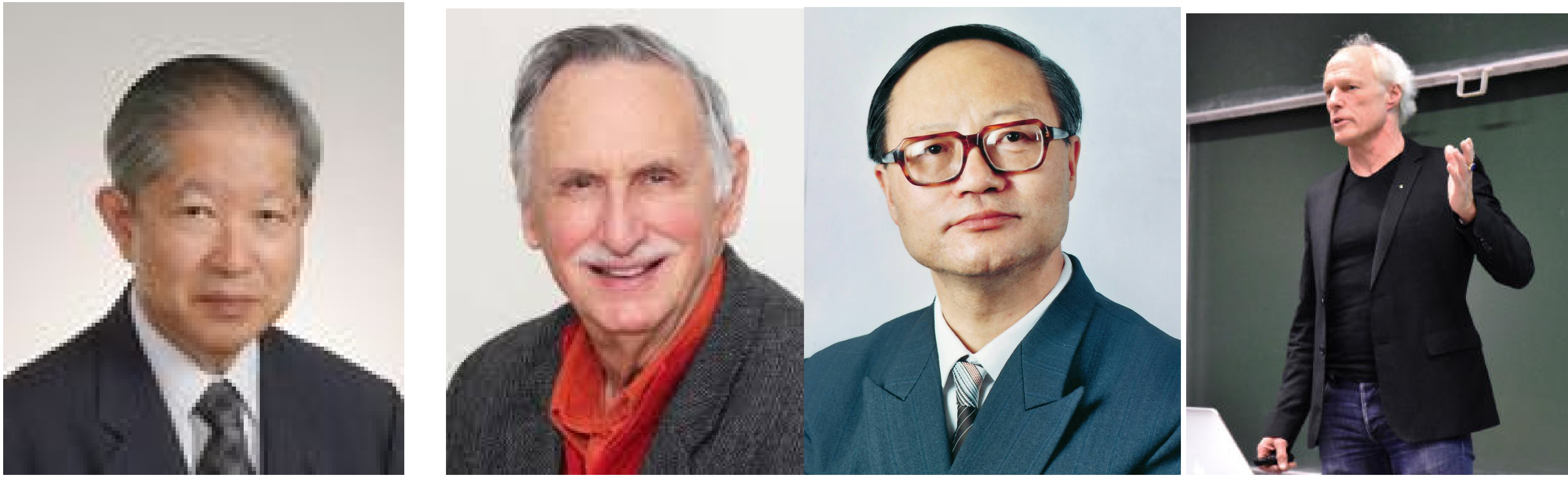
(ii) $D(\mathcal{E})$ is complete w.r.t. $\mathcal{E}_1^{1/2}$-norm, where $\mathcal{E}_1(f,g) := \mathcal{E}(f,g) + (f,g)_{\mathfrak{m}}$ for $f, g \in D(\mathcal{E})$.

(iii) For $f \in D(\mathcal{E})$, $f^\sharp := 0 \vee f \wedge 1 \in D(\mathcal{E})$ & $\mathcal{E}(f^\sharp, f^\sharp) \leq \mathcal{E}(f,f)$.

If $(\mathcal{E}, D(\mathcal{E}))$ is (quasi-)regular, ${}^\exists\mathrm{X} = (\Omega, X_t, \mathsf{P}_x)$ s.t. $T_t f(x) = \mathsf{E}_x[f(X_t)]$ $\mathfrak{m}$-a.e. for $f \in L^2(M;\mathfrak{m}) \cap \mathcal{B}(M)$.

Fukushima ('76), Albeverio-Ma-Röckner ('91,'92,'93)

(From Wikipedia & Personal HP)



See Fukushima-Oshima-Takeda ('10), Oshima ('13),

Ma-Röckner ('92)

3 What is tamed Dirichlet space?

Roughly speaking, tamed Dirichlet space is a

- strongly local Dirichlet space $(\mathcal{E}, D(\mathcal{E}))$ on $L^2(M;\mathfrak{m})$ $\Leftrightarrow \mathrm{X} = (\Omega, X_t, \mathrm{P}_x)$: $\mathfrak{m}$-sym. diffusion process on M.
- $(\mathcal{E}, D(\mathcal{E}))$ has a lower bound κ of Ricci curvature in distribution sense: weak Bakry-Émery condition.
- $\kappa := \kappa^+ - \kappa^-$: signed measure s.t. κ^+ has bounded potential $U_1\kappa^+$, $2\kappa^-$ is of (extended) Kato class.

The notion of tamed Dirichlet space was proposed by Erbar-Rigoni-Sturm-Tamanini ('22) and its vector space calculus was developed by Braun ('22+):

(From Wikipedia & Personal HP)

Very nice framework!, but sub-Riem. mfds, Φ_2^4-model, super process of immigration models are not included in.

4 Precise definition of tamed Dirichlet space

(M,τ): top. Lusin space

$\mathfrak{m}$: σ-finite Borel measure with full support

$(f,g)_{\mathfrak{m}}$: L^2-inner product

$(\mathcal{E},D(\mathcal{E}))$: strongly local quasi-regular

Dirichlet form on $L^2(M;\mathfrak{m})$

$(P_t)_{t\geq 0}$: Markov L^2-semigroup $\Leftrightarrow$ $(\mathcal{E},D(\mathcal{E}))$

$\mathbf{X}=(\Omega,X_t,\mathbb{P}_x)$: $\mathfrak{m}$-sym diffusion process s.t.

$P_tf=\mathbb{E}_{\bullet}[f(X_t)]$ $\mathfrak{m}$-a.e. for $f\in L^2(M;\mathfrak{m})\cap\mathcal{B}(M)$

$\mu_{\langle f,g\rangle} = \Gamma(f,g)\mathrm{d}\mathfrak{m}$, $f, g \in D(\mathcal{E})$: **(signed finite) energy measure** $\mathcal{E}(f,g) = \mu_{\langle f,g\rangle}(M) = \int_M \Gamma(f,g)\mathrm{d}\mathfrak{m}$.

$\kappa \in S(\mathrm{X})$, $\kappa = \kappa^+ - \kappa^-$; **Jordan-Hahn decomposition.**

$$\left\| \mathbf{E}_{\cdot}\left[A_t^{\kappa^+} \right] \right\|_\infty < \infty \ {}^{\exists/\forall} t > 0 \iff \kappa^+ \in S_D(\mathrm{X}).$$

$$\lim_{t\to 0} \left\| \mathbf{E}_{\cdot}\left[A_t^{\kappa^-} \right] \right\|_\infty < \frac{1}{2} \iff 2\kappa^- \in S_{EK}(\mathrm{X}).$$

Define $(\mathcal{E}^{2\kappa}, D(\mathcal{E}^{2\kappa}))$ **by**

$$\mathcal{E}^{2\kappa}(f,g) := \mathcal{E}(f,g) + 2\int_M \tilde{f}\tilde{g}\mathrm{d}\kappa,$$

$$f, g \in D(\mathcal{E}^{2\kappa}) = D(\mathcal{E}).$$

Then this is a closed bilinear form bounded below s.t.

$^{\exists}\alpha_0 > 0$**,** $C > 0$

$$C^{-1}\mathcal{E}_1(f) \leq \mathcal{E}^{2\kappa}_{\alpha_0}(f) \leq C\mathcal{E}_1(f) \quad \textbf{for all} \quad f \in D(\mathcal{E}).$$

Here $\mathcal{E}^{2\kappa}_{\alpha_0}(f,g) = \mathcal{E}^{2\kappa}(f,g) + \alpha_0(f,g)_{\mathfrak{m}}$ **and**

$\mathcal{E}_1(f,g) = \mathcal{E}(f,g) + (f,g)_{\mathfrak{m}}$**.**

Consider a CAF $A_t^{\kappa} := A_t^{\kappa^+} - A_t^{\kappa^-} (= Kt$ if $\kappa = K\mathfrak{m})$ and Feynman-Kac semi-group $(p_t^{2\kappa})_{t\geq 0}$ by

$$p_t^{2\kappa} f(x) := \mathbf{E}_x[e^{-2A_t^{\kappa}} f(X_t)], \quad f \in \mathcal{B}_b(M).$$

Then $(p_t^{2\kappa} f, g)_{\mathfrak{m}} = (f, p_t^{2\kappa} g)_{\mathfrak{m}}$, $f, g \in \mathcal{B}_+(M)$. Moreover, $(p_t^{2\kappa})_{t\geq 0}$ coincides with $(P_t^{2\kappa})_{t\geq 0}$ on $L^2(M; \mathfrak{m})$ associated to $(\mathcal{E}^{2\kappa}, D(\mathcal{E}^{2\kappa}))$. Under such conditions, the stochastic semi-group $(p_t^{\kappa})_{t\geq 0}$ can be extended a semi-group P_t^{κ} on $L^p(M; \mathfrak{m})$ for each $p \in [1, +\infty]$. Let $\Delta^{2\kappa}$ be an L^2-generator associated to $(\mathcal{E}^{2\kappa}, D(\mathcal{E}^{2\kappa}))$.

Def 4.1 (Tamed Dirichlet space, i.e. BE$_2(\kappa, N)$)

κ^+, κ^- as defined before. Fix $N \in [1, +\infty]$.

$(M, \mathcal{E}, \mathfrak{m})$ or M is said to satisfy 2-Bakry-Émery condition (BE$_2(\kappa, N)$ in short), if the following holds:

For ${}^\forall f \in D(\Delta)$ with $\Delta f \in D(\mathcal{E})$ & ${}^\forall \phi \in D(\Delta^{2\kappa}) \cap L^\infty(M; \mathfrak{m})_+$ with $\Delta^{2\kappa}\phi \in L^\infty(M; \mathfrak{m})$,

$$\frac{1}{2}\int_M \Gamma(f)\Delta^{2\kappa}\phi \, d\mathfrak{m} - \int_M \phi\Gamma(f, \Delta f) \, d\mathfrak{m} \geq \frac{1}{N}\int_M \phi(\Delta f)^2 \, d\mathfrak{m}.$$

When $N = +\infty$, the right-hand vanishes.

κ^+, κ^- as defined before. $(M, \mathcal{E}, \mathfrak{m})$ or simply M satisfies BE$_2(\kappa, N)$, we call $(\mathcal{E}, D(\mathcal{E}))$ Tamed Dirichlet space.

Thm 4.1 (Erbar-Rigoni-Sturm-Tamanini ('22))

κ^+, κ^- as defined before. Then BE$_2(\kappa, \infty) \Leftrightarrow$ GE$_1(\kappa, \infty)$.

$$\text{GE}_1(\kappa, \infty): \quad \sqrt{\Gamma(P_t f)} \leq P_t^\kappa \sqrt{\Gamma(f)}, \quad f \in D(\mathcal{E}). \quad (1)$$

Here P_t^κ associates to $p_t^\kappa h(x) := \mathbf{E}_x[e^{-A_t^\kappa} h(X_t)]$. Moreover, the following $\mathrm{Test}(M)$ forms an algebra.

$$\mathrm{Test}(M) := \{f \in D(\Delta) \cap L^\infty(M; \mathfrak{m}) \mid$$
$$\Gamma(f) \in L^\infty(M; \mathfrak{m}), \Delta f \in D(\mathcal{E})\}. \quad (2)$$

Ex 4.1 (Examples of Tamed Dirichlet spaces)

- RCD(K, N)-spaces,
- Abstract Wiener space (B, H, μ),
- cpt R-manifolds with boundary, **ERST** ('22),
- Almost smooth metric measure space with BE$_2$-condition, This is not an RCD-space, **Honda** ('18),
- Infinite particle systems on (M, g) with $\mathrm{Ric} \geq K$ without interaction under Poisson measure: **Albeverio-Kondratiev-Röckner** ('98), **Dello Schiavo-Suzuki** ('22+).

5 Results

Thm **5.1 (Hess-Schrader-Uhlenbrock inequality)**

We have the following: Recall $p_t^{\kappa}h(x) = \mathbf{E}_x[e^{-A_t^{\kappa}}h(X_t)]$.

(1) For ${}^{\forall}\omega \in L^2(T^*M)$ and $\alpha > C_{\kappa}$,

$$|R_{\alpha}^{\mathrm{HK}}\omega| \leq R_{\alpha}^{\kappa}|\omega| \quad \mathfrak{m}\text{-a.e.} \tag{3}$$

(2) For ${}^{\forall}\omega \in L^2(T^*M)$ and every $t \geq 0$,

$$|P_t^{\mathrm{HK}}\omega| \leq P_t^{\kappa}|\omega| \quad \mathfrak{m}\text{-a.e.} \tag{4}$$

Cor 5.1 (C_0-property of $(P_t^{\mathrm{HK}})_{t\geq 0}$ on $L^p(M;\mathfrak{m})$)

Suppose $\underline{p\in[2,+\infty]}$, or $\underline{\kappa^-\in S_K(\mathrm{X})\text{ and }p\in[1,+\infty]}$. Then the heat flow $(P_t^{\mathrm{HK}})_{t\geq 0}$ can be extended to a semigroup on $L^p(T^*M)$ and and for each $t>0$

$$\|P_t^{\mathrm{HK}}\omega\|_{L^p(T^*M)}\leq C(\kappa)e^{C_\kappa t}\|\omega\|_{L^p(T^*M)},\quad \omega\in L^p(T^*M).$$

Moreover, if $\underline{\kappa^-\in S_K(\mathrm{X})\text{ and }p\in[1,+\infty[}$, then $(P_t^{\mathrm{HK}})_{t\geq 0}$ is strongly continuous on $L^p(T^*M)$, i.e., $(P_t^{\mathrm{HK}})_{t\geq 0}$ is a C_0-semigroup on $L^p(T^*M)$, and further $(P_t^{\mathrm{HK}})_{t\geq 0}$ is weakly* continuous on $L^\infty(T^*M)$.

Thm 5.2 (**Esaki-Xu-K** (23+)) The Riesz operator $R_\alpha(\Delta)$ defined by $R_\alpha(\Delta)f := \Gamma((\alpha - \Delta)^{-\frac{1}{2}}f)^{\frac{1}{2}}$ is bounded on $L^p(X;\mathfrak{m})$ under $\kappa^- \in S_K(\mathrm{X})$ and $p \in [2,+\infty[$.

Ex 5.1 (New examples)

- a class of R-mfd with boundary s.t. $\kappa = k\mathfrak{v} + \ell\sigma$, $\mathfrak{v} := \mathrm{vol}_g$, σ: surface measure on ∂M of Kato, $\mathrm{Ric} \geq k$: Kato function, ℓ is a lower bounds of second fundamental form on ∂M.
- Configuration space $(\Upsilon, \mathcal{E}^\Upsilon, \pi)$ without interactions over (M, g) having $\mathrm{Ric} \geq K$.

6 Recent known results

Braun ('22): **Thm** 5.1 & **Cor** 5.1 are proved for $\mathrm{RCD}(K,\infty)$. Note that abstract Wiener space (B,H,μ) (satisfying $\mathrm{CD}(1,\infty)$ by **Fang-Shao-Sturm** ('10)) is not an $\mathrm{RCD}(1,\infty)$!, so not included in this setting.

Braun ('22+): **Thm** 5.1 is proved for tamed Dirichlet space under the that ${}^{\exists}k \in L^1_{\mathrm{loc}}(M;\mathfrak{m})$ s.t. $\kappa = k\mathfrak{m}$ and $|k|\mathfrak{m} \in S_{EK}(\mathrm{X})$ and ${}^{\exists}K \in \mathbb{R}$ s.t. $k \geq K$ on M.

(B,H,μ) is included in this setting.

7 Vector space calculus over tamed space

Vector space calculus was established by **Braun** ('22+), which was a natural extension of the vector space calculus for RCD-space developed by **Gigli** ('18).

The proof for $\kappa = K\mathfrak{m}$ is easy. New point is that κ is not necessarily of constant nor of function! This causes another technical difficulty.

So you can follow the proof below for

(M,g): smooth Riemmannian mfd with $\partial M\neq\emptyset$

$n := \dim(M)$, $\mathfrak{v} = \mathrm{vol}_g$, $\mathrm{Ric}_x \geq k(x)g_x$

$k(x)$: Kato class function on M

σ: surface measure on ∂M of Kato class

ℓ: lower bound of second fundamental form

$\Longrightarrow$ $\mathsf{BE}_2(\kappa, n)$ with $\kappa = k\mathfrak{v} + \ell\sigma$.

But this concrete expression is not so important in the proof. Essential point is the (extended) Kato class condition for $2\kappa^-$!

8 Sketch of proof

Lem 8.1 (**Braun** ('22+)) For $X \in H^{1,2}(TM)$, $|\nabla|X|| \le |\nabla X|_{\mathrm{HS}}$: Kato's inequality.

Lem 8.2 $X \in H^{1,2}(TM)$ implies $|X| \in D(\mathcal{E})$ and

$$\mathcal{E}(|X|, |X|) \le \widetilde{\mathcal{E}}_{\mathrm{cov}}(X, X) < \infty. \tag{5}$$

For $f \in D(\mathcal{E}) \cap L^\infty(M; \mathfrak{m})_+$ with $fX \in H^{1,2}(TM)$, we have $f|X| \in D(\mathcal{E})$ and

$$\mathcal{E}(|X|, f|X|) \le \widetilde{\mathcal{E}}_{\mathrm{cov}}(X, fX). \tag{6}$$

Proof. The proof of (5) can be directly deduced from **Lem** 9.1. Next we show (6). Assume $fX \in H^{1,2}(TM)$ for $f \in D(\mathcal{E}) \cap L^\infty(M;\mathfrak{m})_+$. By (5), we have $f|X| \in D(\mathcal{E})$. Moreover, due to **Braun** ('22+),

$$|P_t^B X| \le P_t|X| \quad \mathfrak{m}\text{-a.e.}$$

we have

$$((I - P_t)|X|, f|X|)_{L^2(M;\mathfrak{m})} \le ((I - P_t^{\mathrm{B}})X, fX)_{L^2(TM)}.$$

Divided by $t > 0$ and letting $t \to 0$, we obtain (6). $\square$

Lem 8.3 Take $\omega \in H^{1,2}(T^*M)(= D(\mathcal{E}^{\mathrm{HK}}))$. Then $|\omega| \in D(\mathcal{E})$ and

$$\mathcal{E}^{\kappa}(|\omega|, |\omega|) \le \mathcal{E}^{\mathrm{HK}}(\omega, \omega). \tag{7}$$

Proof. $\omega \in H^{1,2}(T^*M)$ implies $\omega^\sharp \in H^{1,2}(TM)$ and $|\omega| = |\omega^\sharp| \in D(\mathcal{E})$. Then

$$\begin{aligned}\mathcal{E}(|\omega|, |\omega|) = \mathcal{E}(|\omega^\sharp|, |\omega^\sharp|) &\overset{(5)}{\le} \widetilde{\mathcal{E}}_{\mathrm{cov}}(\omega^\sharp, \omega^\sharp) \\ &\le \mathcal{E}^{\mathrm{HK}}(\omega, \omega) - \langle \kappa, |\omega|^2 \rangle,\end{aligned}$$

which implies the conclusion. The last inequality is due to **Braun** ('22+). ☐

Lem 8.4

(i) $\omega \in H^{1,2}(T^*M) \cap L^\infty(T^*M)$ & $f \in D(\mathcal{E}) \cap L^\infty(M;\mathfrak{m})$

$\Rightarrow f\omega \in H^{1,2}(T^*M) \cap L^\infty(T^*M)$.

(ii) $\omega \in H^{1,2}(T^*M)$ and $f \in \mathrm{Test}(M) \Rightarrow f\omega \in H^{1,2}(T^*M)$.

Proof. The following are due to **Braun** ('22+):

(i) $X \in H^{1,2}(TM) \cap L^\infty(TM)$ & $f \in D(\mathcal{E}) \cap L^\infty(M;\mathfrak{m})$

$\Rightarrow fX \in H^{1,2}(TM) \cap L^\infty(TM)$.

(ii) $X \in H^{1,2}(TM)$ & $f \in \mathrm{Test}(M) \Rightarrow fX \in H^{1,2}(TM)$

$\square$

Lem 8.5 Take $\omega \in H^{1,2}(T^*M)$ and $f \in \mathrm{Test}(M)_+$. Then $f|\omega| \in D(\mathcal{E}^\kappa) = D(\mathcal{E})$ and

$$\mathcal{E}^\kappa(|\omega|, f|\omega|) \leq \mathcal{E}^{\mathrm{HK}}(\omega, f\omega). \qquad (8)$$

Proof. By **Lem** 8.4, we have $f\omega \in H^{1,2}(T^*M)$ & $f|\omega| \in D(\mathcal{E})$. By **Braun** ('22+), we have

$$\mathrm{Ric}(\omega^\sharp, f\omega^\sharp)(M) = \mathcal{E}^{\mathrm{HK}}(\omega, f\omega) - \widetilde{\mathcal{E}}_{\mathrm{cov}}(\omega^\sharp, f\omega^\sharp). \quad (9)$$

Here the LHS is the total mass of Ricci curvature measure defined by **Braun** ('22+).

By **Lem** 8.2, we then have

$$\begin{aligned}
\mathcal{E}^{\kappa}(|\omega|, f|\omega|) &= \mathcal{E}(|\omega|, f|\omega|) + \langle \boldsymbol{\kappa}, f|\omega|^2 \rangle \\
&\leq \widetilde{\mathcal{E}}_{\mathrm{cov}}(\omega^\sharp, f\omega^\sharp) + \langle \boldsymbol{\kappa}, f|\omega|^2 \rangle \\
&\overset{(9)}{=} \mathcal{E}^{\mathrm{HK}}(\omega, f\omega) - \mathrm{Ric}^{\kappa}(\omega^\sharp, f\omega^\sharp)(M) \\
&= \mathcal{E}^{\mathrm{HK}}(\omega, f\omega) - \int_M \widetilde{f} \mathrm{d}\mathrm{Ric}^{\kappa}(\omega^\sharp, \omega^\sharp) \\
&\leq \mathcal{E}^{\mathrm{HK}}(\omega, f\omega).
\end{aligned}$$

Cor 8.1 Take $\omega \in D(\boldsymbol{\Delta}^{\mathrm{HK}}) \cap L^\infty(T^*M)$ and $f \in D(\mathcal{E}) \cap L^\infty(M; \mathfrak{m})_+$. Then $f\omega \in H^{1,2}(T^*M)$, $f|\omega| \in D(\mathcal{E}) \cap L^\infty(M; \mathfrak{m})$, and (8) hold.

Lem 8.6 Take $\omega \in D(\mathbb{\Delta}^{\mathrm{HK}}) \cap L^\infty(T^*M)$. Then

$$\left(-\mathbb{\Delta}^{\mathrm{HK}}\omega, g\frac{\omega}{|\omega|}\right)_{L^2(T^*M)} \geq \mathcal{E}^\kappa(|\omega|, g) \quad {}^\forall g \in \mathrm{Test}(M)_+. \tag{10}$$

Here we set $\omega/|\omega| := 0$ if $\omega = 0$.

Proof. By **Lem** 8.3, we see $|\omega| \in D(\mathcal{E}) \cap L^\infty(M; \mathfrak{m})$. For each $\varepsilon > 0$, we set $|\omega|_\varepsilon := \sqrt{|\omega|^2 + \varepsilon^2}$. Then we see $\frac{|\omega|}{|\omega|_\varepsilon} \in D(\mathcal{E}) \cap L^\infty(M; \mathfrak{m})$. For $g \in \mathrm{Test}(M)_+$, we set $f := g/|\omega|_\varepsilon \in D(\mathcal{E}) \cap L^\infty(M; \mathfrak{m})_+$.

We apply **Cor** 8.1 for $f \in D(\mathcal{E}) \cap L^\infty(M;\mathfrak{m})_+$ so that $f\omega \in H^{1,2}(T^*M)$, $f|\omega| \in D(\mathcal{E})$ and

$$\left(-\mathbb{\Delta}^{\mathrm{HK}}\omega, g\frac{\omega}{|\omega|_\varepsilon}\right)_{L^2(T^*M)} \geq \int_M \frac{|\omega|}{|\omega|_\varepsilon}\Gamma(|\omega|, g)\mathrm{d}\mathfrak{m} + \left\langle \kappa, g\frac{|\omega|^2}{|\omega|_\varepsilon}\right\rangle.$$

Letting $\varepsilon \to 0$, we obtain the conclusion.

Lem 8.7 Suppose $\omega \in H^{1,2}(T^*M)$. Then

$$\int_0^t \int_M |\widetilde{P_s^{\mathrm{HK}}\omega}|^2 \mathrm{d}\kappa^- \mathrm{d}s < \infty. \tag{11}$$

Lem 8.8 For ${}^\forall\omega \in L^2(T^*M)$ and $t > 0$, we have

$$|P_t^{\mathrm{HK}}\omega|^2 \leq P_t^{-2\kappa^-}|\omega|^2 \quad \mathfrak{m}\text{-a.e.} \tag{12}$$

In particular, for $\omega \in L^2(T^*M) \cap L^\infty(T^*M)$ and $\alpha > C_\kappa$,

$$\|R_\alpha^{\mathrm{HK}}\omega\|_{L^\infty(T^*M)} \leq \frac{\sqrt{C(\kappa)}}{\alpha - C_\kappa}\|\omega\|_{L^\infty(T^*M)}. \tag{13}$$

Hence, $R_\alpha^{\mathrm{HK}}\omega \in L^\infty(T^*M)$ for $\omega \in L^2(T^*M) \cap L^\infty(T^*M)$.

Proof. We may assume $\kappa^+ = 0$, because BE$_2(-\kappa^-, \infty)$ is satisfied. We may assume $\omega \in H^{1,2}(T^*M)$.

Take $g \in \mathrm{Test}(M)_+$. and set $g_\alpha R_\alpha g$.

We now set a function $F_n : [0,t] \to \mathbb{R}$ defined by

$$F_n(s) := \int_M P^{2\kappa}_{t-s} g_\alpha \, n P_{\frac{1}{n}} G_n |P_s^{\mathrm{HK}} \omega|^2 \mathrm{d}\mathfrak{m}.$$

After a long calculation,

$$\begin{aligned} \frac{1}{n} F_n'(s) \le\ & 2 \int_M p_{\frac{1}{n}} R_n p^{2\kappa}_{t-s} g_\alpha \, |\widetilde{P_s^{\mathrm{HK}} \omega}|^2 \mathrm{d}\kappa^- \\ & - 2 \int_M (p^{2\kappa}_{t-s} g_\alpha) \, p_{\frac{1}{n}} R_n |\widetilde{P_s^{\mathrm{HK}} \omega}|^2 \mathrm{d}\kappa^-. \end{aligned} \tag{14}$$

$$\begin{aligned} \overline{\lim_{n\to\infty}} F_n'(s) \le\ & 2 \int_M p^{2\kappa}_{t-s} g_\alpha \, |\widetilde{P_s^{\mathrm{HK}} \omega}|^2 \mathrm{d}\kappa^- \\ & - \underline{\lim_{n\to\infty}} \, 2 \int_M p^{2\kappa}_{t-s} g_\alpha \, n p_{\frac{1}{n}} R_n |\widetilde{P_s^{\mathrm{HK}} \omega}|^2 \mathrm{d}\kappa^- \\ \le\ & 0. \end{aligned} \tag{15}$$

By way of Monotone convergence theorem for $f_n(s) := \inf_{\ell \geq n} \left(-F_\ell'(s)\right)$, we get

$$\varlimsup_{n\to\infty} \int_0^t F_n'(s)\mathrm{d}s \leq 0.$$

Thus,

$$\begin{aligned}\varlimsup_{n\to\infty} \Bigg(\int_M g_\alpha\, nP_{\frac{1}{n}}G_n|P_t^{\mathrm{HK}}\omega|^2\mathrm{d}\mathfrak{m} \\ - \int_M (P_{t-s}^{2\kappa}g_\alpha)\, nP_{\frac{1}{n}}G_n|\omega|^2\mathrm{d}\mathfrak{m}\Bigg) \\ \leq \varlimsup_{n\to\infty} \int_0^t F_n'(s)\mathrm{d}s \leq 0.\end{aligned}$$

Since $\lim_{n\to\infty}\|P_{\frac{1}{n}}nG_nf-f\|_{L^1(M;\mathfrak{m})}=0$ for $f\in L^1(M;\mathfrak{m})$, we have

$$\int_M g_\alpha|P_t^{\mathrm{HK}}\omega|^2\mathrm{d}\mathfrak{m}\le\int_M(P_t^{2\kappa}g_\alpha)|\omega|^2\mathrm{d}\mathfrak{m}=\int_M g_\alpha P_t^{2\kappa}|\omega|^2\mathrm{d}\mathfrak{m}.$$

Since $\alpha g_\alpha=\alpha R_\alpha g\in L^\infty(M;\mathfrak{m})$ weakly* converges to g in $L^\infty(M;\mathfrak{m})$ as $\alpha\to\infty$ and $g\in L^2(M;\mathfrak{m})\cap L^\infty(M;\mathfrak{m})\cap\mathcal{B}_+(M)$ is arbitrary, we obtain (12).

Proof. of **Thm** 5.1 Take $g \in \mathrm{Test}(M)_+$. Then, for $\omega \in D(\boldsymbol{\Delta}^{\mathrm{HK}}) \cap L^\infty(T^*M)$

$$\begin{aligned}\int_M (\Delta g - \alpha g)|\omega|\,\mathrm{d}\mathfrak{m} - \int_M g|\omega|\,\mathrm{d}\kappa &\\ \geq \int_M g \left(\left\langle \boldsymbol{\Delta}^{\mathrm{HK}}\omega, \frac{\omega}{|\omega|} \right\rangle - \alpha|\omega| \right) \mathrm{d}\mathfrak{m} &\\ = \int_M g \left\langle \boldsymbol{\Delta}^{\mathrm{HK}}\omega - \alpha\omega, \frac{\omega}{|\omega|} \right\rangle \mathrm{d}\mathfrak{m} &\\ \geq - \int_M g \left| \boldsymbol{\Delta}^{\mathrm{HK}}\omega - \alpha\omega \right| \mathrm{d}\mathfrak{m} &\end{aligned}$$

by **Lem** 8.6, hence

$$\mathcal{E}^\kappa_\alpha(|\omega|, g) \leq \int_M g \left| (\alpha - \boldsymbol{\Delta}^{\mathrm{HK}})\omega \right| \mathrm{d}\mathfrak{m}. \qquad (16)$$

Since $\mathrm{Test}(M)_+$ is dense in $D(\mathcal{E})_+$, (16) holds for any $g \in D(\mathcal{E})_+$. By Lem 8.8, for $\alpha > C_\kappa$, we can set $\omega := \boldsymbol{R}_\alpha^{\mathrm{HK}}\eta \in D(\boldsymbol{\Delta}^{\mathrm{HK}}) \cap L^\infty(T^*M)$ for $\eta \in L^2(T^*M) \cap L^\infty(T^*M)$ and $g := R_\alpha^\kappa\psi$ with $\psi \in L^2(M;\mathfrak{m})_+$. Then we see

$$(\psi, |\boldsymbol{R}_\alpha^{\mathrm{HK}}\eta|)_{\mathfrak{m}} \le (\psi, \boldsymbol{R}_\alpha^{\kappa}|\eta|)_{\mathfrak{m}} \quad \text{for any} \quad \psi \in L^2(M;\mathfrak{m})$$

This implies that for $\alpha > C_\kappa$ and $\eta \in L^2(T^*M) \cap L^\infty(T^*M)$

$$|\boldsymbol{R}_\alpha^{\mathrm{HK}}\eta| \le \boldsymbol{R}_\alpha^{\kappa}|\eta| \quad \mathfrak{m}\text{-a.e.} \tag{17}$$

By approximation, we can deduce that (17) holds for general $\eta \in L^2(T^*M)$.

From (17), we can obtain that $|P_t^{\mathrm{HK}}\eta| \le P_t^{\kappa}|\eta|$ $\mathfrak{m}$-a.e. for each $t > 0$ in view of the following observation:

$$P_t^{\kappa} f = \lim_{n\to\infty}\left(\frac{n}{t}\right)^n (R^{\kappa}_{\frac{n}{t}})^n f, \qquad f \in L^2(M;\mathfrak{m}),$$

$$P_t^{\mathrm{HK}}\theta = \lim_{n\to\infty}\left(\frac{n}{t}\right)^n (R^{\mathrm{HK}}_{\frac{n}{t}})^n \theta \qquad \theta \in L^2(T^*M).$$

Thank you for your attention.

Nous vous remercions de votre attention.

Vielen Dank für Ihre Aufmerksamkeit.

9 Vector space calculus over tamed space

In this section, we summarize the results by **Braun** ('22+). This was a natural extension of the vector space calculus for RCD-space developed by **Gigli** ('18).

Def 9.1 (L^p-normed L^∞-module)

Given $p \in [1, +\infty]$, a real Banach space $(\mathcal{M}, \|\cdot\|_{\mathcal{M}})$, or simply, $\mathcal{M}$ is called an L^p-normed L^∞-module (over $(M, \mathfrak{m})$) if it satisfies

(a) a bilinear map $\cdot : L^\infty(M;\mathfrak{m}) \times \mathcal{M} \to \mathcal{M}$ satisfying

$$(fg) \cdot v = f \cdot (gv),$$

$$1_M \cdot v = v,$$

(b) a nonnegatively valued map $|\cdot|_{\mathfrak{m}} : \mathcal{M} \to L^p(M;\mathfrak{m})$ s.t.

$$|f \cdot v|_{\mathfrak{m}} = |f||v|_{\mathfrak{m}} \quad \mathfrak{m}\text{-a.e.},$$

$$\|v\|_{\mathcal{M}} = \||v|_{\mathfrak{m}}\|_{L^p(M;\mathfrak{m})},$$

for $^\forall f, g \in L^\infty(M;\mathfrak{m})$ and $v \in \mathcal{M}$. If only (a) is satisfied, we call $(\mathcal{M}, \|\cdot\|_{\mathcal{M}})$ or simply $\mathcal{M}$ an $L^\infty(M;\mathfrak{m})$-module.

We always assume that for $^\forall v \in \mathcal{M}$, $|v|_{\mathfrak{m}}$ is Borel. $\mathcal{M}$ is called Hilbert module if is an L^2-normed L^∞-module, in this case, the point-wise norm $|\cdot|_{\mathfrak{m}}$ induces a point-wise scalar product $\langle\cdot,\cdot\rangle_{\mathfrak{m}}:\mathcal{M}\times\mathcal{M}\to L^1(M;\mathfrak{m})$ which is L^∞-bilinear, $\mathfrak{m}$-a.e. nonnegative definite, local in both components, satisfies the point-wise $\mathfrak{m}$-a.e. Cauchy-Schwar inequality.

Def 9.2 (Dual module)

We can define the dual module $\mathcal{M}^*$ by

$$\mathcal{M}^* := \mathrm{Hom}(\mathcal{M}, L^1(M; \mathfrak{m}))$$

and will be endowed with the usual operator norm. The point-wise paring between $v \in \mathcal{M}$ and $L \in \mathcal{M}^*$ is denoted by $L(v) \in L^1(M; \mathfrak{m})$. If $\mathcal{M}$ is L^p-normed, then $\mathcal{M}^*$ is an L^q-normed L^∞-module, where $p, q \in [1, +\infty]$ with $1/p + 1/q = 1$.

Def 9.3 (Tensor products) Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two Hilbert module. We can define the tensor product $\mathcal{M}_1 \otimes \mathcal{M}_2$ the $\|\cdot\|_{\mathcal{M}_1\otimes\mathcal{M}_2}$-completion of the subspace that consists of all $A \in \mathcal{M}_1^0 \odot \mathcal{M}_2^0$ s.t. $\|A\|_{\mathcal{M}_1\otimes\mathcal{M}_2} < \infty$. Here $\mathcal{M}_i^0$ $(i = 1, 2)$ is the L^0-module induced by $\mathcal{M}_i$ and $\mathcal{M}_1^0 \odot \mathcal{M}_2^0$ is the algebraic tensor product.

Def 9.4 (Exterior product) The exterior product $\Lambda\mathcal{M}$ is defined as the completion w.r.t. $\|\cdot\|_{\Lambda\mathcal{M}}$ of the subspace consisting of all $\omega \in \Lambda\mathcal{M}^0$ s.t. $\|\omega\|_{\Lambda\mathcal{M}} < \infty$.

Let $(\mathcal{E}, D(\mathcal{E}))$ be a quasi-regular strongly local Dirichlet form on $L^2(M;\mathfrak{m})$. We define the cotangent module $L^2(T^*M)$, i.e., the space of differential 1-forms that are L^2-integrable in a certain "universal" sense.

Def 9.5 (Pre-cotangent module)

We define the pre-cotangent module Pcm by

$$\mathrm{Pcm} := \left\{ (f_i, A_i)_{i\in\mathbb{N}} \;\middle|\; (A_i)_{i\in\mathbb{N}} \text{ Borel partition of } M, \right. \\ \left. (f_i)_{i\in\mathbb{N}} \subset D(\mathcal{E})_e, \sum_{i\in\mathbb{N}} \int_{A_i} \Gamma(f_i)\mathrm{d}\mathfrak{m} < \infty \right\}$$

Moreover, we define a relation $\sim$ on Pcm by $(f_i, A_i)_{i\in\mathbb{N}} \sim (g_j, B_j)_{j\in\mathbb{N}}$ if and only if $\int_{A_i\cap B_j} \Gamma(f_i - g_j)\mathrm{d}\mathfrak{m} = 0$ for ${}^\forall i, j \in \mathbb{N}$. The relation, in fact forms an equivalence relation. The equivalence class of an element $(f_i, A_i)_{i\in\mathbb{N}} \in \mathrm{Pcm}$ w.r.t. $\sim$ is denoted by $[f_i, A_i]$. The space $\mathrm{Pcm}/\sim$ of equivalence classes becomes a vector space via the well-defined operations

$$[f_i, A_i]+[g_j, B_j] := [f_i+g_j, A_i \cap B_j],\ \lambda[f_i, A_i] := [\lambda f_i, A_i] \tag{18}$$

for ${}^\forall[f_i, A_i], [g_j, B_j] \in \mathrm{Pcm}/\sim$ and $\lambda \in \mathbb{R}$.

Now we define the space $\mathrm{SF}(M;\mathfrak{m}) \subset L^\infty(M;\mathfrak{m})$ of simple functions, i.e., each element $h \in \mathrm{SF}(M;\mathfrak{m})$ attains only a finite number values. For $[f_i, A_i] \in \mathrm{Pcm}/\!\sim$ and $h = \sum_{j=1}^{\ell} a_j 1_{B_j} \in \mathrm{SF}(M;\mathfrak{m})$ with a Borel partition (B_j) of M, we define the product $h[f_i, A_i] \in \mathrm{Pcm}/\!\sim$ as

$$h[f_i, A_i] := [a_j f_i, A_i \cap B_j], \tag{19}$$

where we set $B_j := \emptyset$ and $a_j := 0$ for $^\forall j > \ell$. The definition is well-posed and that the resulting multiplication is a bilinear map from $\mathrm{SF}(M;\mathfrak{m}) \times \mathrm{Pcm}/\!\sim$ into $\mathrm{Pcm}/\!\sim$

s.t. for $^{\forall}[f_i, A_i] \in \mathrm{Pcm}/\!\sim$ and every $h, k \in \mathrm{SF}(M; \mathfrak{m})$

$$(hk)[f_i, A_i] = h(k[f_i, A_i]), \qquad 1[f_i, A_i] = [f_i, A_i]. \tag{20}$$

Moreover, the map $\|\cdot\|_{L^2(T^*M)} : \mathrm{Pcm}/\!\sim\, \to [0, +\infty[$ given by

$$\|[f_i, A_i]\|^2_{L^2(T^*M)} := \sum_{i\in\mathbb{N}} \int_{A_i} \Gamma(f_i)\mathrm{d}\mathfrak{m} < \infty$$

constitutes a norm on $\mathrm{Pcm}/\!\sim$.

Def 9.6 (Cotangent module)

We define the Banach space $(L^2(T^*M), \|\cdot\|_{L^2(T^*M)})$ as the completion of $(\mathrm{Pcm}/\sim, \|\cdot\|_{L^2(T^*M)})$. $L^2(T^*M)$ is called cotangent module, and the elements of $L^2(T^*M)$ are called (differential) 1-forms.

Thm 9.1 (Module property) $L^2(T^*M)$ is an L^2-normed L^∞-module over M w.r.t. $\mathfrak{m}$ whose point-wise norm $|\cdot|_{\mathfrak{m}}$ satisfies, for ${}^\forall[f_i, A_i] \in \mathrm{Pcm}/\sim$,

$$|[f_i, A_i]|_{\mathfrak{m}} = \sum_{i\in\mathbb{N}} 1_{A_i}\Gamma(f_i)^{\frac{1}{2}} \quad \mathfrak{m}\text{-a.e.} \tag{21}$$

Def 9.7 (L^2-differential) The L^2-differential $\mathrm{d}f$ of any function $f \in D(\mathcal{E})_e$ is defined by

$$\mathrm{d}f := [f, X] \in L^2(T^*M),$$

where $[f, X] \in \mathrm{Pcm}/\sim\ \subset L^2(T^*M)$ is the representative of the sequence $(f_i, A_i)_{i\in\mathbb{N}}$ given by $f_i := f$, $A_1 := X$, $f_i := 0$ and $A_i := \emptyset$ for ${}^\forall i \geq 2$.

As usual, we call a 1-form $\omega \in L^2(T^*M)$ exact if, for some $f \in D(\mathcal{E})_e$,

$$\omega = \mathrm{d}f.$$

The L^2-differential d is a linear operator on $D(\mathcal{E})_e$. By (21), the L^∞-module structure induced by $\mathfrak{m}$ according to Theorem 9.1,

$$|\mathrm{d}f|_{\mathfrak{m}} = \Gamma(f)^{\frac{1}{2}} \quad \mathfrak{m}\text{-a.e.}$$

holds for ${}^\forall f \in D(\mathcal{E})_e$.

Def 9.8 (Tangent module) The tangent module $(L^2(TM), \|\cdot\|_{L^2(TM)})$ or simply $L^2(TM)$ is

$$L^2(TM) := L^2(T^*M)^*$$

and it is endowed with the norm $\|\cdot\|_{L^2(TM)}$.

The elements of $L^2(TM)$ will be called vector fields.

As before, the point-wise pairing between $\omega \in L^2(T^*M)$ and $X \in L^2(TM)$ is denoted by $\omega(X) \in L^1(M;\mathfrak{m})$, and, by a slight abuse of notation, $|X| \in L^2(M;\mathfrak{m})$ denotes the point-wise norm of M. By Braun ('22+) $L^2(TM)$ is a separable Hilbert module. Furthermore, in terms of the point-wise scalar product $\langle\cdot,\cdot\rangle$ on $L^2(T^*M)$ and $L^2(TM)$, respectively.

We can define the (Riesz) musical isomorphisms $\sharp : L^2(T^*M) \to L^2(TM)$ and $\flat := \sharp^{-1}$ defined by

$$\langle \omega^\sharp, X\rangle := \omega(X) =: \langle X^\flat, \omega\rangle \quad \mathfrak{m}\text{-a.e.} \qquad (22)$$

Def 9.9 (L^2-gradient) The L^2-gradient ∇f of a function $f \in D(\mathcal{E})_e$ is defined by

$$\nabla f := (\mathrm{d}f)^\sharp.$$

Observe from (22) that $f \in D(\mathcal{E})_e$, is characterized as the unique element $X \in L^2(TM)$ which satisfies

$$\mathrm{d}f(X) = |\mathrm{d}f|^2 = |X|^2 \quad \mathfrak{m}\text{-a.e.}$$

Def 9.10 ($\mathrm{Test}(TM)$ and $\mathrm{Reg}(TM)$)

$$\mathrm{Test}(TM) = \left\{ \sum_{i=1}^{n} g_i \nabla f_i \;\middle|\; n \in \mathbb{N}, f_i, g_i \in \mathrm{Test}(M) \right\},$$

$$\mathrm{Reg}(TM) = \left\{ \sum_{i=1}^{n} g_i \nabla f_i \;\middle|\; n \in \mathbb{N}, f_i, g_i \in \mathrm{Test}(M) \cup \mathbb{R}1_M \right\}$$

Def 9.11 ($\mathrm{Test}(T^*M)$ and $\mathrm{Reg}(T^*M)$)

$$\mathrm{Test}(T^*M) = \left\{ \sum_{i=1}^{n} g_i \mathrm{d} f_i \;\middle|\; n \in \mathbb{N}, f_i, g_i \in \mathrm{Test}(M) \right\},$$

$$\mathrm{Reg}(T^*M) = \left\{ \sum_{i=1}^{n} g_i \mathrm{d} f_i \;\middle|\; n \in \mathbb{N}, f_i, g_i \in \mathrm{Test}(M) \cup \mathbb{R}1_M \right\}$$

$\mathrm{Test}(TM) \hookrightarrow L^p(TM)$ (resp. $\mathrm{Test}(T^*M) \hookrightarrow L^p(T^*M))$ for $p \in [1, +\infty[$, hence $\mathrm{Reg}(TM) \cap L^p(TM) \hookrightarrow L^p(TM)$ (resp. $\mathrm{Reg}(T^*M) \cap L^p(T^*M) \hookrightarrow L^p(T^*M)$). From this, $L^2(TM) \cap L^p(TM) \hookrightarrow L^p(TM)$ (resp. $L^2(T^*M) \cap L^p(T^*M) \hookrightarrow L^p(T^*M)$) for $p \in [1, +\infty[$.

We define

$$L^2((T^*)^{\otimes 2}M) := L^2(T^*M) \otimes L^2(T^*M),$$

$$L^2((T)^{\otimes 2}M) := L^2(TM) \otimes L^2(TM).$$

They are point-wise isometrically module isomorphic: the respective pairing is initially defined by

$$(\omega_1 \otimes \omega_2)(X_1 \otimes X_2) := \omega_1(X_1)\omega_2(X_2) \quad \mathfrak{m}\text{-a.e.}$$

for $\omega_1, \omega_2 \in L^2(T^*M) \cap L^\infty(T^*M)$ and $X_1, X_2 \in L^2(TM)$ $L^\infty(TM)$, and is extended by linearity and continuity to $L^2((T^*)^{\otimes 2}M)$ and $L^2((T)^{\otimes 2}M)$, respectively. By a slight abuse of notation, this pairing induces the (Riesz) musical isomorphisms $\flat : L^2((T)^{\otimes 2}M) \to L^2((T^*)^{\otimes 2}M)$

and $\sharp := \flat^{-1}$ given by

$$\langle A^\sharp \,|\, T\rangle_{\mathfrak{m}} := A(T) =: \langle A \,|\, T^\flat\rangle_{\mathfrak{m}} \quad \mathfrak{m}\text{-a.e.} \tag{23}$$

and write $|A|_{\mathrm{HS}} := \sqrt{\langle A \,|\, A\rangle_{\mathfrak{m}}}$ and $|T|_{\mathrm{HS}} := \sqrt{\langle T \,|\, T\rangle_{\mathfrak{m}}}$ for $A \in L^2((T^*)^{\otimes 2}M)$ and $T \in L^2((T)^{\otimes 2}M)$.

Given any $k \in \mathbb{N} \cup \{0\}$, we set

$$L^2(\Lambda T^*M) : = \Lambda L^2(T^*M),$$

$$L^2(\Lambda TM) : = \Lambda L^2(TM).$$

$$L^2(\Lambda^1 T^*M) = L^2(T^*M), \quad L^2(\Lambda^1 TM) = L^2(TM),$$

$$L^2(\Lambda^0 T^*M) = L^2(\Lambda^0 TM) = L^2(M;\mathfrak{m}).$$

These are naturally Hilbert modules. $L^2(\Lambda T^*M)$ and $L^2(\Lambda TM)$ are pointwise isometrically module isomorphic. For brevity, the induced pointwise pairing between $\omega \in L^2(\Lambda T^*M)$ and $X_1 \wedge X_1 \in L^2(\Lambda TM)$ with $X_1 \in L^2(TM) \cap L^\infty(TM)$, is written by

$$\omega(X_1, X_1) := \omega(X_1 \wedge X_1).$$

$$\mathrm{Test}(\Lambda T^*M) := \left\{ \sum_{i=1}^{n} f_i^0 \mathrm{d}f_i^1 \wedge \cdots \wedge \mathrm{d}f_i \;\middle|\; n \in \mathbb{N}, \right.$$
$$\left. f_i^j \in \mathrm{Test}(M) \text{ for } 0 \le j \le k \right\},$$

$$
\begin{aligned}
\mathrm{Test}(\Lambda TM) := \Bigg\{ \sum_{i=1}^{n} f_i^0 \nabla f_i^1 \wedge \cdots \wedge \nabla f_i \;\Bigg|\; & n \in \mathbb{N}, \\
& f_i^j \in \mathrm{Test}(M) \text{ for } 0 \leq j \leq k \Bigg\}, \\
\mathrm{Reg}(\Lambda T^* M) := \Bigg\{ \sum_{i=1}^{n} f_i^0 \mathrm{d} f_i^1 \wedge \cdots \wedge \mathrm{d} f_i \;\Bigg|\; & n \in \mathbb{N}, \\
f_i^j \in \mathrm{Test}(M) \text{ for } 1 \leq j \leq k,\, f_i^0 \in \mathrm{Test}(M) \cup \mathbb{R} 1_M & \Bigg\}, \\
\mathrm{Reg}(\Lambda TM) := \Bigg\{ \sum_{i=1}^{n} f_i^0 \mathrm{d} f_i^1 \wedge \cdots \wedge \mathrm{d} f_i \;\Bigg|\; & n \in \mathbb{N}, \\
f_i^j \in \mathrm{Test}(M) \text{ for } 1 \leq j \leq k,\, f_i^0 \in \mathrm{Test}(M) \cup \mathbb{R} 1_M & \Bigg\}.
\end{aligned}
$$

Def 9.12 ($(1,2)$-Sobolev space $W^{1,2}(TM)$)

The space $W^{1,2}(TM)$ is defined to consist of all $X \in L^2(TM)$ for which ${}^{\exists}T \in L^2(T^{\otimes 2}M)$ s.t. for ${}^{\forall}g_1, g_2, h \in \mathrm{Test}(M)$,

$$\int_M h\langle T \,|\, \nabla g_1 \otimes \nabla g_2\rangle \mathrm{d}\mathfrak{m}$$
$$= -\int_M \langle X, \nabla g_2\rangle \mathrm{div}(h\nabla g_1)\mathrm{d}\mathfrak{m} - \int_M h\,\mathrm{Hess}\,g_2(X, \nabla g_1)\mathrm{d}\mathfrak{m}.$$

Here $\mathrm{Hess}\,g_2 \in L^2((T^*)^{\otimes 2}M)$ is the Hessian defined for $g_2 \in \mathrm{Test}(M)$. The element T is unique, denoted by ∇X and called the covariant derivative of M.

The space $W^{1,2}(TM)$ endowed with the norm $\|\cdot\|_{W^{1,2}(M)}$ is given by

$$\|X\|^2_{W^{1,2}(TM)} := \|X\|^2_{L^2(TM)} + \|\nabla X\|^2_{L^2(T^{\otimes 2}M)}.$$

We also define the covariant functional $\mathcal{E}_{\mathrm{cov}} : L^2(TM) \to [0, +\infty[$ by

$$\mathcal{E}_{\mathrm{cov}}(X) := \begin{cases} \int_M |\nabla X|^2_{\mathrm{HS}} \mathrm{d}\mathfrak{m} & X \in W^{1,2}(TM), \\ \infty & \text{otherwise.} \end{cases} \tag{24}$$

It is proved in **Braun** ('22+) that $(W^{1,2}(TM), \|\cdot\|_{W^{1,2}(TM)}$ is a separable Hilbert space, ∇ is a closed operator, $\mathrm{Reg}(TM) \subset W^{1,2}(TM)$, $W^{1,2}(TM) \hookrightarrow L^2(TM)$, and lsc of $\mathcal{E}_{\mathrm{cov}} : L^2(TM) \to [0, +\infty[$.

Def 9.13 (($(1,2)$-Sobolev space $H^{1,2}(TM)$)) We define the space $H^{1,2}(TM) \subset W^{1,2}(TM)$ as the $\| \cdot \|_{W^{12,}(TM)}$-closure of $\mathrm{Reg}(TM)$:

$$H^{1,2}(TM) := \overline{\mathrm{Reg}(TM)}^{\|\cdot\|_{W^{1,2}(TM)}}.$$

$H^{1,2}(TM)$ is in general a strict subset of $W^{1,2}(TM)$.

Lem 9.1 (Kato's inequality, **Braun** ('22+)) For ${}^{\forall}X \in H^{1,2}$

$|X| \in D(\mathcal{E})$ and

$$|\nabla|X|| \leq |\nabla X|_{\mathrm{HS}} \quad \mathfrak{m}\text{-a.e.}$$

If $X \in H^{1,2}(TM) \cap L^{\infty}(M;\mathfrak{m})$, then $|X|^2 \in D(\mathcal{E})$.

Def 9.14 (Bochner Laplacian) We define $D(\square^B)$ to consist of all $X \in H^{1,2}(TM)$ for which ${}^{\exists}Z \in L^2(TM)$ s.t. for ${}^{\forall}Y \in H^{1,2}(TM)$,

$$\int_M \langle Y, Z \rangle \mathrm{d}\mathfrak{m} = -\int_M \langle \nabla Y \mid \nabla X \rangle \mathrm{d}\mathfrak{m}.$$

In case of existence, Z is uniquely determined, denoted by $\Box^B X$ and called the Bochner Laplacian (or connection Laplacian or horizontal Laplacian) of M.

observe that $D(\Box^B)$ is a vector space, and that $\Box^B : D(\Box^B) \to L^2(TM)$ is a linear operator. Both are easy to see from the linearity of the covariant derivative.

We modify the functional from (24) with the domain $W^{1,2}(TM)$ by introducing the "augmented" covariant energy functional $\widetilde{\mathcal{E}}_{\mathrm{cov}} : L^2(TM) \to [0,+\infty]$ with

$$\widetilde{\mathcal{E}}_{\mathrm{cov}}(X) := \begin{cases} \displaystyle\int_M |\nabla X|^2_{\mathrm{HS}} \, \mathrm{d}\mathfrak{m} & X \in H^{1,2}(TM), \\ \infty & \text{otherwise.} \end{cases}$$

Clearly, its (non-relabeled) polarization $\widetilde{\mathcal{E}}_{\mathrm{cov}} : H^{1,2}(TM)^2$ $\mathbb{R}$ is a closed, symmetric form, and $\square^B$ is the non-positive, self-adjoint generator uniquely associated to it. We write $\mathcal{E}^B$ instead of $\widetilde{\mathcal{E}}_{\mathrm{cov}}$. Let $(P_t^{\mathrm{B}})_{t\geq 0}$ be the heat semigroup on $L^2(TM)$ formally written by

$$\text{“}P_t^{\mathrm{B}} := e^{t\,\square^{\mathrm{B}}}\text{”}.$$

For $\alpha > 0$ & $X \in L^2(TM)$, $R^{\mathrm{B}}_{\alpha} X := \int_0^{\infty} e^{-\alpha t} P^{\mathrm{B}}_t X \,\mathrm{d}t$.

Lem 9.2 (**Braun**('22+)) We have the following:

(1) $0 \leq \inf \sigma(-\Delta) \leq \inf \sigma(-\Box^{\mathrm{B}})$.

(2) For ${}^{\forall}X \in L^2(TM)$ and every $t \geq 0$,

$$|P^{\mathrm{B}}_t X| \leq P_t |X| \quad \mathfrak{m}\text{-a.e.} \tag{25}$$

Cor 9.1 (**Braun**('22+)) Suppose $p \in [1, +\infty[$. Then the heat flow $(P^{\mathrm{B}}_t)_{t \geq 0}$ can be extended to a contractive semigroup on $L^p(TM)$, which is strongly continuous on $L^p(TM)$ under $p \in [1, +\infty[$ and weakly* continuous on $L^{\infty}(TM)$.

Given $\omega \in L^0(\Lambda T^*M)$ and $X_0, \cdots, X_k, Y \in L^0(TM)$

we use the standard abbreviations: for $1 \leq i < j \leq k$

$$\begin{aligned}
\omega(\widehat{X_i}) &:= \omega(X_0, \cdots, \widehat{X_i}, \cdots, X_k), \\
&:= \omega(X_0 \wedge \cdots \wedge X_{i-1} \wedge X_{i+1} \wedge \cdots \wedge X_k), \\
\omega(Y, \widehat{X_i}, \widehat{Y_j}) &:= \omega(Y, X_0, \cdots, \widehat{X_i}, \cdots, \widehat{X_j}, \cdots, X_k), \\
&:= \omega(Y \wedge X_0 \wedge \cdots \wedge X_{i-1} \wedge X_{i+1} \\
&\qquad \wedge \cdots \wedge Y_{j-1} \wedge Y_{j+1} \wedge \cdots \wedge X_k).
\end{aligned}$$

Def 9.15 (Sobolev space $D(\mathrm{d})$)

$D(\mathrm{d})$: The set of all $\omega \in L^2(\Lambda T^*M)$ for which ${}^\exists\eta \in L^2(\Lambda^{k+1}T^*M)$ s.t. for ${}^\forall X_0, \cdots, X_k \in \mathrm{Test}(M)$,

$$\int_M \eta(X_0, \cdots, X_k)\mathrm{d}\mathfrak{m} = \int_M \sum_{i=0}^{1}(-1)^{i+1}\omega(\widehat{X_i})\mathrm{div}\, X_i \mathrm{d}\mathfrak{m}$$
$$+ \int_M \sum_{i=0}^{1}\sum_{j=i+1}^{1}(-1)^{i+j}\omega([X_i, X_j], \widehat{X_i}, \widehat{X_j})\mathrm{d}\mathfrak{m}.$$

In case of existence, the element η is unique, denoted by $\mathrm{d}\omega$ and called the exterior derivative (or exterior differential) of ω.

We always endow $D(\mathrm{d})$ with the norm $\|\cdot\|_{D(\mathrm{d})}$ given by

$$\|\omega\|^2_{D(\mathrm{d})} := \|\omega\|^2_{L^2(\Lambda T^*M)} + \|\mathrm{d}\omega\|^2_{L^2(\Lambda T^*M)}.$$

We introduce the functional $\mathcal{E}_{\mathrm{d}} : L^2(\Lambda T^*M) \to [0, +\infty]$ with

$$\mathcal{E}_{\mathrm{d}}(\omega) := \begin{cases} \displaystyle\int_M |\mathrm{d}\omega|^2 \mathrm{d}\mathfrak{m} & \omega \in D(\mathrm{d}), \\ \infty & \text{otherwise.} \end{cases}$$

We do not make explicit the dependency of $\mathcal{E}_{\mathrm{d}}$ on the degree k. It will always be clear from the context which one is intended.

It is proved in Braun ('22+) that $(D(\mathrm{d}), \| \cdot \|_{D(\mathrm{d})})$ is a separable Hilbert space, the exterior differential d is a closed operator, $\mathrm{Reg}(\Lambda T^*M) \subset D(\mathrm{d})$, $D(\mathrm{d})$ is dense in $L^2(\Lambda T^*M)$, and the functional $\mathcal{E}_{\mathrm{d}} : L^2(\Lambda T^*M) \to [0, +\infty]$ is lower semi continuous.

Def 9.16 (The space $D_{\mathrm{reg}}(\mathrm{d})$) We define the space $D_{\mathrm{reg}}($

$D(\mathrm{d})$ by the closure of $\mathrm{Reg}(\Lambda T^*M)$ w.r.t. the norm $\| \cdot \|_{D(\mathrm{d})}$:

$$D_{\mathrm{reg}}(\mathrm{d}) := \overline{\mathrm{Reg}(\Lambda T^*M)}^{\| \cdot \|_{D(\mathrm{d})}}.$$

It is proved in Braun ('22+) that for $^\forall\omega \in D_{\mathrm{reg}}(\mathrm{d})$, we have $\mathrm{d}\omega \in D_{\mathrm{reg}}(\mathrm{d}^{k+1})$ with $\mathrm{d}(\mathrm{d}\omega) = 0$.

Def 9.17 (The space $D(\mathrm{d}_*)$) $D(\mathrm{d}_*)$: The set of all $\omega \in L^2(\Lambda T^*M)$ for which $^\exists\rho \in L^2(T^*M)$ s.t. for $^\forall\eta \in \mathrm{Test}(T^*M)$, we have

$$\int_M \langle \rho, \eta \rangle \,\mathrm{d}\mathfrak{m} = \int_M \langle \omega, \mathrm{d}\eta \rangle \,\mathrm{d}\mathfrak{m}.$$

If it exists, ρ is unique, denoted by $\mathrm{d}_*\omega$ and called the codifferential of ω. We simply define $D(\mathrm{d}^0_*) := L^0(M;\mathfrak{m})$ and $\mathrm{d}_* := 0$ on this space.

Def 9.18 (The space $W^{1,2}(\Lambda T^*M)$) Define the space W^1

by $W^{1,2}(\Lambda T^*M) := D(\mathrm{d}) \cap D(\mathrm{d}_*)$. By Braun ('22+), we already know that $W^{1,2}(\Lambda T^*M)$ is a dense subspace of $L^2(\Lambda T^*M)$.

We endow $W^{1,2}(\Lambda T^*M)$ with the norm $\|\cdot\|_{W^{1,2}(\Lambda T^*M)}$ given by

$$\|\omega\|^2_{W^{1,2}(\Lambda T^*M)} : = \|\omega\|^2_{L^2(\Lambda T^*M)} + \|\mathrm{d}\omega\|^2_{L^2(\Lambda^{k+1}T^*M)} + \|\mathrm{d}_*\omega\|^2_{L^2(T^*M)}$$

and we define the contravariant functional

$\mathcal{E}_{\mathrm{con}} : L^2(\Lambda T^*M) \to [0, +\infty]$ **by**

$$\mathcal{E}_{\mathrm{con}}(\omega) := \begin{cases} \displaystyle\int_M \big[|\mathrm{d}\omega|^2 + |\mathrm{d}_*\omega|^2\big]\,\mathrm{d}\mathfrak{m} & \omega \in W^{1,2}(\Lambda T^*M) \\ \infty & \text{otherwise.} \end{cases}$$

Arguing as for Braun ('22+), $W^{1,2}(\Lambda T^*M)$ becomes a separable Hilbert space w.r.t. $\|\cdot\|_{W^{1,2}(\Lambda T^*M)}$. Moreover, the functional $\mathcal{E}_{\mathrm{con}} : L^2(\Lambda T^*M) \to [0, +\infty]$ is clearly lower semi continuous.

By Braun ('22+), $\mathrm{Reg}(\Lambda T^*M) \subset W^{1,2}(\Lambda T^*M)$, so that the following definition makes sense.

Def 9.19 (The space $H^{1,2}(\Lambda T^*M)$)

The space $H^{1,2}(\Lambda T^*M) \subset W^{1,2}(\Lambda T^*M)$ is defined by the closure of $\mathrm{Reg}(\Lambda T^*M)$ w.r.t. $\|\cdot\|_{W^{1,2}(\Lambda T^*M)}$:

$$H^{1,2}(\Lambda T^*M) := \overline{\mathrm{Reg}(\Lambda T^*M)}^{\|\cdot\|_{W^{1,2}(\Lambda T^*M)}}.$$

Def 9.20 (L^2-Hodge-Kodaira Laplacian Δ^{HK}) The space $D(\Delta^{\mathrm{HK}})$ is defined to consist of all $\omega \in H^{1,2}(\Lambda T^*M)$ for which ${}^{\exists}\alpha \in L^2(\Lambda T^*M)$ s.t. for ${}^{\forall}\eta \in H^{1,2}(\Lambda T^*M)$,

$$\int_M \langle \alpha, \eta \rangle \mathrm{d}\mathfrak{m} = -\int_M \left[\langle \mathrm{d}\omega, \mathrm{d}\eta \rangle + \langle \mathrm{d}_*\omega, \mathrm{d}_*\eta \rangle \right] \mathrm{d}\mathfrak{m}.$$

In case of existence, the element α is unique, denoted by

$\mathbf{\Delta}^{\mathrm{HK}}\omega$ and called the Hodge Laplacian, Hodge-Kodaira Laplacian or Hodge-de Rham Laplacian of ω. Formally $\mathbf{\Delta}^{\mathrm{HK}}\omega$ can be written "$\mathbf{\Delta}^{\mathrm{HK}}\omega = -(\mathrm{dd}_* + \mathrm{d}_*\mathrm{d})\omega$".

For the most important case $k = 1$, we write $\mathbf{\Delta}^{\mathrm{HK}}$ instead of $\mathbf{\Delta}_1^{\mathrm{HK}}$. We see $\mathbf{\Delta}_0^{\mathrm{HK}} = \Delta$ the usual L^2-generator associated to the given quasi-regular strongly local Dirichlet form $(\mathcal{E}, D(\mathcal{E}))$. Moreover, the Hodge-Kodaira Laplacian $\mathbf{\Delta}^{\mathrm{HK}}$ is a closed operator.

We define the heat flow $P_t^{\rm HK}$ on 1-forms associated to the functional $\widetilde{\mathcal{E}}_{\rm con} : L^2(T^*M) \to [0,+\infty]$ with

$$\widetilde{\mathcal{E}}_{\rm con}(\omega) := \begin{cases} \displaystyle\int_M \left[|{\rm d}\omega|^2 + |{\rm d}_*\omega|^2\right] {\rm d}\mathfrak{m} & \omega \in H^{1,2}(T^*M), \\ \infty & \text{otherwise.} \end{cases}$$

We write $\mathcal{E}^{\rm HK}$ instead of $\widetilde{\mathcal{E}}_{\rm con}$. Let $(P_t^{\rm HK})_{t\geq 0}$ be the heat semigroup of bounded linear and self-adjoint operator on $L^2(T^*M)$ formally written by

$$\text{“}P_t^{\rm HK} := e^{t\boldsymbol{\Delta}^{\rm HK}}\text{”}.$$

The following are important:

Lem 9.3 (**Braun**('22+)) We have the following:

(1) For $^\forall f \in D(\mathcal{E})$ and every $t > 0$, $\mathrm{d}P_t f \in D(\mathbf{\Delta}^{\mathrm{HK}})$ and

$$P_t^{\mathrm{HK}} \mathrm{d}f = \mathrm{d}P_t f. \tag{26}$$

(2) If $\omega \in D(\mathrm{d}_*)$ and $t > 0$, then $P_t^{\mathrm{HK}} \omega \in D(\mathrm{d}_*)$ and

$$\mathrm{d}_* P_t^{\mathrm{HK}} \omega = P_t \mathrm{d}_* \omega. \tag{27}$$

(3) $\inf \sigma(-\Delta^\kappa) \le \inf \sigma(-\mathbf{\Delta}^{\mathrm{HK}})$.

The formulas (26) and (27) are called intertwining properties, which play a crucial role to prove the L^p-boundedness of Riesz operator.

The origin of wald space and phylogenetic information geometry

Tom M. W. Nye, Newcastle University, UK
`tom.nye@ncl.ac.uk`

ENSAE, Paris, October 2023

Joint work with **Maryam Garba**, **Jonas Lueg** and **Stephan Huckemann**

The wald space collaboration

Garba

Nye & Lueg

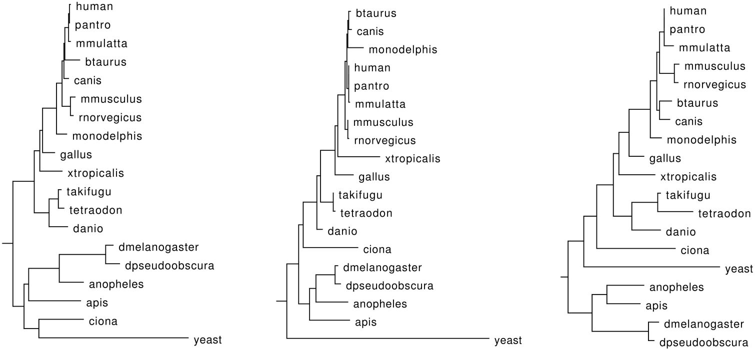
Huckemann

Two papers, two talks

1. Foundations of the wald space for phylogenetic trees,
   Arχiv, 2022
2. Information geometry for phylogenetic trees,
   Journal of Mathematical Biology, 2021

1. Stephan: definition and properties of wald space
2. Me: why wald space? where does the metric come from?

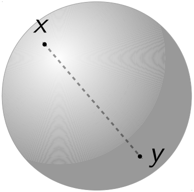
Phylogenetic trees



- Evolutionary trees are constructed from genetic data
- Very typically a collection of trees is obtained: Bayesian posteriors, bootstrap samples, gene trees

Ambient metrics

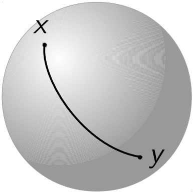
- Suppose (X, d) is a metric space and $Y \subset X$
- Example: consider $S^2 \subset \mathbb{R}^3$



- Standard Euclidean metric on $\mathbb{R}^3$ restricts to give the chordal metric on S^2
- Call metric on $Y \subset X$ an ambient metric

Induced intrinsic metric

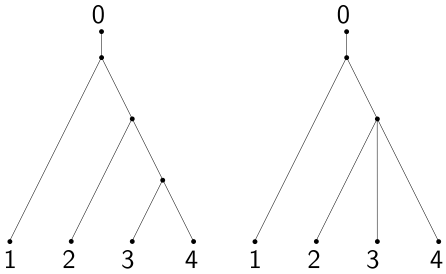
- Measure path length in Y infinitesimally with ambient metric
- Define new metric on Y as infimum of path length between points



- Call this the intrinsic metric induced by d

Notation

- Phylogenetic tree = a connected acyclic graph with no degree 2 vertices
- Leaves are labelled $1, \ldots, N$ and root 0
- Edge weighted – each edge has weight in $\mathbb{R}_{>0}$
- Each tree contains at most $2N$ edges



Metrics between trees

BHV tree space

Trees $\hookrightarrow$ $\mathbb{R}^{2^N-2}$ intrinsic metric induced by Eucl. metric

- Beautiful CAT(0) geometry, but. . .
- Treats trees as geometrical / combinatorial objects

Alternatively

Forests $\hookrightarrow$ distributions on $\{0,1\}^N$

Wald space

Forests $\hookrightarrow$ $\mathcal{S}^+(N)$ intrinsic metric induced by A.I. metric

Metrics between trees

BHV tree space

Trees $\hookrightarrow$ $\mathbb{R}^{2^N-2}$ intrinsic metric induced by Eucl. metric

- Beautiful CAT(0) geometry, but. . .
- Treats trees as geometrical / combinatorial objects

Alternatively

Forests $\hookrightarrow$ distributions on $\{0,1\}^N$

Wald space

Forests $\hookrightarrow$ $\mathcal{S}^+(N)$ intrinsic metric induced by A.I. metric

Metrics between trees

BHV tree space

Trees $\hookrightarrow$ $\mathbb{R}^{2^N-2}$ intrinsic metric induced by Eucl. metric

- Beautiful CAT(0) geometry, but...
- Treats trees as geometrical / combinatorial objects

Alternatively

Forests $\hookrightarrow$ distributions on $\{0,1\}^N$

Wald space

Forests $\hookrightarrow$ $\mathcal{S}^+(N)$ intrinsic metric induced by A.I. metric

Markov substitution models

- Suppose we have a tree T and an alphabet $\Omega = \{A, C, G, T\}$
- Model how each letter in the genome of the species at the root evolves over the tree via a continuous time Markov process with state space Ω
- Tree induces distribution of letters at the leaves
  i.e. distribution on Ω^N
- Use to infer phylogenies from genetic sequence data

Bayes theorem:

$$\Pr(T \mid \text{gene sequence data}) \propto \Pr(\text{gene sequence data} \mid T) \Pr(T)$$

Two state symmetric model

- Take $\Omega = \{0, 1\}$ and let $X(t)$ random variable at $t \in T$
- If $t_1, t_2 \in T$ are path-length ℓ apart

$$\Pr(X(t_2) = X(t_1)) = \frac{1}{2}\left(1 + e^{-\ell}\right)$$
$$\Pr(X(t_2) \neq X(t_1)) = \frac{1}{2}\left(1 - e^{-\ell}\right)$$

- If $X_1, \ldots, X_N$ are the random variables at the leaves then

$$\operatorname{Cov}(X_i, X_j) = \frac{1}{4}\exp(-\ell_{ij})$$

where ℓ_{ij} is the path length between leaves i and j

Embedding trees in a space of distributions

- Given a tree T, the substitution model induces a distribution on letters $X_1, \dots, X_N$ at the leaves
- Let $\mathcal{D}(\Omega^N)$ denote distributions on Ω^N
- Let $p_T(\mathbf{s})$ denote probability mass function associated with tree T, $\mathbf{s} \in \Omega^N$
- For two state symmetric model $T \mapsto \mathcal{D}(\Omega^N)$ is injective

Previous work

- Kim (2001): ‘Slicing hyperdimensional oranges’
- Moulton and Steel (2004): the edge-product space
    - Considered topology of space of tree-like Markov models

Embedding trees in a space of distributions

- Given a tree T, the substitution model induces a distribution on letters $X_1, \ldots, X_N$ at the leaves
- Let $\mathcal{D}(\Omega^N)$ denote distributions on Ω^N
- Let $p_T(\mathbf{s})$ denote probability mass function associated with tree T, $\mathbf{s} \in \Omega^N$
- For two state symmetric model $T \mapsto \mathcal{D}(\Omega^N)$ is injective

Previous work

- Kim (2001): 'Slicing hyperdimensional oranges'
- Moulton and Steel (2004): the edge-product space
  - Considered topology of space of tree-like Markov models

Ambient information metrics

- Any metric d_{am} on $\mathcal{D}(\Omega^N)$ pulls back to give a metric between trees

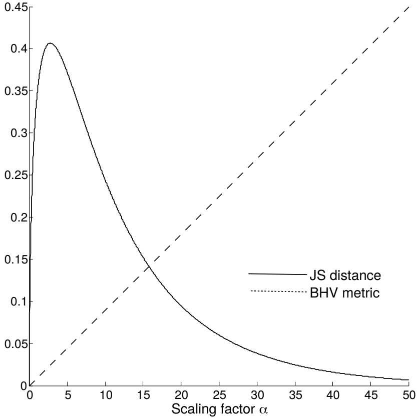
$$d(T_1, T_2) = d_{\text{am}}(p_{T_1}, p_{T_2})$$

- Choice of metric on $\mathcal{D}(\Omega^N)$
  - Jenson-Shannon, Hellinger, (Kullback-Leibler divergence)

E.g. Hellinger distance between $p, q \in \mathcal{D}(\Omega^N)$

$$d_H(p, q)^2 = \frac{1}{2} \sum_{\mathbf{s} \in \Omega^N} \left(\sqrt{p(\mathbf{s})} - \sqrt{q(\mathbf{s})} \right)^2 .$$

Scaling trees



- Pick two random trees and scale all edges by $\alpha > 0$
- As $\alpha \to \infty$, $d(T_1, T_2) \to 0$

Trees and forests

- Letters at leaves separated by infinitely long edges are independent
- Any tree T containing k infinitely long edges can be broken up into a forest $F = T_1 \cup \cdots \cup T_{k+1}$
- Distribution associated to F is

$$p_F = \prod_{i=1}^{k+1} p_{T_i}$$

Remark: By expanding all edges to infinite length, obtain the forest of N isolated vertices

Trees and forests

- Letters at leaves separated by infinitely long edges are independent
- Any tree T containing k infinitely long edges can be broken up into a forest $F = T_1 \cup \cdots \cup T_{k+1}$
- Distribution associated to F is

$$p_F = \prod_{i=1}^{k+1} p_{T_i}$$

Remark: By expanding all edges to infinite length, obtain the forest of N isolated vertices

Fisher information geometry for symmetric 2-state model

- Fix an unweighted binary tree (a tree topology)
- Phylogenies with this topology are parametrized by $\boldsymbol{\ell} \in \mathbb{R}_{>0}^{2N-1}$
- Equip $\mathbb{R}_{>0}^{2N-1}$ with the Fisher information metric

$$g_{ij}(\boldsymbol{\ell}) = \sum_{\mathbf{s}\in\{0,1\}^N} p_{\boldsymbol{\ell}}(\mathbf{s}) \left[\frac{\partial}{\partial_{\ell i}} \log p_{\boldsymbol{\ell}}(\mathbf{s})\right] \left[\frac{\partial}{\partial_{\ell j}} \log p_{\boldsymbol{\ell}}(\mathbf{s})\right]$$

where $p_{\boldsymbol{\ell}}(\mathbf{s})$ is the probability mass function on $\{0,1\}^N$ associated with the tree with edge lengths $\boldsymbol{\ell}$

- Gives $\mathbb{R}_{>0}^{2N-1}$ the structure of a Riemannian manifold
- Solve the geodesic equation for the Riemannian metric numerically

Fisher information geometry for symmetric 2-state model

- Fix an unweighted binary tree (a tree topology)
- Phylogenies with this topology are parametrized by $\boldsymbol{\ell} \in \mathbb{R}_{>0}^{2N-1}$
- Equip $\mathbb{R}_{>0}^{2N-1}$ with the Fisher information metric

$$g_{ij}(\boldsymbol{\ell}) = \sum_{\mathbf{s} \in \{0,1\}^N} p_{\boldsymbol{\ell}}(\mathbf{s}) \left[\frac{\partial}{\partial_{\ell i}} \log p_{\boldsymbol{\ell}}(\mathbf{s}) \right] \left[\frac{\partial}{\partial_{\ell j}} \log p_{\boldsymbol{\ell}}(\mathbf{s}) \right]$$

where $p_{\boldsymbol{\ell}}(\mathbf{s})$ is the probability mass function on $\{0,1\}^N$ associated with the tree with edge lengths $\boldsymbol{\ell}$

- Gives $\mathbb{R}_{>0}^{2N-1}$ the structure of a Riemannian manifold
- Solve the geodesic equation for the Riemannian metric numerically

Why the Fisher information metric?

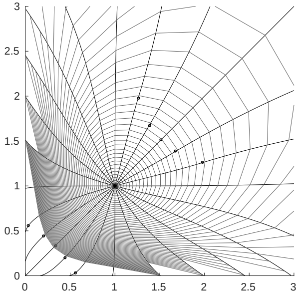
Lemma

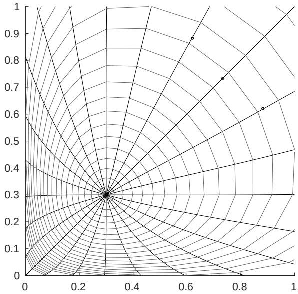
Consider a small perturbation $\delta\boldsymbol{\ell} = (\delta\ell^1, \ldots, \delta\ell^{2N})$ of the edge lengths of a tree $(\tau, \boldsymbol{\ell})$.
Then (using Einstein summation notation)

$$\frac{1}{2}\delta\ell^i g_{ij}(\boldsymbol{\ell})\delta\ell^j \simeq d_{\text{ambient}}(p_{\boldsymbol{\ell}}, p_{\boldsymbol{\ell}+\delta\boldsymbol{\ell}}).$$

i.e. the norm of the perturbation, as measured with respect to the Riemannian inner product, is proportional to the ambient metric

Information geodesics in an orthant for $N = 4$





- Plots substantially different from equivalent for BHV
- Unshown: pendant edge lengths change non-trivially

A continuous Markov model

Problems:

1. Geodesics expensive to compute – sum over Ω^N
2. What about geodesics between trees with different topologies?

Solution: consider Gaussian process $Z(t)$, $t \in T$, which approximates 2-state symmetric process

$$Z(t_2) \mid Z(t_1) = z \sim N(ze^{-\ell_{t_1 t_2}}, 1 - e^{-2\ell_{t_1 t_2}})$$

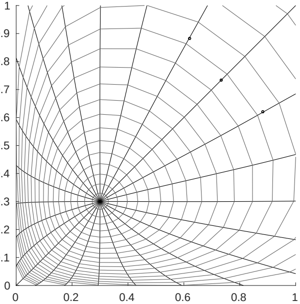
where $\ell_{t_1 t_2}$ is path-length between $t_1, t_2 \in T$

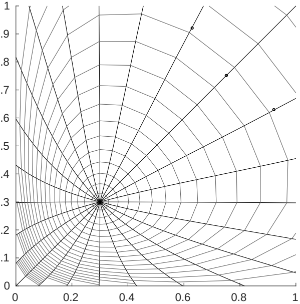
- Induced distribution p_T on $\Omega^N = \mathbb{R}^N$ is $N(0, \Sigma)$ where

  $$\mathrm{Cov}\,(Z_i, Z_j) = \Sigma_{ij} = \exp(-\ell_{ij})$$

  and ℓ_{ij} is path length between leaves i, j on T
- Correlation matrix Σ matches that for 2-state symmetric model on T, and can be shown to be strictly positive definite
- In the Fisher information matrix, summation over Ω^N is replaced by tractable integrals
- This is the well-known affine invariant geometry on symmetric positive definite matrices

Comparison of geodesics for discrete and continuous models





The origin of wald space

1. Aim to construct a geometry for phylogenetic trees by regarding them as probability models for genetic sequences
2. Calculation and properties of ambient information metrics
3. Induced intrinsic metric given by the Fisher information Riemannian metric
4. Replace $\Omega = \{0, 1\}$ with $\Omega = \mathbb{R}$ and use Gaussian process on each tree T
   - Distribution p_T is multivariate normal $N(0, \Sigma)$
   - Sums over Ω^N replaced with tractable integrals
   - This is the affine invariant geometry on symmetric positive definite $N \times N$ matrices Σ

References

- Billera, Holmes and Vogtmann, *Geometry of the space of phylogenetic trees*, Advances in Applied Mathematics, 2001
- Garba, Nye and Boys, *Probabilistic distances between trees*, Systematic Biology, 2017
- Garba, Nye, Lueg and Huckemann, *Information geometry for phylogenetic trees*, Journal of Mathematical Biology, 2021
- Moulton and Steel, *Peeling phylogenetic 'oranges'*, Advances in Applied Mathematics, 2004

Discrete-time gradient flows in Gromov hyperbolic spaces

Shin-ichi OHTA

Osaka Univ./RIKEN AIP

12/Oct/2023 (ENSAE, Paris)

Ref: arXiv:2205.03156 / Israel J. Math. (to appear)

Outline of the talk

As a generalization of the successful theory of convex optimization in $\mathrm{CAT}(0)$-spaces, we consider (geodesic) Gromov hyperbolic spaces.

1. Background
2. Gromov hyperbolic spaces
3. Contraction estimates

§1 Background

Motivating question

Convex optimization (analysis of gradient flows for convex functions) in CAT(0)-spaces made remarkable progress since the mid-1990s.
(Jost, Mayer, Ambrosio–Gigli–Savaré, Bačák, etc.)

Can one generalize to some non-Riemannian spaces?

CAT(0)-spaces

A metric space (X, d) is a CAT(0)-space if

1. geodesic, i.e., $\forall x, y \in X$, $\exists \gamma \colon [0, 1] \longrightarrow X$ (minimal geodesic) s.t.
$$\gamma(0) = x, \quad \gamma(1) = y, \quad d(\gamma(s), \gamma(t)) = |t - s| d(x, y).$$
2. $\forall x, y, z \in X$, $\forall$min. geod. $\gamma \colon [0, 1] \longrightarrow X$ from y to z,
$$d^2(x, \gamma(t)) \le (1-t) d^2(x, y) + t d^2(x, z) - (1-t) t d^2(y, z).$$

The latter condition means that $[d^2(x, \gamma(\cdot))]'' \ge 2d^2(y, z)$, thus $\mathrm{Hess}[d^2] \ge 2$, in the weak sense (along geodesics).

CAT(0) is

- a synthetic notion of nonpositive curvature:
  A complete, simply connected Riemannian manifold is CAT(0) iff its sectional curvature is ≤ 0.
- a "Riemannian" condition:
  Among Banach spaces, only Hilbert spaces are CAT(0). In particular, non-Riemannian Finsler manifolds cannot be CAT(0).

Gradient flows

As usual, we employ the proximal operator constructing discrete-time gradient curves:

Resolvent/Proximal operator

(X,d): a metric space, $f : X \longrightarrow \mathbb{R}$, $\tau > 0$, $x \in X$:

$$\mathsf{J}^f_\tau(x) := \operatorname*{arg\,min}_{y \in X} \left\{ f(y) + \frac{d^2(x,y)}{2\tau} \right\}.$$

$\rightsquigarrow$ The iteration $[\mathsf{J}^f_{t/k}]^k(x)$ converges to a gradient curve $\xi(t)$ for f as $k \to \infty$.

Contraction property

The convexity of d^2 & the Riemannian property are essential in the theory of gradient flows in CAT(0)-sp.'s.

Let (X,d) be CAT(0) and f be K-convex ($K \in \mathbb{R}$), i.e.,

$$f(\gamma(t)) \le (1-t)f(\gamma(0)) + tf(\gamma(1)) - \frac{K}{2}(1-t)td^2(\gamma(0),\gamma(1))$$

∀min. geod. $\gamma\colon [0,1] \longrightarrow X$, $\forall t \in (0,1)$.

Contraction property

For any gradient curves ξ and ζ for f,

$$d(\xi(t),\zeta(t)) \le \mathrm{e}^{-Kt} d(\xi(0),\zeta(0)) \quad \forall t > 0.$$

The contraction property was generalized to:

- CAT(1)-spaces
  (metric spaces of sectional curvature ≤ 1)
- Alexandrov spaces
  (metric spaces of sectional curvature bounded below)
- RCD(K, ∞)-spaces
  (metric measure spaces of Ricci curvature bounded below)

These are all Riemannian!!
(Non-Riemannian Finsler manifolds are excluded.)

Finsler case?

In fact, contraction property fails for Finsler manifolds and normed spaces (O.–Sturm 2012).

Open problem

Any weaker contraction property for convex functions on Finsler manifolds or normed spaces?

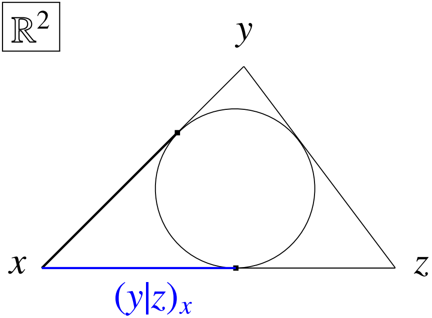
As a class including some non-Riemannian Finsler manifolds, we consider Gromov hyperbolic spaces.

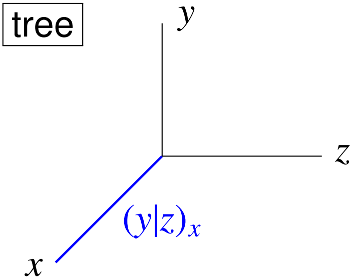
§2 Gromov hyperbolic spaces

Let (X, d) be a metric space.

For $x, y, z \in X$, define the Gromov product:

$$(y|z)_x := \frac{1}{2}\{d(x,y) + d(x,z) - d(y,z)\} \geq 0.$$





δ-hyperbolic spaces

(X, d) is δ-hyperbolic ($\delta \geq 0$) if

$$(x|z)_p \geq \min\{(x|y)_p, (y|z)_p\} - \delta \quad \forall p, x, y, z \in X.$$

(X, d) is Gromov hyperbolic if it is δ-hyperbolic $\exists \delta \geq 0$.

Equality holds with $\delta = 0$ in trees. Thus, trees are 0-hyperbolic.

Roughly speaking, a δ-hyperbolic space is close to a tree up to some “local” perturbations of scale $\leq \delta$.

Other examples

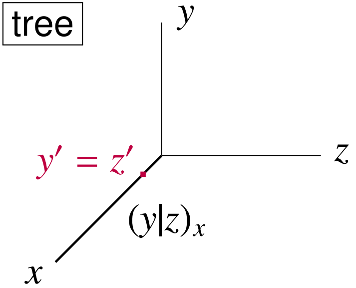
- Complete, simply connected Riem. manifolds of sect. curvature ≤ -1 are Gromov hyperbolic.
- Metric spaces with diameter $\leq \delta$ are δ-hyperbolic.
- Hilbert geometry on a sufficiently smooth & convex domain $D \subset \mathbb{R}^n$ is Gromov hyp. (Karlsson–Noskov 2002); it is non-Riemannian unless D is an ellipsoid.

We shall use the following two fundamental properties of δ-hyperbolic spaces (compare them with triangles in trees).

Lemma A (Tripod lemma)

Let $\gamma, \eta \colon [0,1] \longrightarrow X$ be geodesics emanating from the same point x and put $y = \gamma(1)$, $z = \eta(1)$. Then, for any y' on γ and z' on η with $d(x,y') = d(x,z') \le (y|z)_x$, we have

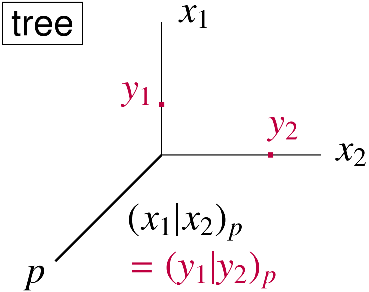
$$d(y',z') \le 4\delta.$$



Lemma B

Let γ_i be a geodesic from p to x_i, $i = 1, 2$. Then, for y_i on γ_i s.t. $\min_{i=1,2} d(p, y_i) \geq (x_1|x_2)_p - \sigma$ with $\sigma \geq 0$, we have

$$|(x_1|x_2)_p - (y_1|y_2)_p| \leq 6\delta + \sigma.$$



§3 Contraction estimates

Setting

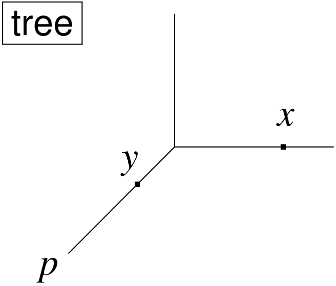
Let (X, d) be a proper (i.e., bounded closed sets are compact), geodesic, δ-hyperbolic space, $f \colon X \longrightarrow \mathbb{R}$ be K-convex ($K \geq 0$). Moreover, assume that f is L-Lipschitz and $\inf_X f$ is attained at some $p \in X$.

(K-convexity along geodesics seems a strong condition $\rightsquigarrow$ related to next talk)

Recall the resolvent/proximal operator:

$$\mathrm{J}_\tau^f(x) := \arg\min_{y \in X} \left\{ f(y) + \frac{d^2(x,y)}{2\tau} \right\}, \quad \tau > 0.$$

Note that $\mathrm{J}_\tau^f(x) \neq \emptyset$ by the properness.



If X is a tree, $\forall y \in \mathrm{J}_\tau^f(x)$, $d(p,y) = d(p,x) - d(x,y)$, i.e., this algorithm (PPA) goes straight to the closest minimizer of f.

Because of inevitable local perturbations of scale $\leq \delta$, the δ-hyperbolicity provides a meaningful information only in a large scale. Thus we consider J_{τ}^{f} with large τ relative to δ (“giant steps”).

Our main results are the following contraction estimates. (We use Lemma A/B to prove Theorem A/B, respectively.)

Theorem A (Tendency towards a minimizer p)

In the setting above, $\forall x \in X$, $\forall y \in \mathsf{J}^f_\tau(x)$, we have

$$d(p,y) \le d(p,x) - d(x,y) + \frac{4\sqrt{2\tau L\delta}}{\sqrt{K\tau+1}}.$$

If $K > 0$ and $\tau > K^{-1}$, we further obtain

$$d(p,y) \le d(p,x) - \left(1 - \frac{1}{K\tau}\right)\frac{f(x)-f(p)}{L} + \frac{4\sqrt{2\tau L\delta}}{\sqrt{K\tau+1}}.$$

(If $f(x) \gg f(p)$, then $d(p,y) \ll d(p,x)$.)

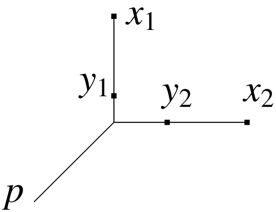
Cf: The case of trees.

Theorem B (Contraction estimate)

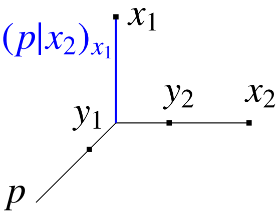
Let $x_1, x_2 \in X$, $y_i \in \mathsf{J}_\tau^f(x_i)$ $(i = 1, 2)$, $d(p, y_1) \le d(p, y_2)$.

1. If $d(p, y_1) \ge (x_1|x_2)_p$, then we have
$$d(y_1, y_2) \le d(x_1, x_2) - d(x_1, y_1) - d(x_2, y_2) + \frac{8\sqrt{2\tau L\delta}}{\sqrt{K\tau + 1}} + 12\delta.$$
2. If $d(p, y_1) < (x_1|x_2)_p$, then we have
$$d(y_1, y_2) \le d(x_1, x_2) - (p|x_2)_{x_1} + C(K, L, D, \tau, \delta),$$
where $D := \max\{d(p, x_1), d(p, x_2)\}$ and $C(K, L, D, \tau, \delta) = O_{K,L,D,\tau}(\delta^{1/4})$ as $\delta \to 0$.

(i) y_1, y_2 do not reach the branching point.



(ii) Essentially reduced to the 1D case (on $p \sim x_2$).



Barycenters and a law of large numbers in Gromov hyperbolic spaces

Shin-ichi OHTA

Osaka Univ./RIKEN AIP

13/Oct/2023 (ENSAE, Paris)

Ref: arXiv:2211.00193

Outline of the talk

- As a generalization of the successful theory of convex optimization in $\mathrm{CAT}(0)$-spaces, we consider (geodesic) Gromov hyperbolic spaces.
- In this talk, we study barycenters of probability measures.

1. Background
2. Barycenters
3. A law of large numbers

§1 Background

The class of geodesically convex functions seems restrictive, compared with the local flexibility of Gromov hyperbolic spaces.

$\rightsquigarrow$ We'd like to build the theory of "roughly convex" functions on δ-hyperbolic spaces, including the (squared) distance function.

$\rightsquigarrow$ For this purpose, we first consider the case of distance function, thus barycenters.

§2 Barycenters

Let (X, d) be a δ-hyperbolic space. Given a Borel probability measure $\mu \in \mathcal{P}^2(X)$ on X of finite second moment, define the variance of μ by

$$\mathbf{V}(\mu) := \inf_{x \in X} \int_X d^2(x, z)\, \mu(dz) = \inf_{x \in X} W_2^2(\delta_x, \mu).$$

If $x \in X$ attains the $\inf$, we call it a barycenter of μ.
($W_p = L^p$-Wasserstein distance on $\mathcal{P}^p(X)$.)

Since μ may not have any barycenter, we consider

$$\mathcal{B}(\mu, \varepsilon) := \left\{ x \in X \,\middle|\, W_2^2(\delta_x, \mu) \le \mathbf{V}(\mu) + \varepsilon \right\}, \quad \varepsilon \ge 0.$$

Remark (Extension to $\mathcal{P}^1(X)$)

One can in fact discuss barycenters of $\mu \in \mathcal{P}^1(X)$ of finite first moment, by considering

$$\inf_{x \in X} \int_X \{d^2(x,z) - d^2(x_0,z)\}\,\mu(dz)$$

for arbitrarily fixed $x_0 \in X$ (indep. of the choice of x_0).

Fact (CAT(0)-case)

In a complete CAT(0)-space, any $\mu \in \mathcal{P}^1(X)$ admits a unique barycenter, denoted by $\beta_\mu \in X$.

In fact, $\forall x, y \in X$, the midpoint w of x and y satisfies (by integrating the CAT(0)-inequality in μ)

$$W_2^2(\delta_w, \mu) \leq \frac{1}{2}W_2^2(\delta_x, \mu) + \frac{1}{2}W_2^2(\delta_y, \mu) - \frac{1}{4}d^2(x, y).$$

$\rightsquigarrow$ Any minimizing sequence of $W_2^2(\delta_\cdot, \mu)$ is a Cauchy sequence and converges to the unique barycenter.

Extending the $\mathrm{CAT}(0)$-inequality with an additional term depending on δ leads the following.

Proposition (Size of $\mathcal{B}(\mu, \varepsilon)$)

Let (X, d) be a geodesic δ-hyperbolic space. For any $\mu \in \mathcal{P}^1(X)$ and $x, y \in \mathcal{B}(\mu, \varepsilon)$, we have

$$d(x, y) \leq 2\sqrt{2\delta\{W_1(\delta_x, \mu) + W_1(\delta_y, \mu)\} + 4\delta^2 + \varepsilon}.$$

In particular, for $\varepsilon = 0$, we have

$$d(x, y) \leq O\left(\sqrt{\delta}\right).$$

Wasserstein contraction property

Fact (CAT(0)-case)

In a complete CAT(0)-space, $\forall \mu, \nu \in \mathcal{P}^1(X)$, we have

$$d(\beta_\mu, \beta_\nu) \le W_1(\mu, \nu).$$

In other words, the map

$$\beta \colon (\mathcal{P}^1(X), W_1) \longrightarrow X, \quad \beta(\mu) := \beta_\mu,$$

is non-expanding (giving a "projection" from $\mathcal{P}^1(X)$ to X; clearly $\beta(\delta_x) = x$).

Theorem (Wasserstein contraction)

Let (X, d) be a geodesic δ-hyperbolic space. For any $\mu, \nu \in \mathcal{P}^1(X)$, $x \in \mathcal{B}(\mu, \varepsilon_1)$ and $y \in \mathcal{B}(\nu, \varepsilon_2)$, we have

$$d(x, y) \le W_1(\mu, \nu) + 8\delta \vee \sqrt{54D\sqrt{D+\delta}\sqrt{\delta}} + 3(\varepsilon_1 + \varepsilon_2),$$

where $D := \mathrm{diam}(\mathrm{supp}\,\mu \cup \mathrm{supp}\,\nu \cup \{x, y\})$ and $a \vee b := \max\{a, b\}$.

In particular, for $\varepsilon_1 = \varepsilon_2 = 0$, we have

$$d(x, y) \le W_1(\mu, \nu) + O\Big(\delta^{1/4}\Big).$$

§3 A law of large numbers

How to approximate barycenters? In a complete CAT(0)-space, Sturm established the following.

Sturm's law of large numbers (2003)

Take $\mu \in \mathcal{P}(X)$ with bounded support, and let $(Z_i)_{i\geq 1}$ be a sequence of i.i.d. random variables with distribution μ. We recursively choose

$$S_1 := Z_1, \qquad S_k := \gamma(k^{-1}) \ (k \geq 2),$$

where $\gamma\colon [0,1] \longrightarrow X$ is the min. geod. from S_{k-1} to Z_k. Then S_k converges to β_μ almost surely.

Note that the above recursive choice S_k requires no knowledge of the construction of barycenters.

Some generalizations

- CAT(1)-spaces of diameter $\leq \pi/2$ (O–Pálfia 2015),
- CAT(1)-spaces of radii $\leq \pi/2$ (Yokota 2018),
- finite dimensional Alexandrov spaces of curvature bounded below (O–Pálfia 2015).

In a geodesic δ-hyperbolic space (X, d), due to the local flexibility, we fix a rate instead of k^{-1} going to 0.

Theorem (A law of large numbers): Setting

Take $\mu \in \mathcal{P}(X)$ having a barycenter $p \in X$. Let $(Z_i)_{i\geq 1}$ be a sequence of i.i.d. random variables with distribution μ. Given $\tau > 0$, take recursively

$$S_1 := Z_1, \qquad S_k := \gamma(2\tau/(2\tau+1)) \ (k \geq 1),$$

where $\gamma\colon [0,1] \longrightarrow X$ is a min. geod. from S_{k-1} to Z_k. Assume that $\operatorname{supp}\mu$, p and $(S_k)_{k\geq 1}$ are all included in a bounded set $\Omega \subset X$.

Theorem (A law of large numbers): Assertion

Then, $\forall \varepsilon > 0$, we have

$$\mathbb{E}[d^2(p, S_{k_0})] \leq 8D_\Omega^2\tau + C(D_\Omega, \tau, \delta)\delta + \varepsilon$$

for some $k_0 < D_\Omega^2/(\tau\varepsilon)$, where $D_\Omega := \mathrm{diam}(\Omega)$.

Hence, after enough iteration (sublinear in ε), S_k likely passes close to p, which makes it possible to restrict the region we explore barycenters.

When we assume $\delta \leq D_\Omega/2$ and choose $\tau = \sqrt{\delta/D_\Omega}$, we have

$$\mathbb{E}[d^2(p, S_{k_0})] \leq \varepsilon + O\left(\sqrt{\delta}\right).$$

Deterministic approximation

A “deterministic” counterpart to the “stochastic” LLN:

Theorem (Deterministic approximation)

Let $(z_i)_{i=1}^n \subset X$ and $p \in X$ be a minimizer of $\sum_{i=1}^n d^2(z_i, \cdot)$.
For $\tau > 0$ and an arbitrary initial point $y_0 \in X$, we take

$$y_{kn+i} := \gamma(2\tau/(2\tau + 1)),$$

where $\gamma\colon [0,1] \longrightarrow X$ is a min. geod. from y_{kn+i-1} to z_i.
Assume that $\{p, z_i, y_{kn+i}\}$ is included in a bounded set Ω.
Then, $\forall \varepsilon > 0$, $\exists k_0 < d^2(p, y_0)/(2\tau\varepsilon)$ such that

$$d^2(p, y_{k_0 n}) \le C(D_\Omega, \tau, \delta, n)(\delta + \tau) + \frac{2\varepsilon}{n}.$$

Further problems

- Improvements by comparing the case of trees (instead of CAT(0)-spaces as above)?
- Introduce an appropriate class of "roughly convex" functions on (possibly non-geodesic) δ-hyperbolic spaces, including (squared) distance functions.
- Study optimization (discrete-time gradient flows) for functions in the above class, possibly with random noise (a kind of simulated annealing).
  $\rightsquigarrow$ Any applications to optimization theory?
- Any connections with geometric group theory?

Gradient flows and calculus of variations in CAT(1)-spaces

Miklós Pálfia

Corvinus University of Budapest and University of Szeged

October 13, 2023

miklos.palfia@uni-corvinus.hu

Based on joint works with Shin-ichi Ohta and Hedvig Gál

Introduction

Let (X, d) be a complete metric space. Consider a lower semi-continuous (lsc) function $\phi : X \longrightarrow (-\infty, \infty]$ such that

$$D(\phi) := X \setminus \phi^{-1}(\infty) \neq \emptyset.$$

If X is Riemannian, a gradient curve $\xi : [0, \infty) \longrightarrow X$ of ϕ with initial condition $\xi(0) := x_0 \in D(\phi)$ is a solution of

$$\dot{\xi} = -\nabla\phi(\xi).$$

We are interested in constructing gradient curves or finding minimizers of ϕ. Classically, the first is related to the Crandall-Liggett theory of contraction semigroups in Banach spaces generated by monotone nonlinear operators. Secondly, discrete approximations of gradient curves leads us to optimization techniques, such as proximal point methods. All can be treated in a unified manner as instances of (contractive) evolution systems in Banach spaces.

Given $x \in X$ and $\tau > 0$, the *Moreau–Yosida approximation* is $\phi_\tau(x) := \inf_{z \in X} \left\{ \phi(z) + \frac{d^2(x,z)}{2\tau} \right\}$ and set

$$J_\tau^\phi(x) := \left\{ z \in X \,\middle|\, \phi(z) + \frac{d^2(x,z)}{2\tau} = \phi_\tau(x) \right\}.$$

For $x \in D(\phi)$ and $z \in J_\tau^\phi(x)$ we have $d^2(x,z) \leq 2\tau\{\phi(x) - \phi(z)\}$.

Assumption

(1) (*coercivity*) There exists $\tau_*(\phi) \in (0,\infty]$ such that $\phi_\tau(x) > -\infty$ and $J_\tau^\phi(x) \neq \emptyset$ for all $x \in X$ and $\tau \in (0, \tau_*(\phi))$.

(2) (*compactness*) For any $Q \in \mathbb{R}$, bounded subsets of the sub-level set $\{x \in X \,|\, \phi(x) \leq Q\}$ are relatively compact in X.

Remark

If $\operatorname{diam} X < \infty$ *and* (2) *holds, then the lsc of* ϕ *implies that every sub-level set* $\{x \in X \,|\, \phi(x) \leq Q\}$ *is (empty or) compact. Thus* ϕ *is bounded below and we can take* $\tau_*(\phi) = \infty$*.*

To construct discrete approximations of gradient curves of ϕ, we consider a *partition* of the interval $[0, \infty)$:

$$\mathscr{P}_{\boldsymbol{\tau}} = \{0 = t_{\boldsymbol{\tau}}^0 < t_{\boldsymbol{\tau}}^1 < \cdots\}, \qquad \lim_{k\to\infty} t_{\boldsymbol{\tau}}^k = \infty,$$

and set

$$\tau_k := t_{\boldsymbol{\tau}}^k - t_{\boldsymbol{\tau}}^{k-1} \quad \text{for } k \in \mathbb{N}, \qquad |\boldsymbol{\tau}| := \sup_{k\in\mathbb{N}} \tau_k.$$

We will always assume $|\boldsymbol{\tau}| < \tau_*(\phi)$. Given an initial point $x_0 \in D(\phi)$,

$x_{\boldsymbol{\tau}}^0 := x_0$ and recursively choose arbitrary $x_{\boldsymbol{\tau}}^k \in J_{\tau_k}^{\phi}(x_{\boldsymbol{\tau}}^{k-1})$ for each $k \in \mathbb{N}$.

We call $\{x_{\boldsymbol{\tau}}^k\}_{k\in\mathbb{N}}$ a *discrete solution* of the variational scheme associated with the partition $\mathscr{P}_{\boldsymbol{\tau}}$, which is thought of as a *discrete-time gradient curve* for the potential function ϕ.

Convergence of discrete solutions

Let $\phi : (-\infty, \infty] \longrightarrow X$ be λ-*convex* for some $\lambda \in \mathbb{R}$ in the sense that

$$\phi\bigl(\gamma(t)\bigr) \leq (1-t)\phi(x) + t\phi(y) - \frac{\lambda}{2}(1-t)td^2(x,y)$$

for any $x, y \in D(\phi)$ and some minimal geodesic $\gamma : [0,1] \longrightarrow X$ from x to y.

We remark that the compactness (2) in the Assumption implies the coercivity in this case; we even have $\tau_*(\phi) = \infty$ if $\lambda \geq 0$).

Fix an initial point $x_0 \in D(\phi)$. Take a sequence of partitions $\{\mathscr{P}_{\boldsymbol{\tau}_i}\}_{i\in\mathbb{N}}$ such that $\lim_{i\to\infty} |\boldsymbol{\tau}_i| = 0$ and associated discrete solutions $\{x^k_{\boldsymbol{\tau}_i}\}_{k\in\mathbb{N}}$ with $x^0_{\boldsymbol{\tau}_i} = x_0$. Under Assumption (2), by the compactness argument, a subsequence of the interpolated curves

$$\bar{\boldsymbol{x}}_{\boldsymbol{\tau}_i}(0) := x_0, \qquad \bar{\boldsymbol{x}}_{\boldsymbol{\tau}_i}(t) := x^k_{\boldsymbol{\tau}_i} \quad \text{for } t \in (t^{k-1}_{\boldsymbol{\tau}_i}, t^k_{\boldsymbol{\tau}_i}]$$

converges to a curve $\xi : [0,\infty) \longrightarrow D(\phi)$ point-wise in $t \in [0,\infty)$.

In general, under the coercivity and λ-convexity of ϕ (but without the compactness), if a curve ξ is obtained as above (called a *generalized minimizing movement*), then it is locally Lipschitz on $(0, \infty)$ and satisfies $\lim_{t\downarrow 0} \xi(t) = x_0$ as well as the *energy dissipation identity*:

$$\phi\big(\xi(T)\big) = \phi\big(\xi(S)\big) - \frac{1}{2}\int_S^T \{|\dot{\xi}|^2 + |\nabla\phi|^2(\xi)\}\, dt.$$

Here

$$|\dot{\xi}|(t) := \lim_{s\to t} \frac{d(\xi(s), \xi(t))}{|t-s|}$$

is the *metric speed* existing at almost all t, and

$$|\nabla\phi|(x) := \limsup_{y\to x} \frac{\max\{\phi(x)-\phi(y), 0\}}{d(x,y)}$$

is the (descending) *local slope*. We remark that $|\nabla\phi|$ is lower semi-continuous and $\lim_{i\to\infty} \phi(\bar{\boldsymbol{x}}_{\boldsymbol{\tau}_i}(t)) = \phi(\xi(t))$ for all $t \geq 0$.

CAT(1)-spaces

Given three points $x, y, z \in X$ with $d(x,y) + d(y,z) + d(z,x) < 2\pi$, we can take corresponding points $\tilde{x}, \tilde{y}, \tilde{z}$ in the 2-dimensional unit sphere $\mathbb{S}^2$ such that

$$d_{\mathbb{S}^2}(\tilde{x},\tilde{y}) = d(x,y), \qquad d_{\mathbb{S}^2}(\tilde{y},\tilde{z}) = d(y,z), \qquad d_{\mathbb{S}^2}(\tilde{z},\tilde{x}) = d(z,x).$$

We call $\triangle\tilde{x}\tilde{y}\tilde{z}$ a *comparison triangle* of $\triangle xyz$ in $\mathbb{S}^2$.

Definition (CAT(1)-spaces)

A geodesic metric space (X, d) is called a $\mathrm{CAT}(1)$-*space* if, for any $x, y, z \in X$ with $d(x,y) + d(y,z) + d(z,x) < 2\pi$ and any minimal geodesic $\gamma : [0,1] \longrightarrow X$ from y to z, we have

$$d\big(x, \gamma(t)\big) \le d_{\mathbb{S}^2}\big(\tilde{x}, \tilde{\gamma}(t)\big)$$

at all $t \in [0,1]$, where $\triangle\tilde{x}\tilde{y}\tilde{z} \subset \mathbb{S}^2$ is a comparison triangle of $\triangle xyz$ and $\tilde{\gamma} : [0,1] \longrightarrow \mathbb{S}^2$ is the minimal geodesic from $\tilde{y}$ to $\tilde{z}$.

Lemma (Semi-convexity of distance functions)

Let (X, d) *be a* $\mathrm{CAT}(1)$*-space and take* $R \in (0, \pi)$*. Then there exists* $K = K(R) \in \mathbb{R}$ *such that the squared distance function* $d^2(x, \cdot)$ *is* K*-convex on the open* R*-ball* $B(x, R)$ *for all* $x \in X$*.*

We define the *angle* between two geodesics γ and η emanating from $\gamma(0) = \eta(0) = x$ by $\angle_x(\gamma, \eta) := \lim_{s,t\downarrow 0} \angle \widetilde{\gamma(s)}\tilde{x}\widetilde{\eta(t)}$, where $\angle \widetilde{\gamma(s)}\tilde{x}\widetilde{\eta(t)}$ is the angle at $\tilde{x}$ of $\triangle \widetilde{\gamma(s)}\tilde{x}\widetilde{\eta(t)}$ in $\mathbb{S}^2$.

Theorem (First variation formula)

Let $\gamma : [0, 1] \longrightarrow X$ *be a geodesic from* x *to* z*, and take* $y \in X$ *with* $0 < d(x, y) < \pi$*. Then we have*

$$\lim_{s\downarrow 0} \frac{d(\gamma(s), y) - d(x, y)}{s} = -d(x, z) \cos \angle_x(\gamma, \eta),$$

where $\eta : [0, 1] \to X$ *is the unique minimal geodesic from* x *to* y*.*

Key lemma

Let (X, d) be a complete $\mathrm{CAT}(1)$-space and $\phi : X \longrightarrow (-\infty, \infty]$ satisfy the λ-convexity for some $\lambda \in \mathbb{R}$ and Assumption (1).

Lemma (Key lemma)

Let $x \in D(\phi)$ and $\tau \in (0, \min\{\pi^2/(2C), \tau_*(\phi)/8\})$ with $C = C(x, \tau_*(\phi), \phi(x), \tau_*(\phi)/8)$. Take $x_\tau \in J_\tau^\phi(x)$. Then we have, for any $y \in D(\phi) \cap B(x_\tau, R - d(x, x_\tau))$ with $R < \pi$ and for $K = K(R)$,

$$\begin{aligned} d^2(x_\tau, y) &\leq d^2(x, y) - \lambda\tau d^2(x_\tau, y) + 2\tau\{\phi(y) - \phi(x_\tau)\} - \frac{K}{2} d^2(x, x_\tau) \\ &\leq d^2(x, y) - \lambda\tau d^2(x_\tau, y) + 2\tau\{\phi(y) - \phi(x_\tau)\} \\ &\quad + \max\{0, -K\} \cdot \tau\{\phi(x) - \phi(x_\tau)\}. \end{aligned}$$

proof of the Key lemma

We have $d^2(x, x_\tau) \leq 2C\tau < \pi^2$ by an a priori lemma of Ambrosio-Gigli-Savaré and the choice of τ. Let $\gamma : [0,1] \longrightarrow X$ be the minimal geodesic from x_τ to y, and $\eta : [0,1] \longrightarrow X$ from x_τ to x. For any $s \in (0,1)$, by the definition of $J_\tau^\phi(x)$ and the λ-convexity of ϕ, we have

$$\begin{aligned}\phi(x_\tau) + \frac{d^2(x, x_\tau)}{2\tau} \leq& \phi(\gamma(s)) + \frac{d^2(x, \gamma(s))}{2\tau} \\ \leq& (1-s)\phi(x_\tau) + s\phi(y) - \frac{\lambda}{2}(1-s)sd^2(x_\tau, y) \\ &+ \frac{d^2(x, \gamma(s))}{2\tau}.\end{aligned}$$

Hence

$$\phi(x_\tau) \le \phi(y) + \frac{1}{2\tau}\frac{d^2(x,\gamma(s)) - d^2(x,x_\tau)}{s} - \frac{\lambda}{2}(1-s)d^2(x_\tau,y).$$

Applying the first variation formula twice, we observe the *commutativity*:

$$\lim_{s\downarrow 0}\frac{d^2(x,\gamma(s)) - d^2(x,x_\tau)}{s} = \lim_{t\downarrow 0}\frac{d^2(\eta(t),y) - d^2(x_\tau,y)}{t},$$

since both sides equal $-2d(x_\tau,x)d(x_\tau,y)\cos\angle_{x_\tau}(\gamma,\eta)$.
Notice that η is contained in $B(y,R)$ by the choice of y. Thus it follows from the K-convexity of $d^2(\cdot,y)$ in $B(y,R)$ that

$$\begin{aligned}\lim_{t\downarrow 0}\frac{d^2(\eta(t),y) - d^2(x_\tau,y)}{t} &\le d^2(x,y) - d^2(x_\tau,y) - \frac{K}{2}d^2(x,x_\tau)\\ &\le d^2(x,y) - d^2(x_\tau,y) + \max\{0,-K\}\cdot\tau\{\phi(x) - \phi(x_\tau)\}.\end{aligned}$$

□

Remark

(a) Used before by Mayer, Ambrosio-Gigli-Savaré and Bačák is the direct application of the convexity of ϕ and $d^2(x, \cdot)$ along γ, which implies in our setting

$$\frac{K}{2} d^2(x_\tau, y) \leq d^2(x, y) - \lambda\tau d^2(x_\tau, y) + 2\tau\{\phi(y) - \phi(x_\tau)\} - d^2(x, x_\tau).$$

This coincides with our estimate when $K = 2$. The commutativity was used to move the coefficient $K/2$ from $d^2(x_\tau, y)$ to $d^2(x, x_\tau)$.
(b) The Riemannian nature of the space (i.e., the angle) is essential in the commutativity. In fact, on a Finsler manifold (M, F), commutativity (written using only the distance) implies

$$g_v(v, w) = g_w(v, w) \qquad \textit{for all } v, w \in T_x M \setminus \{0\},\ x \in M,$$

and the parallelogram identity on $T_x M$ and hence F is Riemannian.

Applications to gradient flows

Our argument covers two cases. In both cases, (X, d) is complete, $\phi : X \longrightarrow (-\infty, \infty]$ is lower semi-continuous, λ-convex and $D(\phi) \neq \emptyset$.

Case (I)

(X, d) is a $\mathrm{CAT}(1)$-space.

Case (II)

(X, d) satisfies the commutativity and the K-convexity of the squared distance function, and ϕ satisfies the coercivity condition (Assumption (1)).

We stress that both $\lambda, K \in \mathbb{R}$ can be negative.

Interpolations

Given an initial point $x_0 \in D(\phi)$ and a partition $\mathscr{P}_{\boldsymbol{\tau}}$ with $|\boldsymbol{\tau}| < \tau_*(\phi)$, we fix a discrete solution $\{x_{\boldsymbol{\tau}}^k\}_{k\in\mathbb{N}}$. Let us also take a point $y \in X$. We interpolate the discrete data $x_{\boldsymbol{\tau}}^k$, $d(x_{\boldsymbol{\tau}}^k, y)$ and $\phi(x_{\boldsymbol{\tau}}^k)$ as follows:
For $t \in (t_{\boldsymbol{\tau}}^{k-1}, t_{\boldsymbol{\tau}}^k]$, $k \in \mathbb{N}$, define

$$\bar{\boldsymbol{x}}_{\boldsymbol{\tau}}(t) := x_{\boldsymbol{\tau}}^k \in J_{\tau_k}^{\phi}(x_{\boldsymbol{\tau}}^{k-1}) \quad (\bar{\boldsymbol{x}}_{\boldsymbol{\tau}}(0) := x_0),$$

$$\bar{\boldsymbol{d}}_{\boldsymbol{\tau}}(t; y) := \left\{ d^2(x_{\boldsymbol{\tau}}^{k-1}, y) + \frac{t - t_{\boldsymbol{\tau}}^{k-1}}{\tau_k} \{ d^2(x_{\boldsymbol{\tau}}^k, y) - d^2(x_{\boldsymbol{\tau}}^{k-1}, y) \} \right\}^{1/2},$$

$$\bar{\boldsymbol{\phi}}_{\boldsymbol{\tau}}(t) := \phi(x_{\boldsymbol{\tau}}^{k-1}) + \frac{t - t_{\boldsymbol{\tau}}^{k-1}}{\tau_k} \{ \phi(x_{\boldsymbol{\tau}}^k) - \phi(x_{\boldsymbol{\tau}}^{k-1}) \}.$$

Recall that $\tau_k = t_{\boldsymbol{\tau}}^k - t_{\boldsymbol{\tau}}^{k-1}$ and note that $\bar{\boldsymbol{\phi}}_{\boldsymbol{\tau}}$ is non-increasing.

Theorem (Discrete evolution variational inequality)

Assuming $|\boldsymbol{\tau}| < \tau_*(\phi)$, *we have*

$$\frac{1}{2}\frac{d}{dt}\left[\bar{\boldsymbol{d}}_{\boldsymbol{\tau}}^{2}(t;y)\right] + \frac{\lambda}{2}d^2\big(\bar{\boldsymbol{x}}_{\boldsymbol{\tau}}(t),y\big) + \bar{\boldsymbol{\phi}}_{\boldsymbol{\tau}}(t) - \phi(y) \leq \mathscr{R}_{\boldsymbol{\tau},K}(t)$$

for almost all $t \in (0,T)$ *and all* $y \in D(\phi)$, *where for* $t \in (t_{\boldsymbol{\tau}}^{k-1}, t_{\boldsymbol{\tau}}^{k}]$

$$\mathscr{R}_{\boldsymbol{\tau},K}(t) := \left(\frac{t_{\boldsymbol{\tau}}^{k}-t}{\tau_k} + \frac{\max\{0,-K\}}{2}\right)\{\phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^{k})\}.$$

Convergence of discrete solutions

Theorem (Unique limits of discrete solutions)

Fix an initial point $x_0 \in D(\phi)$ *and consider discrete solutions* $\{x_{\boldsymbol{\tau}_i}^k\}_{k\in\mathbb{N}}$ *with* $x_{\boldsymbol{\tau}_i}^k = x_0$ *associated with a sequence of partitions* $\{\mathscr{P}_{\boldsymbol{\tau}_i}\}_{i\in\mathbb{N}}$ *such that* $\lim_{i\to\infty} |\boldsymbol{\tau}_i| = 0$. *Then the interpolated curve* $\bar{\boldsymbol{x}}_{\boldsymbol{\tau}_i} : [0,\infty) \longrightarrow X$ *converges to a curve* $\xi : [0,\infty) \longrightarrow X$ *with* $\xi(0) = x_0$ *as* $i \to \infty$ *uniformly on each bounded interval* $[0,T]$. *In particular, the limit curve* ξ *is independent of the choice of the sequence of partitions nor discrete solutions.*

We can define the *gradient flow* operator

$$\mathcal{G} : [0,\infty) \times D(\phi) \longrightarrow D(\phi) \tag{4.1}$$

by $\mathcal{G}(t,x_0) := \xi(t)$, where $\xi : [0,\infty) \longrightarrow X$ is the unique gradient curve with $\xi(0) = x_0$. Then the semigroup property holds:

$$\mathcal{G}\big(t,\mathcal{G}(s,x_0)\big) = \mathcal{G}(s+t,x_0) \quad \text{for all } s,t \geq 0.$$

Contraction property

Theorem (Contraction property)

Take $x_0, y_0 \in D(\phi)$ and put $\xi(t) := \mathcal{G}(t, x_0)$ and $\zeta(t) := \mathcal{G}(t, y_0)$. Then we have, for any $t > 0$,

$$d\big(\xi(t), \zeta(t)\big) \leq e^{-\lambda t} d(x_0, y_0).$$

The contraction property allows us to take the continuous limit

$$\mathcal{G} : [0, \infty) \times \overline{D(\phi)} \longrightarrow \overline{D(\phi)}$$

of the gradient flow operator, which again enjoys the semigroup property as well as the contraction property.

Evolution variational inequality

Theorem (Evolution variational inequality)

Take $x_0 \in D(\phi)$ *and put* $\xi(t) := \mathcal{G}(t, x_0)$. *Then we have*

$$\limsup_{\varepsilon \downarrow 0} \frac{d^2(\xi(t+\varepsilon), y) - d^2(\xi(t), y)}{2\varepsilon} + \frac{\lambda}{2} d^2\big(\xi(t), y\big) + \phi\big(\xi(t)\big) \le \phi(y)$$

for all $y \in D(\phi)$ *and* $t > 0$. *In particular,*

$$\frac{1}{2}\frac{d}{dt}\left[d^2\big(\xi(t), y\big)\right] + \frac{\lambda}{2} d^2\big(\xi(t), y\big) + \phi\big(\xi(t)\big) \le \phi(y)$$

for all $y \in D(\phi)$ *and almost all* $t > 0$.

Stationary points and large time behavior of the flow

Theorem

A point $x_0 \in D(\phi)$ satisfies $|\nabla\phi|(x_0) = 0$ if and only if $\mathcal{G}(t, x_0) = x_0$ for all $t > 0$.

Theorem (Large time behavior)

Take $x_0 \in D(\phi)$, put $\xi(t) := \mathcal{G}(t, x_0)$ and assume $\lim_{t\to\infty} \phi(\xi(t)) > -\infty$. Then we have $\lim_{t\to\infty} |\nabla\phi|(\xi(t)) = 0$.

Corollary

Take $x_0 \in D(\phi)$, put $\xi(t) := \mathcal{G}(t, x_0)$ and assume that there is a sequence $\{t_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} t_n = \infty$ and $\{\xi(t_n)\}_{n\in\mathbb{N}}$ converges to a point $\bar{x}$. Then $\bar{x}$ is a stationary point of ϕ and $\lim_{t\to\infty} \phi(\xi(t)) = \phi(\bar{x})$.

A Trotter–Kato product formula

Assumption

Let (X, d) be a complete metric space in either Case (I) or Case (II), and assume additionally $D := \operatorname{diam} X < \infty$. For $i = 1, 2$, we consider lsc, λ_i-convex function $\phi_i : X \longrightarrow (-\infty, \infty]$ $(\lambda_i \in \mathbb{R})$ satisfying $D(\phi_1) \cap D(\phi_2) \neq \emptyset$ and the compactness (Assumption (2)).

Given $z_0 \in D(\phi) = D(\phi_1) \cap D(\phi_2)$ and a partition $\mathscr{P}_{\boldsymbol{\tau}}$, we consider the discrete variational schemes for ϕ_1 and ϕ_2 in turn, namely

$$z_{\boldsymbol{\tau}}^0 := z_0, \text{ choose arbitrary } \hat{z}_{\boldsymbol{\tau}}^k \in J_{\tau_k}^{\phi_1}(z_{\boldsymbol{\tau}}^{k-1}) \text{ and } z_{\boldsymbol{\tau}}^k \in J_{\tau_k}^{\phi_2}(\hat{z}_{\boldsymbol{\tau}}^k) \text{ for } k \in \mathbb{N}.$$

The *Trotter–Kato product formula* asserts that $\{z_{\boldsymbol{\tau}}^k\}_{k \geq 0}$ converges to the gradient curve of $\phi := \phi_1 + \phi_2$ emanating from z_0 in an appropriate sense.

Assumption

Given $z_0 \in D(\phi)$ and a partition $\mathscr{P}_{\boldsymbol{\tau}}$, set

$$\delta_{\boldsymbol{\tau}}^k(z_0) := \max\{0, \phi_2(\hat{z}_{\boldsymbol{\tau}}^k) - \phi_2(z_{\boldsymbol{\tau}}^{k-1}), \phi_1(z_{\boldsymbol{\tau}}^k) - \phi_1(\hat{z}_{\boldsymbol{\tau}}^k)\}$$

for $k \in \mathbb{N}$ by suppressing the dependence on the choice of $\{\hat{z}_{\boldsymbol{\tau}}^k, z_{\boldsymbol{\tau}}^k\}_{k\in\mathbb{N}}$. Assume that, for any $\varepsilon, T > 0$, there is $\Delta_\varepsilon^T(z_0) < \infty$ such that

$$\sum_{k=1}^{N} \delta_{\boldsymbol{\tau}}^k(z_0) \leq \Delta_\varepsilon^T(z_0)$$

for any $\mathscr{P}_{\boldsymbol{\tau}}$ with $|\boldsymbol{\tau}| < \varepsilon$, $N \in \mathbb{N}$ with $t_{\boldsymbol{\tau}}^N \leq T$, and for any solution $\{\hat{z}_{\boldsymbol{\tau}}^k, z_{\boldsymbol{\tau}}^k\}_{k\in\mathbb{N}}$. This in particular guarantees that $\hat{z}_{\boldsymbol{\tau}}^k \in D(\phi)$ and $z_{\boldsymbol{\tau}}^k \in D(\phi)$.

Introduce the interpolated curve $\bar{\boldsymbol{z}}_{\boldsymbol{\tau}}$:

$$\bar{\boldsymbol{z}}_{\boldsymbol{\tau}}(0) := z_0, \qquad \bar{\boldsymbol{z}}_{\boldsymbol{\tau}}(t) := z_{\boldsymbol{\tau}}^k \quad \text{for } t \in (t_{\boldsymbol{\tau}}^{k-1}, t_{\boldsymbol{\tau}}^k].$$

Theorem (A Trotter–Kato product formula)

Let the above assumptions be satisfied. Given $z_0 \in D(\phi)$*, the curve* $\bar{\boldsymbol{z}}_{\boldsymbol{\tau}}$ *converges to the gradient curve* $\xi := \mathcal{G}(\cdot, z_0)$ *of* ϕ *(constructed in the previous section) as* $|\boldsymbol{\tau}| \to 0$ *uniformly on each bounded interval* $[0, T]$*.*

Nonsmooth convex optimization

Definition (Proximal Point Algorithm)

Let (X, d) be a complete Alexandrov space either with curvature bounded above or below by κ, and $G \subset X$ be a closed, geodesically convex set satisfying the following:

(1) In the upper curvature bound case, $\mathrm{diam}\, G < \pi/(2\sqrt{\kappa})$ if $\kappa > 0$;

(2) In the lower curvature bound case, $\dim X < \infty$, $\partial X = \emptyset$, and $\mathrm{diam}\, G < \infty$ if $\kappa < 0$. Also $J_\lambda^f(x) := \mathrm{gexp}_x(\lambda \nabla(-f)(x))$.

Let $f_i : G \to (-\infty, \infty]$ be convex, lsc for $i = 1, \ldots, n$. Set $f(x) := \sum_{i=1}^n f_i(x)$ and suppose it is proper. Take $\lambda_k > 0$ s.t. $\sum_{k=0}^\infty \lambda_k = +\infty$, $\sum_{k=0}^\infty \lambda_k^2 < +\infty$. Given $x_0 \in G$ and for each $k \geq 0$ and $1 \leq i \leq n$, we set

$$x_{kn+i} := J_{\lambda_k}^{f_i}(x_{kn+i-1}).$$

Theorem

Let (X, d), $G \subset X$, $f = \sum_{i=1}^{n} f_i$ and $\{\lambda_k\}_{k \geq 0}$ be as above. Assume further that X is locally compact, f_i is L-Lipschitz for some $L \geq 1$ and all i, and that $\inf_G f$ is attained at some point. Then x_m converges to a minimizer of f in G as $m \to \infty$.

Proposition

Let (X, d), $G \subset X$, $f = \sum_{i=1}^{n} f_i$ be as above and further assume that f_i is L-Lipschitz, and that f is K-convex for some $K > 0$. Take $\lambda_k > 0$ with $\lambda_k K < 1$, $\lambda_k \to 0$ and $\sum_{k=0}^{\infty} \lambda_k = +\infty$, and consider the sequence $\{x_m\}_{m \geq 0}$ generated by the above. Then x_m converges to the unique minimizer $y \in G$ of f as $m \to \infty$.

An application: Sturm's law of large numbers

Theorem (Sturm 2002, Annals of Prob.)

Let (X, d) be a $CAT(0)$-space and let $\mathcal{P}^2(X)$ denote the set of all probability measures μ s.t. $\int_X d^2(x,a)d\mu(a) < \infty$. Let $a\#_t b$ denote the unique geodesic between $a, b \in X$. Then for $\mu \in \mathcal{P}^2(X)$

$$\Lambda(\mu) := \operatorname*{arg\,min}_{x \in X} \int_X d^2(x,a)d\mu(a)$$

exists and is unique. Moreover consider an i.i.d. sequence of random variables $\{Y_i\}_{i\in\mathbb{N}}$ with law μ and define

$$\begin{aligned} S_1 &:= Y_1, \\ S_{k+1} &:= S_k \#_{\frac{1}{k+1}} Y_{k+1}. \end{aligned}$$

Then S_k converges to $\Lambda(\mu)$ almost surely, if $\mathrm{supp}(\mu)$ is bounded.

An application: Nodice theorem for the Karcher mean

A deterministic version of Sturm's law (cf. also Holbrook 2012):

Theorem (Lim-Pálfia 2014, Bull. LMS)

Let (X,d) be a CAT(0)-space and let $\mu := \sum_{i=0}^{n-1} \frac{1}{n}\delta_{a_i}$ with $a_i \in X$. Consider the deterministic sequence $\{S_k\}_{k\in\mathbb{N}}$ defined as the inductive sequence of geometric means

$$S_1 := a_0,$$
$$S_{k+1} := S_k \#_{\frac{1}{k+1}} a_{\overline{k}}$$

where $\overline{k} := k \mod (n)$. Then $S_k \to \Lambda(\mu)$ with rate $d(S_k, \Lambda(\mu)) = O(1/k)$.

The above along with Sturm's slln even generalizes to CAT(κ) spaces (Ohta-Pálfia 2015, Yokota 2018) and positive operators (Lim-Pálfia 2020, 2021).

Abstract law of large numbers

Let $G \subset X$ be a closed, geodesically convex set. We assume that (G, d) is separable. Consider the set of all lower semi-continuous, convex functions $f : G \to (-\infty, \infty]$ not identically $+\infty$, denoted by $F(G)$. For $K > 0$, we denote by $F_K(G)$ the subset of all lower semi-continuous, K-convex functions $f : G \to (-\infty, \infty]$ not identically $+\infty$.
Denote by $\mathfrak{P}(F_K(G))$ the set of all complete probability measures on $F_K(G)$ with σ-field generated by the topology of one-sided uniform convergence, such that $g(x) := \int_{F_K(G)} f(x) d\mu(f)$ is lsc $(-\infty, +\infty]$-valued K-convex and there exists $x \in G$ so that $g(x) < +\infty$.

Definition (Variance)

We define the *variance* of $\mu \in \mathfrak{P}(F_K(G))$ by

$$\mathrm{var}(\mu) := \inf_{x \in G} \int_{F_K(G)} f(x) d\mu(f).$$

A fixed $\mu \in \mathfrak{P}(F_K(G))$ can be viewed as the distribution of an $F_K(G)$-valued random variable. $\mathbb{E}\varphi := \int_{F_K(G)} \varphi(f) d\mu(f)$

Definition (Expectation)

Let $\mu \in \mathfrak{P}(F_K(G))$. We define the *expectation* of μ as

$$\mathbb{E}\mu := \arg\min_{x \in G} \int_{F_K(G)} f(x) d\mu(f),$$

which is indeed uniquely determined by the K-convexity of $g(x) = \int_{F_K(G)} f(x) d\mu(f)$.

The above is motivated by the definition of Sturm of the expectation as $\mathbb{E}\nu := \arg\min_{x\in G}\int_G d(x,a)^2 d\nu(a)$ of a probability measure ν supported over G.
Note that $g(\mathbb{E}\mu) = \mathrm{var}(\mu)$. Let L_x denote the evaluation operator at $x \in G$ defined as $L_x f := f(x)$. Clearly L_x is a linear functional on the cone $F_K(G)$.

Proposition (Variance inequality)

Let $\mu \in \mathfrak{P}(F_K(G))$. Then, for all $x \in G$, we have

$$d(x,\mathbb{E}\mu)^2 \leq \frac{2}{K}\mathbb{E}\left(L_x - L_{\mathbb{E}\mu}\right) = \frac{2}{K}\int_{F_K(G)}[f(x) - f(\mathbb{E}\mu)]d\mu(f).$$

Theorem (Law of large numbers)

Let (X, d) *and* $G \subset X$ *be as above. Fix* $\mu \in \mathfrak{P}(F_K(G))$ *supported on L-Lipschitz functions and let* $\{f_k\}_{k\geq 0}$ *denote a sequence of i.i.d. random variables taking values in* $F_K(G)$ *with distribution* μ*. Take a positive sequence* $\{\lambda_k\}_{k\geq 0}$ *with* $\lambda_k K < 1$*,* $\lambda_k \to 0$ *and* $\sum_{k=0}^{\infty} \lambda_k = +\infty$*. Define the sequence* $S_k \in G$ *recursively as*

$$S_{k+1} := J_{\lambda_k}^{f_k}(S_k), \quad k \geq 0,$$

with an arbitrary starting point $S_0 \in G$*, assuming that* $S_k \in G$ *for all* $k \geq 0$ *in the lower curvature bound case. Then* $S_k \to \mathbb{E}\mu$ *almost surely.*

Calculus of variations in CAT(1)

Variational result for convex lsc potential functions: Kuwae-Shioya 2009, Bačák 2015 proves in $CAT(0)$ spaces that continuity in Mosco implies continuity of resolvent and thus continuity of gradient flows.

Definition (Weak convergence)

x_n converges weakly to x, if $P_\gamma(x_n) \to x$ for any geodesic $\gamma : [0,1] \mapsto X$ with $\gamma(0) = x$.

Weak convergence makes sense on geodesically convex sets in $CAT(1)$ and sequences included in convex balls have weak cluster points.

Lemma (CAT(1) variant of Bačák's Lemma)

Let (X, d) be a CAT(1) space. Let $x_n, x \in X$ such that $d(x_n, x) < \pi/2$ for all $n \in \mathbb{N}$. Then $x_n \to x$ if and only if $x_n \overset{w}{\to} x$ and $d(x_n, y) \to d(x, y)$ for some $y \in X$ such that $d(x, y) < \pi/2$.

Lemma

Let (X, d) be a CAT(1) space with $\mathrm{diam}(X) < \pi$. Then if $(C_i)_{i \in I}$ is a non-increasing family of bounded closed convex sets in X for an index set I, we have $\cap_{i \in I} C_i \neq \emptyset$.

Lemma

Let $\mathrm{diam}(X) < \pi$. Let $f : X \mapsto (-\infty, \infty]$ be a convex lsc function. Then f is bounded below on bounded sets.

Theorem (Theorem 3.5., Kell 2014)

Let $\mathrm{diam}(X) < \pi$. Then closed convex sets are weakly closed.

Lemma (Proposition 3.8., Kell 2014)

Let $\mathrm{diam}(X) < \pi$. Let $f : X \mapsto (-\infty, \infty]$ be a quasiconvex lsc function. Then f is weakly lsc. In particular $x \to d^2(a, x)$ is weakly lsc on $B_a(\pi/2)$.

Theorem (Yokota's Theorem A, 2016)

Let $\mathrm{diam}(X) < \pi$. *There exists a jointly* κ*-convex lsc function* $\Phi : X \times X \to [0, \infty)$ *for some* $\kappa > 0$.

Lemma (Ekeland principle, 1979)

Given $x_0 \in X$ *and a lsc function* $f : X \mapsto (-\infty, \infty]$ *that is bounded below, there exist* $\alpha, \beta \geq 0$ *such that for all* $x \in X$

$$f(x) \geq -\alpha d(x, x_0) - \beta.$$

Definition (Mosco convergence)

A sequence of lsc functions $\phi_n : X \mapsto \overline{\mathbb{R}}$ said to converge to $\phi : X \mapsto \overline{\mathbb{R}}$ in the sense of Mosco if, for any $x \in X$, we have

(M1) $f(x) \leq \liminf_{n\to\infty} f_n(x_n)$ whenever $x_n \xrightarrow{w} x$,

(M2) there exists an $(y_n) \subseteq X$, such that $y_n \to x$ and $f_n(y_n) \to f(x)$.

Mosco convergence implies uniform minimization

Proposition (Ekeland principle, bounded case)

Let $\mathrm{diam}(X) < \pi$. *Given* $x_0 \in X$ *and a uniformly proper sequence of lsc* λ*-convex functions* $f_n : X \mapsto (-\infty, \infty]$ *that is Mosco converging to* $f : X \mapsto (-\infty, \infty]$, *there exist* $\alpha, \beta \geq 0$ *such that*

$$f_n(x) \geq -\alpha d(x, x_0) - \beta$$

for all $x \in X$ *and* $n \in \mathbb{N}$.

Theorem

Let $\mathrm{diam}(X) < \pi$ *and* $f_n : X \mapsto (-\infty, \infty]$ *a uniformly proper sequence of lsc* λ*-convex functions that is Mosco converging to* $f : X \mapsto (-\infty, \infty]$. *Then for any small enough* $\tau > 0$ *and* $x \in D(f)$

$$\lim_{n\to\infty} (f_n)_\tau(x) = f_\tau(x), \quad \lim_{n\to\infty} J_\tau^{f_n}(x) = J_\tau^f(x). \tag{a}$$

Remark

If $J_\tau^f(x)$ *is not unique in the above Theorem, then it still follows that all weak cluster points of* $J_\tau^{f_n}(x)$ *are in fact strong cluster points and are in* $J_\tau^f(x)$*.*

Theorem

Let $f_n : X \mapsto (-\infty, \infty]$ *be a uniformly proper, uniformly lower bounded sequence of lsc functions that is Mosco converging to* $f : X \mapsto (-\infty, \infty]$*. Then* (a) *holds for any small enough* $\tau > 0$ *and* $x \in D(f)$*.*

Theorem

Let $f_n : X \mapsto (-\infty, \infty]$ *be a uniformly proper sequence of L-Lipschitz functions that is Mosco converging to* $f : X \mapsto \overline{\mathbb{R}}$*. Then* (a) *holds for any small enough* $\tau > 0$*.*

- H. Gál and M. Pálfia, *Convergence of semi-convex functions in $CAT(1)$ spaces*, in preparation.
- Y. Lim and M. Pálfia, *Strong law of large numbers for the L^1-Karcher mean*, Journal of Functional Analysis 279:7 (2020), 108672.
- Y. Lim and M. Pálfia, *Existence and uniqueness of the L^1-Karcher mean*, Adv. Math. 376 (2021), 107435.
- S. Ohta and M. Pálfia, *Gradient flows and a Trotter-Kato formula of semi-convex functions on CAT(1)-spaces*, Amer. J. Math. 139:4 (2017), pp. 937–965.
- S. Ohta and M. Pálfia, *Discrete-time gradient flows and law of large numbers in Alexandrov spaces*, Calc. Var. PDE 54:2 (2015), pp. 1591–1610.

- L. Ambrosio, N. Gigli and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures.* Second edition, Birkhäuser Verlag, Basel, 2008.
- M. Bačák, *Convergence Of Nonlinear Semigroups Under Nonpositive Curvature*, Trans. Amer. Math. Soc. 367:6 (2015), pp. 3929–3953.
- K. Kuwae and T. Shioya, *Variational convergence over metric spaces*, Trans. Amer. Math. Soc. 360:1 (2008), pp. 35–75.
- K.-T. Sturm, *Nonlinear martingale theory for processes with values in metric spaces of nonpositive curvature*, Ann. Probab. **30** (2002), pp. 1195–1222.

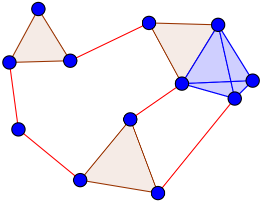
Thank you for your kind attention!

Computing homology robustly: The geometry of normed chain complexes

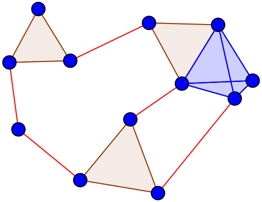
Pierre Pansu, Université Paris-Saclay

October 11th, 2023

A simplicial complex is made of simplices of various dimensions. *Simplicial chains* are linear combinations of simplices. The *boundary* of a simplex is a chain, whence a linear map ∂ and its adjoint d, which satisfy $\partial \circ \partial = 0$ and $d \circ d = 0$.

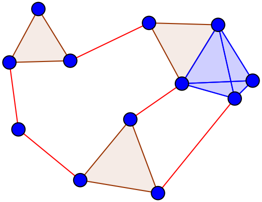


A simplicial complex is made of simplices of various dimensions. *Simplicial chains* are linear combinations of simplices. The *boundary* of a simplex is a chain, whence a linear map ∂ and its adjoint d, which satisfy $\partial \circ \partial = 0$ and $d \circ d = 0$.



The homology (resp. cohomology) of the simplicial complex is $\mathrm{Ker}(\partial)/\mathrm{Im}(\partial)$ (resp. $\mathrm{Ker}(d)/\mathrm{Im}(d)$).

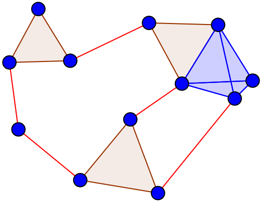
A simplicial complex is made of simplices of various dimensions. *Simplicial chains* are linear combinations of simplices. The *boundary* of a simplex is a chain, whence a linear map ∂ and its adjoint d, which satisfy $\partial \circ \partial = 0$ and $d \circ d = 0$.



The homology (resp. cohomology) of the simplicial complex is $\mathrm{Ker}(\partial)/\mathrm{Im}(\partial)$ (resp. $\mathrm{Ker}(d)/\mathrm{Im}(d)$).

Simplicial chains and cochains can be equipped with ℓ^p norms.

A simplicial complex is made of simplices of various dimensions. *Simplicial chains* are linear combinations of simplices. The *boundary* of a simplex is a chain, whence a linear map ∂ and its adjoint d, which satisfy $\partial \circ \partial = 0$ and $d \circ d = 0$.



The homology (resp. cohomology) of the simplicial complex is $\mathrm{Ker}(\partial)/\mathrm{Im}(\partial)$ (resp. $\mathrm{Ker}(d)/\mathrm{Im}(d)$).

Simplicial chains and cochains can be equipped with ℓ^p norms.

In general, a *normed chain complex* is a normed vector space B equipped with a linear map $d : B \to B$ such that $d \circ d = 0$.

When $F : B_1 \to B_2$ is a linear bijection, the robustness of the resolution of the equation

$$Fx = y$$

is governed by the *conditioning number*

$$\kappa(F) = |F||F^{-1}|.$$

When $F : B_1 \to B_2$ is a linear bijection, the robustness of the resolution of the equation

$$Fx = y$$

is governed by the *conditioning number*

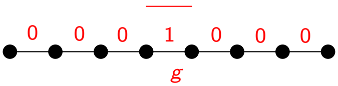
$$\kappa(F) = |F||F^{-1}|.$$

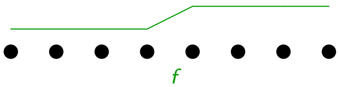
For normed chain complexes, we first turn d into a bijection $\bar{d} : B/\mathrm{Ker}(d) \to \mathrm{Im}(d)$, and set

$$\kappa(B) := |\bar{d}||\bar{d}^{-1}|.$$

Example. The n-stick satisfies $H^1 = 0$. The 1-cochain g equal to $\bar{1}$ on the central edge and 0 elsewhere can be written df where

$$\|g\|_p = 1, \quad \|f\|_p \sim n^{1/p}.$$





Example. The n-stick satisfies $H^1 = 0$. The 1-cochain g equal to $\bar{1}$ on the central edge and 0 elsewhere can be written df where

$$\|g\|_p = 1, \quad \|f\|_p \sim n^{1/p}.$$

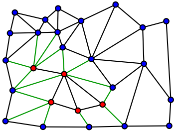
When n is large, solving $df = g$ is unstable. The computation of cohomology is ill-conditioned.

Definition

The **conditionning number** *of a graph X is $\kappa(X, p, \mathbf{k}) = |\bar{d}||\bar{d}^{-1}|$ where $\bar{d} : C^0(X, \mathbf{k})/\mathrm{Ker}(d) \to dC^0(X, \mathbf{k})$. (It depends on p and on the field $\mathbf{k}$).*

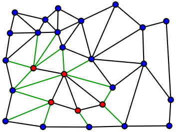
Isoperimetry = the art of cutting space apart.

$|A| = 5$, $|\partial A| = 15$.



Isoperimetry = the art of cutting space apart.

$|A| = 5$, $|\partial A| = 15$.



Definition

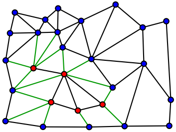
Cheeger's constant *$h(X)$ of a graph X is the largest h such that for every set A of vertices such that $|A| \leq \frac{1}{2}|X|$,*

$$|\partial A| \geq h\,|A|.$$

Here, ∂A is the set of edges connecting A to its complement.

Isoperimetry = the art of cutting space apart.

$|A| = 5$, $|\partial A| = 15$.



Definition

Cheeger's constant $h(X)$ *of a graph* X *is the largest* h *such that for every set* A *of vertices such that* $|A| \leq \frac{1}{2}|X|$*,*

$$|\partial A| \geq h\,|A|.$$

Here, ∂A *is the set of edges connecting* A *to its complement.*

Proposition

$$h(X) = \frac{2}{\kappa(X,1,\mathbb{F}_2)} = 2(\|\bar{d}\|_{1\to 1}\|\bar{d}^{-1}\|_{1\to 1})^{-1} \text{ over } \mathbb{F}_2.$$

Proposition

Let Δ be the self-adjoint operator corresponding to the quadratic form $f \mapsto \|df\|_2^2 = \langle f, \Delta f\rangle$. Let $\lambda_1 \leq \lambda_2 \leq \cdots$ denote its eigenvalues. If the graph X is connected, then $\lambda_1 = 0$ and

$$\lambda_2 = (\|\bar{d}^{-1}\|_{2\to 2})^{-2}.$$

Proposition

Let Δ *be the self-adjoint operator corresponding to the quadratic form* $f \mapsto \|df\|_2^2 = \langle f, \Delta f\rangle$*. Let* $\lambda_1 \leq \lambda_2 \leq \cdots$ *denote its eigenvalues. If the graph* X *is connected, then* $\lambda_1 = 0$ *and*

$$\lambda_2 = (\|\bar{d}^{-1}\|_{2\to 2})^{-2}.$$

In particular,

$$2\lambda_2^{-1} \leq \kappa_0(X, 2, \mathbb{R})^2 \leq 4\lambda_2^{-1}.$$

Proposition

Let Δ be the self-adjoint operator corresponding to the quadratic form $f \mapsto \|df\|_2^2 = \langle f, \Delta f \rangle$. Let $\lambda_1 \leq \lambda_2 \leq \cdots$ denote its eigenvalues. If the graph X is connected, then $\lambda_1 = 0$ and

$$\lambda_2 = (\|\bar{d}^{-1}\|_{2\to 2})^{-2}.$$

In particular,

$$2\lambda_2^{-1} \leq \kappa_0(X, 2, \mathbb{R})^2 \leq 4\lambda_2^{-1}.$$

λ_2 is known as the *spectral gap* of the graph.
It governs the speed at which a random walk on the graph is mixing. In particular, the possibility of picking a vertex at random.

Proposition

Let Δ be the self-adjoint operator corresponding to the quadratic form $f \mapsto \|df\|_2^2 = \langle f, \Delta f \rangle$. Let $\lambda_1 \leq \lambda_2 \leq \cdots$ denote its eigenvalues. If the graph X is connected, then $\lambda_1 = 0$ and

$$\lambda_2 = (\|\bar{d}^{-1}\|_{2\to 2})^{-2}.$$

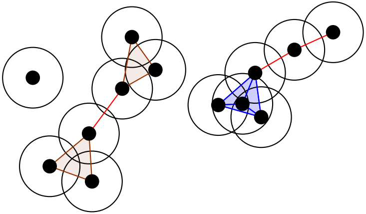
In particular,

$$2\lambda_2^{-1} \leq \kappa_0(X, 2, \mathbb{R})^2 \leq 4\lambda_2^{-1}.$$

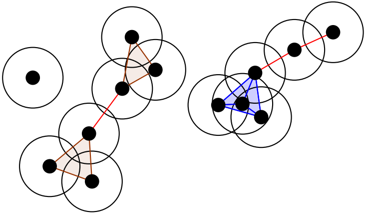
λ_2 is known as the *spectral gap* of the graph.
It governs the speed at which a random walk on the graph is mixing. In particular, the possibility of picking a vertex at random.

Morality. Normed chain complexes contain interesting information, beyond their mere homology.

Given a metric space X, a finite subset $Y \subset X$ and $r > 0$, the *Čech simplicial complex* Y_r has a simplex $(y_0, \dots, y_k)$ each time $\bigcap_i B(y_i, r) \neq \emptyset$. Let $C_\cdot^r$ denote the simplicial chains of Y_r.



Given a metric space X, a finite subset $Y \subset X$ and $r > 0$, the *Čech simplicial complex* Y_r has a simplex $(y_0, \ldots, y_k)$ each time $\bigcap_i B(y_i, r) \neq \emptyset$. Let $C_\cdot^r$ denote the simplicial chains of Y_r.



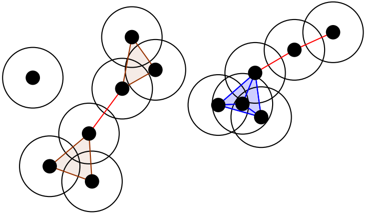
Theorem (Bobrowski-Weinberger 2017)

Fix $r < \frac{1}{2}$ and $1 \leq k \leq d$. Let Y be an n-sample picked at random on the standard d-torus. Then, with high probability, the k-homology of Y_r coincides with the homology of the torus as soon as

$$\omega_d r^d\, n \gg \log n + k \log \log n,$$

and this fails if $\omega_d r^d\, n \ll \log n + (k-2) \log \log n$. If $k = 0$, the threshold is $2^{-d} \log n$.

Given a metric space X, a finite subset $Y \subset X$ and $r > 0$, the *Čech simplicial complex* Y_r has a simplex $(y_0, \dots, y_k)$ each time $\bigcap_i B(y_i, r) \neq \emptyset$. Let $C_\cdot^r$ denote the simplicial chains of Y_r.



Theorem (Bobrowski-Weinberger 2017)

Fix $r < \frac{1}{2}$ and $1 \leq k \leq d$. Let Y be an n-sample picked at random on the standard d-torus. Then, with high probability, the k-homology of Y_r coincides with the homology of the torus as soon as

$$\omega_d r^d\, n \gg \log n + k \log\log n,$$

and this fails if $\omega_d r^d\, n \ll \log n + (k-2)\log\log n$. If $k = 0$, the threshold is $2^{-d}\log n$.

Question. Can one say that the chain complexes $C_\cdot^r$ converge to some chain complex attached to the torus?

In order to define a distance between normed chain complexes, the first idea is to measure conditioning numbers of isomorphisms.

Definition

Let $B_1 \overset{d_1}{\to} B_1$ *and* $B_2 \overset{d_2}{\to} B_2$ *be normed chain complexes. The* **Banach-Mazur distance** $\mathrm{BMDist}(B_1, B_2)$ *is the infimum of* $\log(|F||F^{-1}|)$ *over all isomorphisms* $F : B_1 \to B_2$ *duch that* $Fd_1 = d_2F$.

In order to define a distance between normed chain complexes, the first idea is to measure conditioning numbers of isomorphisms.

Definition

Let $B_1 \overset{d_1}{\to} B_1$ *and* $B_2 \overset{d_2}{\to} B_2$ *be normed chain complexes. The* **Banach-Mazur distance** $\mathrm{BMDist}(B_1, B_2)$ *is the infimum of* $\log(|F||F^{-1}|)$ *over all isomorphisms* $F : B_1 \to B_2$ *duch that* $Fd_1 = d_2F$*.*

This is too restrictive: this implies $\dim(B_1) = \dim(B_2)$.

The second idea is too measure the size of homotopies.

Definition

Let $B_1 \overset{d_1}{\to} B_1$ *and* $B_2 \overset{d_2}{\to} B_2$ *be normed chain complexes. Consider all bounded homotopies, i.e.*

- *bounded morphisms* $F_1 : B_1 \to B_2$ *and* $F_2 : B_2 \to B_1$ *such that*

$$d_2 F_1 = F_1 d_1, \quad d_1 F_2 = F_2 d_2,$$

- *bounded operators* $Q_1 : B_1 \to B_1$ *and* $Q_2 : B_2 \to B_2$ *such that*

$$1 - F_2 F_1 = d_1 Q_1 + Q_1 d_1, \quad 1 - F_1 F_2 = d_2 Q_2 + Q_2 d_2.$$

Let $q = \max\{|Q_1|, |Q_2|\}$*,* $f = \max\{1, |F_1||F_2|\}$*. The* **homotopy distance** $\mathrm{HomDist}(B_1, B_2)$ *is the infimum over all homotopies of* $\min\{\frac{q}{f} + \log f, \frac{f}{q} + \log q\}$.

The second idea is too measure the size of homotopies.

Definition

Let $B_1 \overset{d_1}{\to} B_1$ *and* $B_2 \overset{d_2}{\to} B_2$ *be normed chain complexes. Consider all bounded homotopies, i.e.*

- *bounded morphisms* $F_1 : B_1 \to B_2$ *and* $F_2 : B_2 \to B_1$ *such that*

$$d_2 F_1 = F_1 d_1, \quad d_1 F_2 = F_2 d_2,$$

- *bounded operators* $Q_1 : B_1 \to B_1$ *and* $Q_2 : B_2 \to B_2$ *such that*

$$1 - F_2 F_1 = d_1 Q_1 + Q_1 d_1, \quad 1 - F_1 F_2 = d_2 Q_2 + Q_2 d_2.$$

Let $q = \max\{|Q_1|, |Q_2|\}$, $f = \max\{1, |F_1||F_2|\}$. *The* **homotopy distance** $\mathrm{HomDist}(B_1, B_2)$ *is the infimum over all homotopies of* $\min\{\frac{q}{f} + \log f, \frac{f}{q} + \log q\}$.

The weird expression guarantees a triangle inequality.

Definition

Let Null denote the set of null normed chain complexes (i.e. with $d = 0$). Denote by

$$\mathrm{ND}(B) = \mathrm{HomDist}(B, \mathit{Null}), \quad \mathrm{NH}(B) = |\bar{d}^{-1}|.$$

Definition

Let Null denote the set of null normed chain complexes (i.e. with $d = 0$). Denote by

$$\mathrm{ND}(B) = \mathrm{HomDist}(B, \mathit{Null}), \quad \mathrm{NH}(B) = |\bar{d}^{-1}|.$$

Fact. ND is continuous. NH is continuous on the complement of *Null*.

Definition

Let Null denote the set of null normed chain complexes (i.e. with $d = 0$). Denote by

$$\mathrm{ND}(B) = \mathrm{HomDist}(B, \mathit{Null}), \quad \mathrm{NH}(B) = |\bar{d}^{-1}|.$$

Fact. ND is continuous. NH is continuous on the complement of *Null*.

Remark. ND is a function of NH for prehilbertian complexes, but not in general.

Definition

Let Null denote the set of null normed chain complexes (i.e. with $d = 0$). Denote by

$$\mathrm{ND}(B) = \mathrm{HomDist}(B, \mathit{Null}), \quad \mathrm{NH}(B) = |\bar{d}^{-1}|.$$

Fact. ND is continuous. NH is continuous on the complement of *Null*.

Remark. ND is a function of NH for prehilbertian complexes, but not in general.

Definition

Let B be a normed chain complex. Let $\bar{B} = B/\mathrm{Ker}(d)$ and $\bar{d} : \bar{B} \to \mathrm{Im}(d)$.
The **singular values** *of B are the numbers*

$$\sigma_j = \inf\{s \geq 0\,;\, \exists L \subset \bar{B} \text{ subvectorspace such that } \dim(L) \geq j \text{ and } \forall \bar{x} \in L,\ |\bar{d}\bar{x}| \leq s|\bar{x}|\}.$$

Definition

Let Null denote the set of null normed chain complexes (i.e. with $d = 0$). Denote by

$$\mathrm{ND}(B) = \mathrm{HomDist}(B, Null), \quad \mathrm{NH}(B) = |\bar{d}^{-1}|.$$

Fact. ND is continuous. NH is continuous on the complement of *Null*.

Remark. ND is a function of NH for prehilbertian complexes, but not in general.

Definition

Let B be a normed chain complex. Let $\bar{B} = B/\mathrm{Ker}(d)$ and $\bar{d} : \bar{B} \to \mathrm{Im}(d)$.
The **singular values** *of B are the numbers*

$$\sigma_j = \inf\{s \geq 0\,;\, \exists L \subset \bar{B} \textit{ subvectorspace such that } \dim(L) \geq j \textit{ and } \forall \bar{x} \in L,\ |\bar{d}\bar{x}| \ \leq s|\bar{x}|\}.$$

Fact. Each σ_j is continuous in homotopy distance.

Definition

Say a normed chain complex B is **precompact** *if it is not null and belongs to the closure of finite dimensional normed chain complexes.*

Definition

Say a normed chain complex B is **precompact** *if it is not null and belongs to the closure of finite dimensional normed chain complexes.*

Example. The (de Rham) complex of smooth differential forms on a smooth compact Riemannian manifold, in its L^2 norm, is precompact.

Definition

Say a normed chain complex B is **precompact** *if it is not null and belongs to the closure of finite dimensional normed chain complexes.*

Example. The (de Rham) complex of smooth differential forms on a smooth compact Riemannian manifold, in its L^2 norm, is precompact.

Fact. A prehilbertian chain complex is precompact $\iff$ its singular values form a finite sequence that tends to $+\infty$.

Proposition

Let B_i be precompact prehilbertian chain complexes. Then B_i converges to B $\iff$ for every j, $\sigma_j(B_i)$ tends to $\sigma_j(B)$.

Analogy between normed chain complexes and metric spaces.

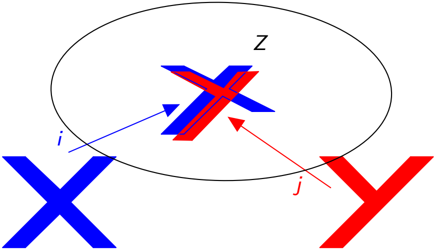
Metric space	Normed chain complex
Gromov-Hausdorff distance	Homotopy distance
Point	?
Bounded	?
Precompact	?
Compactness criterion (Gromov)	?

Analogy between normed chain complexes and metric spaces.

Metric space	Normed chain complex
Gromov-Hausdorff distance	Homotopy distance
Point	?
Bounded	?
Precompact	?
Compactness criterion (Gromov)	?

Let X, Y be metric spaces.

$$GHDist(X,Y) = \inf\{HDist_Z(i(X),j(Y))\,;\, Z \text{ metric space},\\ i : X \to Z,\, j : Y \to Z \text{ isometric embeddings}\}.$$



Analogy between normed chain complexes and metric spaces.

Metric space	Normed chain complex
Gromov-Hausdorff distance	Homotopy distance
Point	Null complex (i.e. $d = 0$)
Bounded	?
Precompact	?
Compactness criterion (Gromov)	?

Analogy between normed chain complexes and metric spaces.

Metric space	Normed chain complex
Gromov-Hausdorff distance	Homotopy distance
Point	Null complex (i.e. $d = 0$)
Bounded	Homotopic to a null complex
Precompact	?
Compactness criterion (Gromov)	?

B is homotopic to a null complex $\iff \mathrm{ND}(B) < \infty$.

One can think of $\mathrm{ND}(B) = \mathrm{HomDist}(B, \mathit{Null})$ as an analogue of diameter.

Analogy between normed chain complexes and metric spaces.

Metric space	Normed chain complex
Gromov-Hausdorff distance	Homotopy distance
Point	Null complex (i.e. $d = 0$)
Bounded	Homotopic to a null complex
Precompact	In the closure of finite dim. complexes
Compactness criterion (Gromov)	?

B is precompact $\implies$ B has a finite sequence of singular values that tends to $+\infty$ ($\iff$ if B is prehilbertian).

Analogy between normed chain complexes and metric spaces.

Metric space	Normed chain complex
Gromov-Hausdorff distance	Homotopy distance
Point	Null complex (i.e. $d = 0$)
Bounded	Homotopic to a null complex
Precompact	In the closure of finite dim. complexes
Compactness criterion (Gromov)	????

Definition

X precompact metric space, $\epsilon > 0$. The **covering number** *$N(X, \epsilon)$ is the minimal number of ϵ-balls that can cover X.*

Analogy between normed chain complexes and metric spaces.

Metric space	Normed chain complex
Gromov-Hausdorff distance	Homotopy distance
Point	Null complex (i.e. $d = 0$)
Bounded	Homotopic to a null complex
Precompact	In the closure of finite dim. complexes
Compactness criterion (Gromov)	????

Definition

X precompact metric space, $\epsilon > 0$. The **covering number** *$N(X, \epsilon)$ is the minimal number of ϵ-balls that can cover X.*

Theorem (Gromov's compactness criterion)

A collection $\mathcal{T}$ of precompact metric spaces is precompact in Gromov-Hausdorff distance if and only if there is a function ν which serves as a covering number for all spaces in $\mathcal{T}$, i.e.

$$\forall \epsilon > 0, \quad \forall X \in \mathcal{T}, \quad N(X, \epsilon) \leq \nu(\epsilon).$$

Analogy between normed chain complexes and metric spaces.

Metric space	Normed chain complex
Gromov-Hausdorff distance	Homotopy distance
Point	Null complex (i.e. $d = 0$)
Bounded	Homotopic to a null complex
Precompact	In the closure of finite dim. complexes
Compactness criterion (Gromov)	????

Definition

Let (B, d) be a normed chain complex that belongs to the closure of finite dimensional normed complexes. Its **profile** *is the smallest function $\pi = (\pi_d, \pi_c) : (0, +\infty) \to (0, +\infty)^2$ with the following property. For every $\epsilon > 0$, there exists a finite-dimensional normed complex (B', d') such that*

$$HomDist(B, B') < \epsilon, \quad \dim(B') \leq \pi_d(\epsilon), \quad \kappa(B', d') \leq \pi_c(\epsilon).$$

Analogy between normed chain complexes and metric spaces.

Metric space	Normed chain complex
Gromov-Hausdorff distance	Homotopy distance
Point	Null complex (i.e. $d = 0$)
Bounded	Homotopic to a null complex
Precompact	In the closure of finite dim. complexes
Compactness criterion (Gromov)	????

Definition

Let (B, d) be a normed chain complex that belongs to the closure of finite dimensional normed complexes. Its **profile** *is the smallest function $\pi = (\pi_d, \pi_c) : (0, +\infty) \to (0, +\infty)^2$ with the following property. For every $\epsilon > 0$, there exists a finite-dimensional normed complex (B', d') such that*

$$HomDist(B, B') < \epsilon, \quad \dim(B') \leq \pi_d(\epsilon), \quad \kappa(B', d') \leq \pi_c(\epsilon).$$

Theorem

A collection of nonnull normed chain complexes is precompact if and only if a same profile serves for all and the distances to null complexes are bounded below.

Lemma

Let B be a prehilbertian chain complex. Then the profile of B is determined by the asymptotics of eigenvalues,

$$\pi_d(\epsilon) \leq \mathrm{Card}\{\lambda \in \mathrm{spectrum}(d^* d)\,;\, \lambda < \frac{1}{\epsilon^2}\}, \quad \pi_c(\epsilon) \leq \frac{1}{\epsilon\sqrt{\lambda_2}}.$$

Lemma

Let B be a prehilbertian chain complex. Then the profile of B is determined by the asymptotics of eigenvalues,

$$\pi_d(\epsilon) \leq \mathrm{Card}\{\lambda \in \mathrm{spectrum}(d^* d)\,;\,\lambda < \frac{1}{\epsilon^2}\}, \quad \pi_c(\epsilon) \leq \frac{1}{\epsilon\sqrt{\lambda_2}}.$$

Example

Let M be a smooth compact Riemannian manifold. Consider the (de Rham) complex of smooth differential forms on M in its L^2 norm. Its profile satisfies $\pi_d(\epsilon) \leq C\,\epsilon^{-N}$, where $N = \dim(M)$.

Lemma

Let B be a prehilbertian chain complex. Then the profile of B is determined by the asymptotics of eigenvalues,

$$\pi_d(\epsilon) \leq \mathrm{Card}\{\lambda \in \mathrm{spectrum}(d^* d)\,;\, \lambda < \frac{1}{\epsilon^2}\}, \quad \pi_c(\epsilon) \leq \frac{1}{\epsilon\sqrt{\lambda_2}}.$$

Example

Let M be a smooth compact Riemannian manifold. Consider the (de Rham) complex of smooth differential forms on M in its L^2 norm. Its profile satisfies $\pi_d(\epsilon) \leq C\,\epsilon^{-N}$, where $N = \dim(M)$.

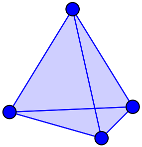
Conjecture

Consider finer and finer triangulations of a fixed compact manifold. The corresponding complexes of simplicial cochains in their weighted ℓ^p norms form a precompact family.

Here, the weight of a simplex is a function of its volume.

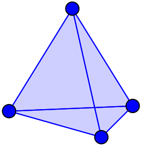
Let Y be a finite metric space. The *complete simplicial complex* Δ_Y on Y takes as simplices all tuples of points of Y. Pick a function of the diameter as a weight. Use weighted ℓ^p norms on cochains. This gives a normed chain complex $C^{\cdot}(Y)$.

Let Y be a finite metric space. The *complete simplicial complex* Δ_Y on Y takes as simplices all tuples of points of Y. Pick a function of the diameter as a weight. Use weighted ℓ^p norms on cochains. This gives a normed chain complex $C^\cdot(Y)$.



The complete simplicial complex on 4 points.

Let Y be a finite metric space. The *complete simplicial complex* Δ_Y on Y takes as simplices all tuples of points of Y. Pick a function of the diameter as a weight. Use weighted ℓ^p norms on cochains. This gives a normed chain complex $C^{\cdot}(Y)$.

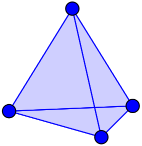


The complete simplicial complex on 4 points.

Let (X,μ) be a metric measure space. Same construction with the same weight w and $L^p(\mu^{\otimes \cdot})$ norms yields a normed chain complex $C^{\cdot}(X)$.
Example. 1-cochains are functions c on $X\times X$. The squared weighted L^2 norm is

$$\int_{X\times X} w(|x-x'|)|c(x,x')|^2\,d\mu(x)\,d\mu(x').$$

Let Y be a finite metric space. The *complete simplicial complex* Δ_Y on Y takes as simplices all tuples of points of Y. Pick a function of the diameter as a weight. Use weighted ℓ^p norms on cochains. This gives a normed chain complex $C^{\cdot}(Y)$.



The complete simplicial complex on 4 points.

Let (X, μ) be a metric measure space. Same construction with the same weight w and $L^p(\mu^{\otimes \cdot})$ norms yields a normed chain complex $C^{\cdot}(X)$.
Example. 1-cochains are functions c on $X \times X$. The squared weighted L^2 norm is

$$\int_{X\times X} w(|x-x'|)|c(x,x')|^2 \, d\mu(x) \, d\mu(x').$$

Question. Given a metric measure space (X, μ) and a finite sample $Y \subset X$. Does $C^{\cdot}(Y)$ converge to $C^{\cdot}(X)$?

Online Learning with Exponential Weights in Metric Spaces under the Measure Contraction Property

Quentin Paris

HSE University

Outline

Introduction

- Online Learning in $\mathbb{R}^p$
- Exponentially Weighted Average (EWA) forecaster
- Performance analysis

Exponential Weights in Metric spaces

- Barycenters
- EWA forecaster in metric spaces
- Measure Contraction Property
- Performance of the EWB forecaster

Jensen Inequality

- Alexandrov curvature bounds
- Connection with MCP property
- $\text{Alex}(M) \geq \kappa \Rightarrow$ Jensen's inequality
- Open question

Introduction

Introduction

Classical setup

Introduction

Classical setup → $M \subset \mathbb{R}^p$ convex

Classical setup

→ $M \subset \mathbb{R}^P$ convex

→ $\mathcal{L}$: Set of convex "loss" functions

$\ell : M \to \mathbb{R}$

Introduction

Classical setup

→ $M \subset \mathbb{R}^p$ convex

→ $\mathcal{L}$: Set of convex "loss" functions

$\ell : M \to \mathbb{R}$

Repeated game

For all $t \geq 1$

Classical setup

→ $M \subset \mathbb{R}^p$ convex

→ $\mathcal{L}$: Set of convex "loss" functions

$$\ell : M \to \mathbb{R}$$

Repeated game

For all $t \geq 1$ → Player chooses $x_t \in M$

Introduction

Classical setup

→ $M \subset \mathbb{R}^p$ convex

→ $\mathcal{L}$: Set of convex "loss" functions

$\ell : M \to \mathbb{R}$

Repeated game

For all $t \geqslant 1$

→ Player chooses $x_t \in M$

→ "Environment" reveals $\ell_t \in \mathcal{L}$

Online Learning in $\mathbb{R}^p$

Classical setup

→ $M \subset \mathbb{R}^p$ convex

→ $\mathcal{L}$: Set of convex "loss" functions

$\ell : M \to \mathbb{R}$

Repeated game

For all $t \geq 1$

→ Player chooses $x_t \in M$

→ "Environment" reveals $\ell_t \in \mathcal{L}$

→ Player incurs loss $\ell_t(x_t)$

and moves on to next round

Introduction

Performance measure : Regret

$$R_n := \sup_{(\ell_1,\dots,\ell_n)\in\mathcal{L}^n} \left\{ \sum_{t=1}^{n} \ell_t(x_t) - \inf_{x\in M} \sum_{t=1}^{n} \ell_t(x) \right\}$$

Introduction

Performance measure : Regret

$$R_n := \sup_{(\ell_1,\dots,\ell_n)\in\mathcal{L}^n} \left\{ \underbrace{\sum_{t=1}^{n} \ell_t(x_t)}_{\textcircled{1}} - \inf_{x\in M} \sum_{t=1}^{n} \ell_t(x) \right\}$$

Encodes ① Cumulative loss of player

Introduction

Performance measure : Regret

$$R_n := \sup_{(\ell_1,\dots,\ell_n)\in\mathcal{L}^n} \Big\{ \underbrace{\sum_{t=1}^{n} \ell_t(x_t)}_{①} - \underbrace{\inf_{x\in M} \sum_{t=1}^{n} \ell_t(x)}_{②} \Big\}$$

Encodes ① Cumulative loss of player

② Competitive benchmark

Introduction

Performance measure : Regret

$$R_n := \underbrace{\sup_{(\ell_1,\dots,\ell_n)\in\mathcal{L}^n}}_{③} \left\{ \underbrace{\sum_{t=1}^{n} \ell_t(x_t)}_{①} - \underbrace{\inf_{x\in M} \sum_{t=1}^{n} \ell_t(x)}_{②} \right\}$$

Encodes
① Cumulative loss of player
② Competitive benchmark
③ "Worst case" point of view

Introduction

Exponentially Weighted Average (EWA) forecaster

Introduction

Exponentially Weighted Average (EWA) forecaster

- Assume $M \subset \mathbb{R}^p$ is a convex body

Exponentially Weighted Average (EWA) forecaster

- Assume $M \subset \mathbb{R}^p$ is a convex body

- Define

$$x_t := \int_M x \, m_t(dx)$$

Exponentially Weighted Average (EWA) forecaster

- Assume $M \subset \mathbb{R}^p$ is a convex body
- Define

$$x_t := \int_M x \, m_t(dx)$$

$$\begin{cases} m_1 := \mathrm{Unif}_M \\ \end{cases}$$

Exponentially Weighted Average (EWA) forecaster

- Assume $M \subset \mathbb{R}^p$ is a convex body

- Define

$$x_t := \int_M x \, m_t(dx)$$

$$\begin{cases} m_1 := \mathrm{Unif}_M \\ m_{t+1}(dx) := \dfrac{\exp(-\beta \ell_t(x))}{Z_{t+1}} \, m_t(dx) \end{cases}$$

Exponentially Weighted Average (EWA) forecaster

- Assume $M \subset \mathbb{R}^p$ is a convex body
- Define

$$x_t := \int_M x \, m_t(dx)$$

$\beta > 0$, parameter of the algorithm

$$\begin{cases} m_1 := \mathrm{Unif}_M \\ m_{t+1}(dx) := \dfrac{\exp\left(-\beta \ell_t(x)\right)}{Z_{t+1}} \, m_t(dx) \end{cases}$$

Remarks

$$m_{t+1}(dx) := \frac{\exp\left(-\beta \ell_t(x)\right)}{Z_{t+1}} \, m_t(dx)$$

Remarks

$$m_{t+1}(dx) := \frac{\exp\left(-\beta \ell_t(x)\right)}{Z_{t+1}} m_t(dx)$$

→ Naturally promotes decisions with small cumulative loss

Remarks

$$m_{t+1}(dx) := \frac{\exp\left(-\beta \ell_t(x)\right)}{Z_{t+1}} m_t(dx)$$

→ Naturally promotes decisions with small cumulative loss

→ Performance analysis simplified by the "exp"

Remarks

$$m_{t+1}(dx) := \frac{\exp\left(-\beta \ell_t(x)\right)}{Z_{t+1}} \, m_t(dx)$$

→ Naturally promotes decisions with small cumulative loss

→ Performance analysis simplified by the "exp"

→ Theoreticaly attractive for non-euclidean generalization

Regret upper-bound

Thm (Hazan, Agarwal & Kale, 2007)

Assume every $\ell \in \mathcal{L}$ is β - expconcave
Then $\forall n \geq 1$, the EWA forecaster with
parameter β satisfies

$$R_n \lesssim \frac{p}{\beta} \ln n$$

Regret upper-bound

Thm (Hazan, Agarwal & Kale, 2007)

Assume every $\ell \in \mathcal{L}$ is β - expconcave
Then $\forall\, n \geq 1$, the EWA forecaster with parameter β satisfies

$$R_n \lesssim \frac{p}{\beta} \ln n$$

Def. ℓ β- expconcave
if $e^{-\beta \ell}$ is concave

Expconcavity vs Convexity

Rmk

$(\beta\text{- expconcave}\ ,\ \beta > 0) \Rightarrow$ Convex

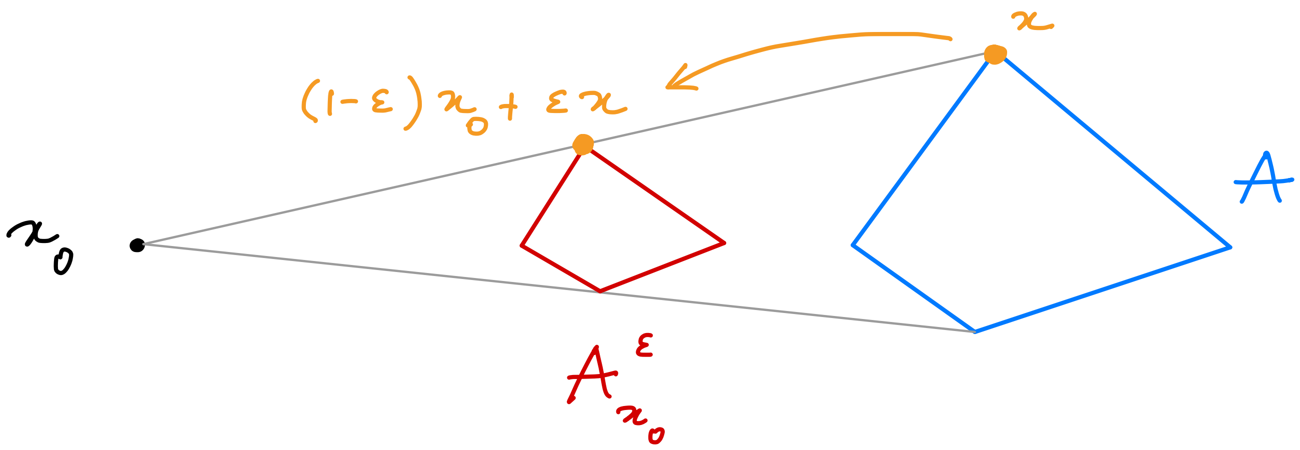
$\left(\begin{array}{c}\alpha\text{- strongly convex}\\ +\\ L\text{ - Lipschitz}\end{array}\right) \Rightarrow \frac{\alpha}{L^2}\text{ - expconcave}$

Introduction

Classical analysis relies upon :

1. Gibbs variational principle
2. Jensen's inequality
3. Properties of the Lebesgue measure

$$\left(\begin{array}{l} A \subset \mathbb{R}^p, \ x_0 \in \mathbb{R}^p, \ \varepsilon \in [0,1]. \\ A^{\varepsilon}_{x_0} := \{ (1-\varepsilon)x_0 + \varepsilon x : x \in A \} \end{array} \right) \Longrightarrow \boxed{\lambda_p(A^{\varepsilon}_{x_0}) = \varepsilon^p \lambda_p(A)}$$



Exponential Weights in Metric Spaces

Exponential weights in (M, d)

Consider → (M, d) metic space

→ Family $\mathcal{L}$ of loss functions $\ell : M \to \mathbb{R}$

Exponential weights in (M, d)

Consider

→ (M, d) metic space

→ Family $\mathcal{L}$ of loss functions $\ell : M \to \mathbb{R}$

Online Learning Problem

$\forall t \geq 1$

→ Player picks $x_t \in M$

→ Environment reveals $\ell_t \in \mathcal{L}$

→ Player incurs loss $\ell_t(x_t)$ and moves on to next round

Exponential weights in (M, d)

Consider

→ (M, d) metic space

→ Family $\mathcal{L}$ of loss functions $\ell : M \to \mathbb{R}$

Online Learning Problem

$\forall t \geqslant 1$

→ Player picks $x_t \in M$

→ Environment reveals $\ell_t \in \mathcal{L}$

→ Player incurs loss $\ell_t(x_t)$ and moves on to next round

Questions

→ Reasonable x_t ?

→ Which $(M, d) \Rightarrow$ Small R_n ?

Exponential weights in (M, d)

Consider → (M, d) metric space
→ m probability measure on M

Exponential weights in (M,d)

Consider → (M,d) metric space
→ m probability measure on M

Def $m \in \mathcal{P}_2(M)$ if $\forall x \in M$:

$$\int_M d^2(x,y)\, m(dy) < +\infty$$

Exponential weights in (M,d)

Consider → (M,d) metric space
→ m probability measure on M

Def $m \in \mathcal{P}_2(M)$ if $\forall x \in M$:

$$\int_M d^2(x,y)\, m(dy) < +\infty$$

Def $x^* \in M$ barycenter of $m \in \mathcal{P}_2(M)$ if

$$x^* \in \operatorname*{argmin}_{x \in M} \int_M d^2(x,y)\, m(dy)$$

Exponential weights in (M, d)

EWA forecaster in metric spaces

Exponential weights in (M, d)

EWA forecaster in metric spaces

Select a prior $m \in \mathcal{P}_2(M)$

Exponential weights in (M, d)

EWA forecaster in metric spaces

Select a prior $m \in \mathcal{P}_2(M)$ and select

x_t barycenter of m_t

Exponential weights in (M, d)

EWA forecaster in metric spaces

Select a prior $m \in \mathcal{P}_2(M)$ and select

x_t barycenter of m_t where

$$\begin{cases} m_1 := m \\ m_{t+1}(dx) := \dfrac{\exp\left(-\beta \ell_t(x)\right)}{Z_{t+1}} \, m_t(dx) \end{cases}$$

Exponential weights in (M,d)

EWA forecaster in metric spaces

Select a prior $m \in \mathcal{P}_2(M)$ and select

x_t barycenter of m_t where

$$\begin{cases} m_1 := m \\ m_{t+1}(dx) := \dfrac{\exp(-\beta \ell_t(x))}{Z_{t+1}} \, m_t(dx) \end{cases}$$

Question

Performance (Regret) of EWA in terms of geometric properties of (M, d, m) ?

Exponential weights in (M, d)

Geodesic spaces

Def (M, d) called geodesic if :

$\forall x_0, x_1 \in M$, $\exists$ $\gamma : [0,1] \rightarrow M$ s.t.

→ $\gamma(0) = x_0$, $\gamma(1) = x_1$

→ $\forall s,t$: $d(\gamma(s), \gamma(t)) = |s-t| \, d(x_0, x_1)$

Exponential weights in (M, d)

Geodesic homothety

Exponential weights in (M, d)

Geodesic homothety

Consider (M, d, m) and suppose

Exponential weights in (M, d)

Geodesic homothety

Consider (M, d, m) and suppose

→ (M, d) geodesic

Exponential weights in (M,d)

Geodesic homothety

Consider (M,d,m) and suppose

→ (M,d) geodesic

→ Negligeable cut-loci : $\forall x \in M$

$$m\left(\left\{ y \in M : \begin{array}{l} \exists \text{ unique } \gamma_{x,y} : [0,1] \to M \\ \text{geod. from } x \text{ to } y \end{array} \right\}\right) = 1$$

Exponential weights in (M,d)

Geodesic homothety

Consider (M,d,m) and suppose

→ (M,d) geodesic

→ Negligeable cut-loci : $\forall x \in M$

$$m\left(\left\{ y \in M : \begin{array}{l} \exists \text{ unique } \gamma_{x,y} : [0,1] \to M \\ \text{geod. from } x \text{ to } y \end{array} \right\}\right) = 1$$

Def (Geodesic homothety)

For $A \subset M$, $x \in M$ and $\varepsilon \in [0,1]$

$$A_x^\varepsilon := \left\{ \gamma_{x,y}(\varepsilon) \ : y \in M \right\}$$

Exponential weights in (M,d)

Geodesic homothety

Consider (M, d, m) and suppose

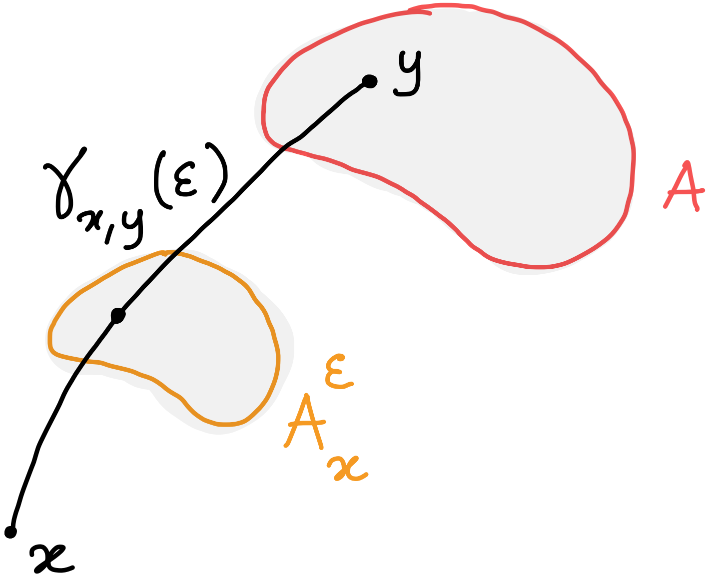
→ (M, d) geodesic
→ Negligeable cut-loci : $\forall x \in M$

$$m\left(\left\{ y \in M : \begin{array}{l} \exists \text{ unique } \gamma_{x,y} : [0,1] \to M \\ \text{geod. from } x \text{ to } y \end{array} \right\}\right) = 1$$

Def (Geodesic homothety)

For $A \subset M$, $x \in M$ and $\varepsilon \in [0,1]$

$$A_x^{\varepsilon} := \left\{ \gamma_{x,y}(\varepsilon) : y \in M \right\}$$



Exponential weights in (M,d)

Measure Contraction Property (MCP)

Def. (S.-I. Ohta, 2006)
For $K \in \mathbb{R}$ and $p > 1$, (M, d, m) satisfies the MCP (K, p) property if:
$\forall x \in M$, $\forall \varepsilon \in (0,1)$, $\forall A \subset M$
(if $K > 0$, $\forall A \subset B(x, \pi \sqrt{(p-1)/K})$)

$$m(A_x^{\varepsilon}) \geqslant \varepsilon \int_A \left(\frac{s_K\left(\frac{\varepsilon d(x,y)}{\sqrt{p-1}}\right)}{s_K\left(\frac{d(x,y)}{\sqrt{p-1}}\right)} \right)^{p-1} m(dy)$$

Exponential weights in (M,d)

Measure Contraction Property (MCP)

Def. (S.-I. Ohta, 2006)
For $K \in \mathbb{R}$ and $p > 1$, (M,d,m) satisfies the MCP(K,p) property if:
$\forall x \in M$, $\forall \varepsilon \in (0,1)$, $\forall A \subset M$
(if $K > 0$, $\forall A \subset B(x, \pi\sqrt{(p-1)/K})$)

$$m(A_x^\varepsilon) \geqslant \varepsilon \int_A \left(\frac{\underline{s_K}\left(\frac{\varepsilon d(x,y)}{\sqrt{p-1}}\right)}{\underline{s_K}\left(\frac{d(x,y)}{\sqrt{p-1}}\right)} \right)^{p-1} m(dy)$$

$$s_K(r) = \begin{cases} \dfrac{\sin(r\sqrt{K})}{\sqrt{K}} , & \text{if } K > 0 \\ r , & \text{if } K = 0 \\ \dfrac{\sinh(r\sqrt{-K})}{\sqrt{-K}} , & \text{if } K < 0 \end{cases}$$

Exponential weights in (M,d)

Measure Contraction Property (MCP)

Def. (S.-I. Ohta, 2006)
For $K \in \mathbb{R}$ and $p > 1$, (M, d, m) satisfies the MCP(K, p) property if :
$\forall x \in M$, $\forall \varepsilon \in (0,1)$, $\forall A \subset M$
(if $K > 0$, $\forall A \subset B(x, \pi \sqrt{(p-1)/K})$)

$$m(A_x^\varepsilon) \geqslant \varepsilon \int_A \left(\frac{s_K\left(\frac{\varepsilon d(x,y)}{\sqrt{p-1}}\right)}{s_K\left(\frac{d(x,y)}{\sqrt{p-1}}\right)} \right)^{p-1} m(dy)$$

Rem. Similar definitions introduced by
- Kuwae & Shioya (2001, 2003)
- Sturm (2006)

Exponential weights in (M,d)

Measure Contraction Property (MCP)

Rem. Inequality

$$m(A_x^\varepsilon) \geqslant \varepsilon \int_A \left(\frac{s_K\left(\frac{\varepsilon\, d(x,y)}{\sqrt{p-1}}\right)}{s_K\left(\frac{d(x,y)}{\sqrt{p-1}}\right)} \right)^{p-1} m(dy)$$

Becomes "$=$" when (M,d,m) is the p-dimensional Riemanian space form of constant sectional curvature K, with Riemanian distance d and volume measure m

Exponential weights in (M, d)

Main intuition

« MCP property = Synthetic Ricci Curvature Lower Bound »

Exponential weights in (M, d)

Main intuition

<< MCP property = Synthetic Ricci Curvature Lower Bound >>

Formally

Thm (Ohta, 2006)

Assume
- M complete Riem. mfd
- d Riem. distance
- m Volume measure

Then

$$\begin{array}{l} \mathrm{Ric}_M \geq K \\ \dim M \leq p \end{array} \iff \begin{array}{l} (M, d, m) \text{ satisfies} \\ \mathrm{MCP}(K, p) \end{array}$$

Exponential weights in (M, d)

Consider → (M, d, m)

→ EWA forecaster, parameter β and prior m

Exponential weights in (M, d)

Consider → (M, d, m)
→ EWA forecaster, parameter β and prior m

Thm (P., 2021)

Suppose that

- All $\ell \in \mathcal{L}$ are geodesically β- expconcave
- (M, d, m) satisfies $MCP(K, p)$,

Then $\forall n \geq 1$

$$R_n \lesssim C_K \frac{p}{\beta} \ln n$$

Exponential weights in (M, d)

Example: Log-concave priors on $\mathbb{R}^p$

Suppose $(M, d, m) = (\mathbb{R}^p, \|\cdot - \cdot\|_2, e^{-V} dx)$

Exponential weights in (M, d)

Example: Log-concave priors on $\mathbb{R}^p$

Suppose $(M, d, m) = (\mathbb{R}^p, \|\cdot - \cdot\|_2, e^{-V} dx)$

Fact If potential V is η-expconcave, then $(\mathbb{R}^p, \|\cdot - \cdot\|_2, e^{-V} dx)$ satisfies the $MCP(0, p + \frac{1}{\eta})$ property

Exponential weights in (M, d)

Example: Log-concave priors on $\mathbb{R}^p$

Suppose $(M, d, m) = (\mathbb{R}^p, \|\cdot - \cdot\|_2, e^{-V} dx)$

Fact If potential V is η-expconcave, then $(\mathbb{R}^p, \|\cdot - \cdot\|_2, e^{-V} dx)$ satisfies the $MCP(0, p + \frac{1}{\eta})$ property

EWA with prior $e^{-V} dx$ will satisfy

$$R_n \lesssim \frac{p + \frac{1}{\eta}}{\beta} \ln n$$

for β-expconcave losses

Jensen Inequality

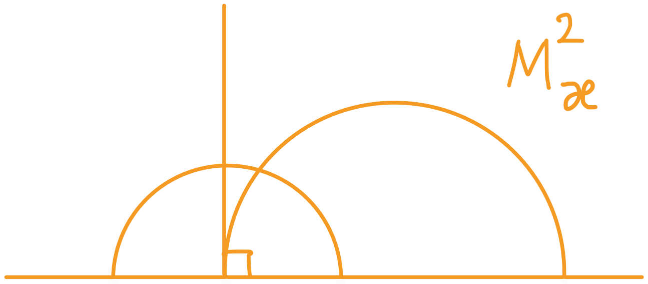
Jensen Compatibility

Alexandrov curvature bounds

Def (Model spaces) $\forall \varkappa \in \mathbb{R}$, let $(M^2_\varkappa, d_\varkappa)$ be the unique 2-dim. complete and simply connected Riemannian manifold with constant sectional curvature $\varkappa$

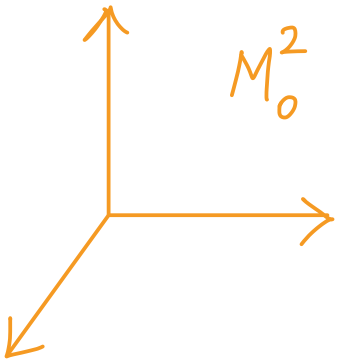
$\underline{\varkappa < 0}$

Hyperbolic plane with distance multiplied by $1/\sqrt{-\varkappa}$



$\underline{\varkappa = 0}$

Euclidean plane



$\underline{\varkappa > 0}$

Euclidean sphere of radius $1/\sqrt{\varkappa}$ with angular distance

Jensen Compatibility

Alexandrov curvature bounds

Def Let (M,d) be geodesic and $\varkappa \in \mathbb{R}$.
$\mathrm{Alex}(M) \geq \varkappa$ if $\forall\, p, x, y \in M$ and
$\forall\ \gamma : [0,1] \to M$ geodesic connecting x to y :

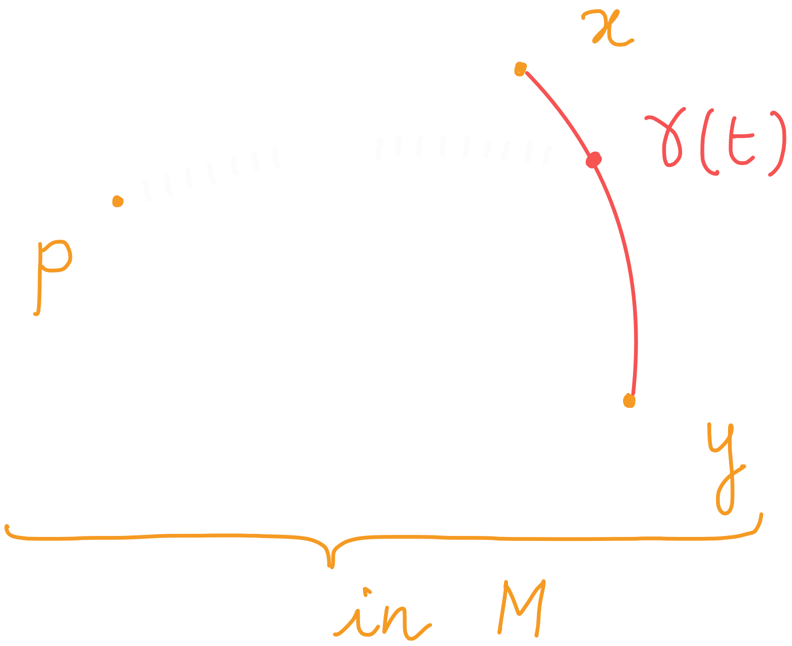
$$d(p, \gamma(t)) \geq d_{\varkappa}(\bar{p}, \bar{\gamma}(t))$$

Jensen Compatibility

Alexandrov curvature bounds

Def Let (M, d) be geodesic and $\varkappa \in \mathbb{R}$.
$\mathrm{Alex}(M) \geq \varkappa$ if $\forall\, p, x, y \in M$ and
$\forall\, \gamma : [0,1] \to M$ geodesic connecting x to y :

$$d(p, \gamma(t)) \geq d_\varkappa(\bar{p}, \bar{\gamma}(t))$$

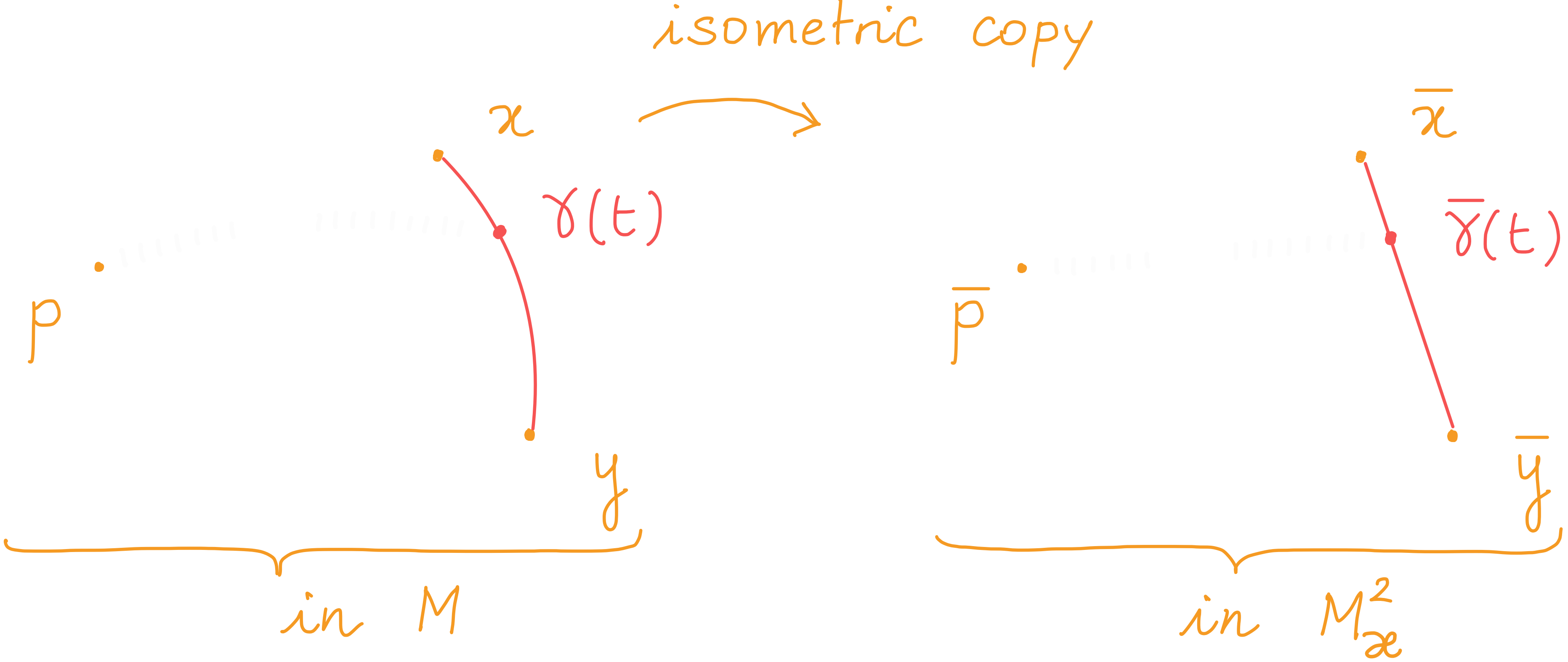


Jensen Compatibility

Alexandrov curvature bounds

Def Let (M,d) be geodesic and $\varkappa \in \mathbb{R}$.
$\mathrm{Alex}(M) \geq \varkappa$ if $\forall\, p, x, y \in M$ and
$\forall\ \gamma : [0,1] \to M$ geodesic connecting x to y :

$$d(p, \gamma(t)) \geq d_\varkappa(\bar{p}, \bar{\gamma}(t))$$



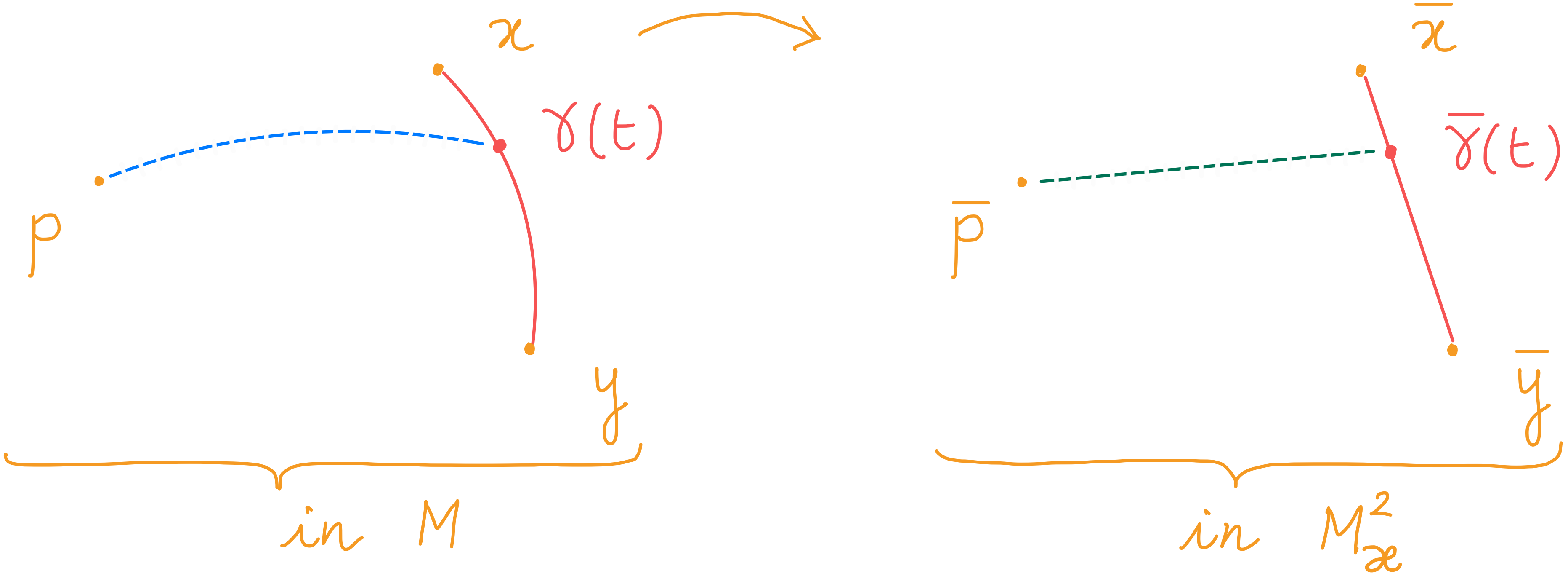
Jensen Compatibility

Alexandrov curvature bounds

Def Let (M, d) be geodesic and $\varkappa \in \mathbb{R}$.
$\mathrm{Alex}(M) \geq \varkappa$ if $\forall\, p, x, y \in M$ and
$\forall\, \gamma : [0,1] \to M$ geodesic connecting x to y :

$$\underbrace{d(p, \gamma(t))} \geq \underbrace{d_\varkappa(\bar{p}, \bar{\gamma}(t))}$$

isometric copy



Jensen Compatibility

Connection to MCP property

Thm (Kuwae & Shioya, 2001 and Ohta 2006)

Assume that
- $\mathrm{Alex}(M) \geqslant \varkappa$, $\varkappa \in \mathbb{R}$
- M compact
- M has finite Hausdorff dimension $p > 1$

Then $(M, d, \mathcal{H}_p)$ satisfies $\mathrm{MCP}((p-1)\varkappa, p)$

Connection to MCP property

Thm (Kuwae & Shioya, 2001 and Ohta 2006)

Assume that
- $\mathrm{Alex}(M) \geq \varkappa$, $\varkappa \in \mathbb{R}$
- M compact
- M has finite Hausdorff dimension $p > 1$

Then $(M, d, \mathcal{H}_p)$ satisfies $\mathrm{MCP}((p-1)\varkappa, p)$

Essentially

$$\left.\begin{array}{c} \mathrm{Alex}(M) \geq \varkappa \\ + \\ \dim(M) = p \end{array}\right\} \Rightarrow \mathrm{MCP}((p-1)\varkappa, p)$$

Jensen's inequality

Def. Let (M,d) be geodesic.
M is called Jensen compatible if
→ $\forall f : M \to \mathbb{R}$ geod. convex
→ $\forall \mu \in \mathcal{P}_2(M)$
→ $\forall x^*$ barycenter of μ
We have

$$f(x^*) \leq \int_M f(x)\,\mu(dx)$$

Jensen Compatibility

Jensen's inequality in metric spaces

- → Kendall (1990)
- → Émery & Mokobodzki (1991)
- → Sturm (2003) : Alex $(M) \leqslant 0$
- → Kuwae (2009) : Alex $(M) \leqslant \kappa$ + small radius
- → Kuwae (2014) : Convex spaces
- → Yokota (2016) : Alex $(M) \leqslant \kappa$ + small radius
- → Kim & Pass (2016) : Wasserstein space

Jensen Compatibility

$\mathrm{Alex}(M) \geqslant \varkappa \implies$ Validity of Jensen's inequality

Thm (P., 2021)

Suppose (M, d) is Polish and geodesic and such that $\mathrm{Alex}(M) \geqslant \varkappa$, $\varkappa \in \mathbb{R}$.

Then if

- f is geodesically convex
- $\nu \in \mathcal{P}_2(M)$
- x^* barycenter of ν
- f locally Lipschitz at x^*,

we have

$$f(x^*) \leqslant \int_M f \, d\nu$$

Jensen Compatibility

Observation

$\mathrm{Alex}(M) \geq \varkappa$ seems asking too much for our pbl

Question

Suppose (M, d, m) satisfies $\mathrm{MCP}(K, p)$ and consider

- $f : M \to \mathbb{R}$ geodesically convex
- $\mu \ll m$ with $\frac{d\mu}{dm} = e^{-V}$, V geod. convex
- x^* barycenter of μ

Then, do we have

$$f(x^*) \leq \int f \, d\mu$$

For more details

→ Online learning with exponential weights in metric spaces

arxiv : 2103.14389

→ Jensen's inequality in geodesic spaces with lower bounded curvature

arxiv : 2011.08597

Merci !

Exponential weights in (M,d)

Let (M,d) be geodesic

Consider the EWB forecaster with
→ Parameter $\beta > 0$
→ Prior $m \in \mathcal{P}_2(M)$

Thm (Demidova & P., 2021)

Suppose
- All $\ell \in \mathcal{L}$ are geodesically β- expconcave
- (M,d) is Jensen compatible
- (M,d,m) satisfies $MCP(K,p)$, $\boxed{K \geqslant 0}$

Then the EWB forecaster with parameter β and prior m satisfies

$$\forall n \geqslant 1, \quad R_n \leqslant \frac{2}{\beta} + \frac{p}{\beta} \ln n$$

Exponential weights in (M,d)

Thm (Demidova & P., 2021)

Suppose
- All $\ell \in \mathcal{L}$ are geodesically β- expconcave
- (M,d) is Jensen compatible
- (M,d,m) satisfies $MCP(K,p)$, $\boxed{K<0}$

and $c := \inf_{x \in M} \int \psi\left(d(x,y)\sqrt{\frac{-K}{p-1}}\right) m(dy) < +\infty$

where $\psi(r) := r\coth(r)\exp(-r\coth(r))$

Then the EWB forecaster with parameter β and prior m satisfies

$$\forall n \geqslant 1, \quad R_n \leqslant \frac{1}{\beta}\left(2 + \ln\frac{1}{c}\right) + \frac{p}{\beta}\ln n$$

Taylor expansion of geodesic triangles in Riemannian manifolds: a central tool to study the effect of curvature in geometric statistics

Xavier Pennec

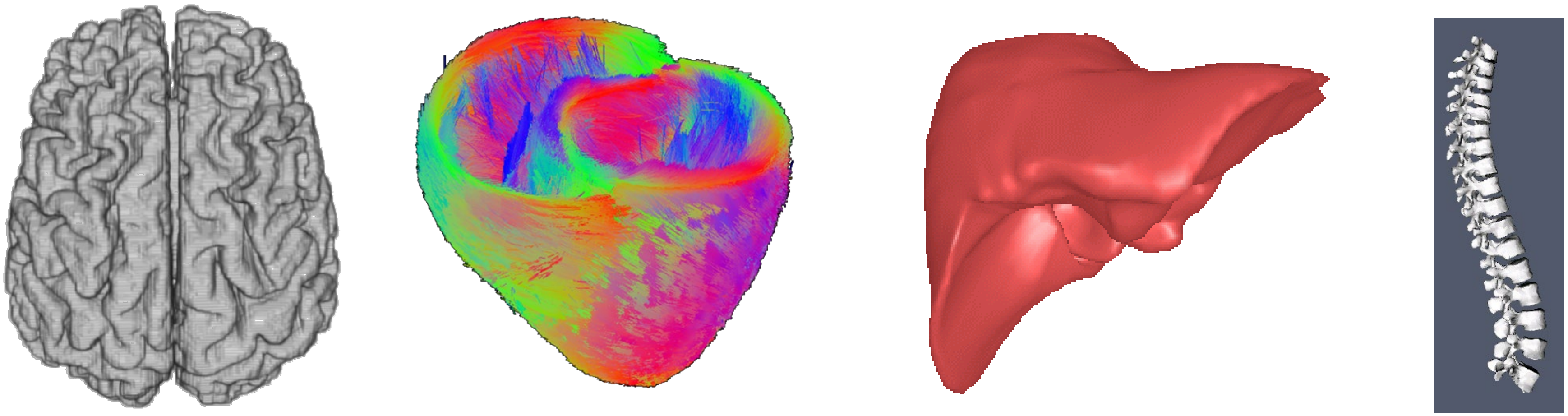
Université Côte d'Azur and Inria, France

Statistics in Metric Spaces
ENSAE, Palaiseau, 11-13/10/2023

ERC AdG 2018-2023 *G-Statistics*

Freely adapted from "Women teaching geometry", in Adelard of Bath translation of Euclid's elements, 1310.

Application context: Computational Anatomy

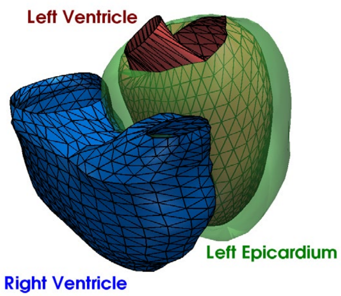


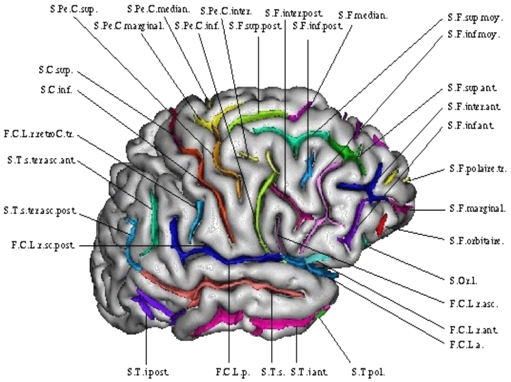
- **Methods to compute statistics of organ shapes across subjects in species, populations, diseases…**
    - Mean shape (atlas), subspace of normal vs pathologic shapes
    - Shape variability (Covariance)
    - Model development across time (growth, ageing, ages…)
- **Use for personalized medicine (diagnostic, follow-up, etc)**
    - Classical use: atlas-based segmentation

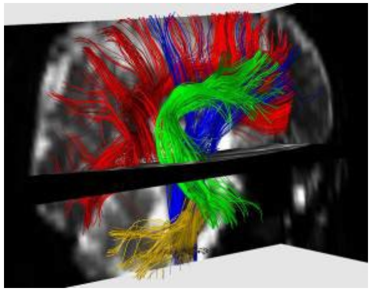
Impact of geometry on statistical learning

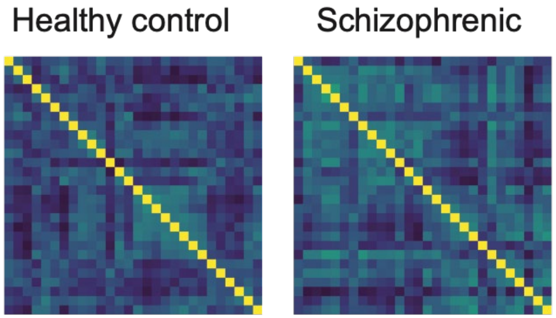
Non-linearity is everywhere in data analysis

- Images, shapes, transformations, texture, segmentations…
- Computational anatomy : Brain, heart, liver,
- Other applications: shape of molecules, Gram matrices…









Connectomics

Modeling at the population level:

- Simple statistics on non-linear Riemannian manifolds
- Frechet Mean, tPCA, PGA or GPCA

Statistical Analysis of the Scoliotic Spine

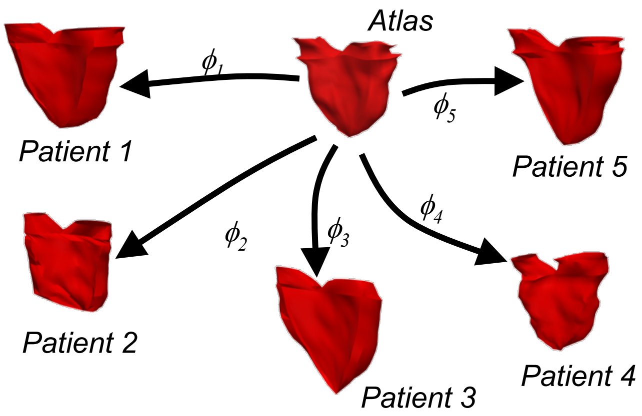
[J. Boisvert et al. ISBI'06, AMDO'06 and IEEE TMI 27(4), 2008]
AMDO'06 best paper award, Best French-Quebec joint PhD 2009

tPCA on SE(3)^{16} with left-invariant metric

4 first variation modes have clinical meaning

- Mode 1: King's class I or III
- Mode 2: King's class I, II, III
- Mode 3: King's class IV + V
- Mode 4: King's class V (+II)

Diffeomorphometry



Lift statistics to transformation groups

- [D’Arcy Thompson 1917, Grenander & Miller]
- LDDMM = right invariant kernel metric (Trouvé, Younes, Joshi, etc.)

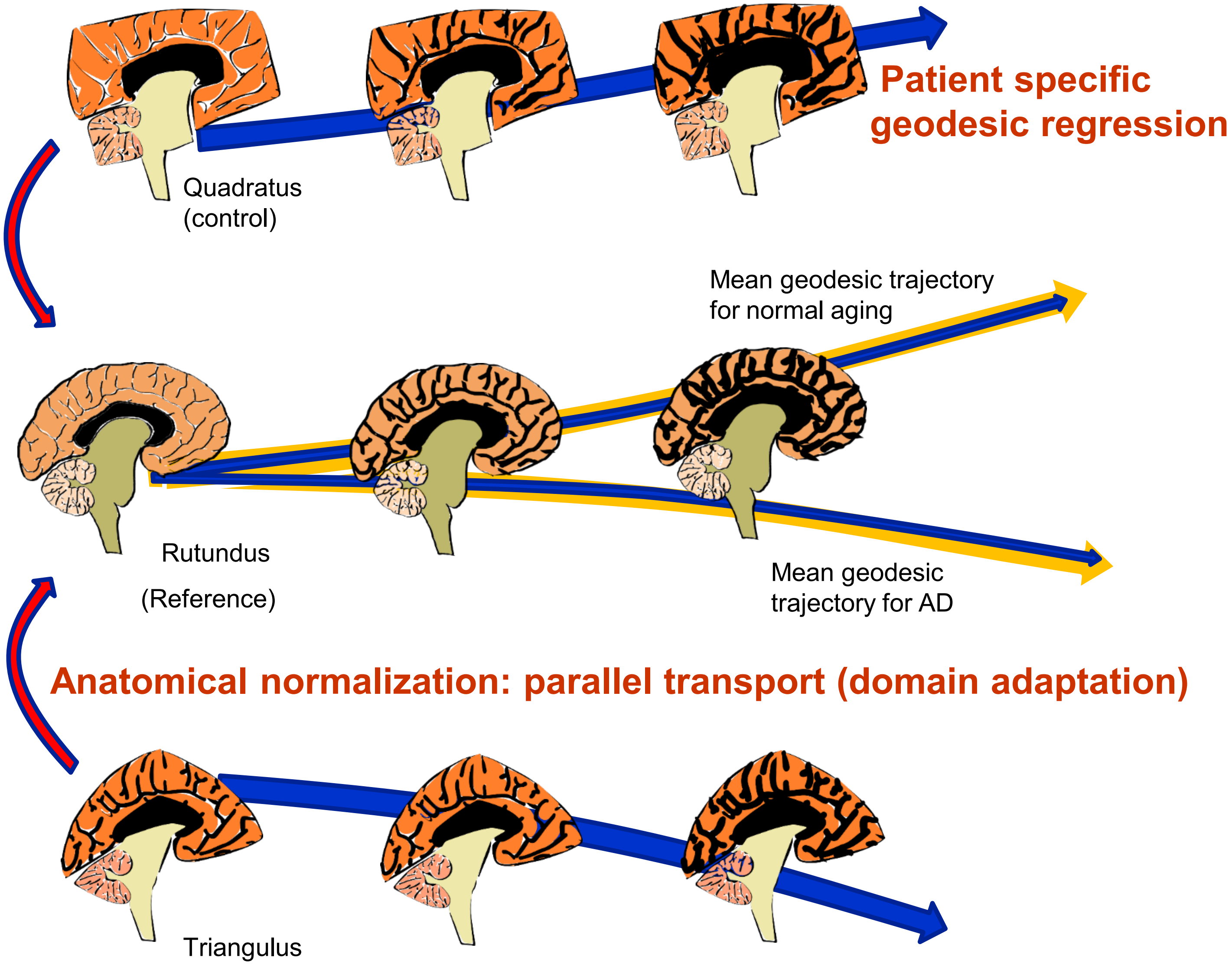
No bi-invariant metric in general for Lie groups

- Partial compatibility of Fréchet mean with the group structure:
  - Frechet mean is not right invariant nor inverse consistent
- Examples with simple 2D rigid transformations

A natural bi-invariant affine symmetric space structure

- Symmetric bi-invariant Cartan-Schouten connection (non-metric)
- Geodesics through Id = one-parameter subgroups: M(t) = exp(t.V)
  - Diffeomorphisms : flow of Stationary Velocity Fields (SVFs)
    [XP & Arsigny, 2012 ; XP & Lorenzi, IJCV 2013, Beyond Riemannian Geometry, 2019]
- Automatically “inverse-consistent”

Normal/AD modeling: Statistics on diffeomorphisms



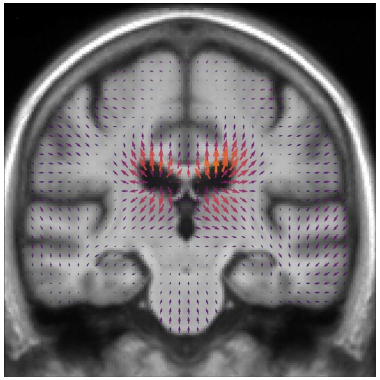
Patient specific geodesic regression

Mean geodesic trajectory for normal aging

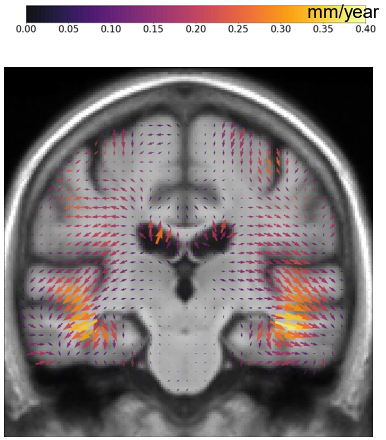
Mean geodesic trajectory for AD

Anatomical normalization: parallel transport (domain adaptation)

SVF parametrizing the deformation trajectory



Normal aging

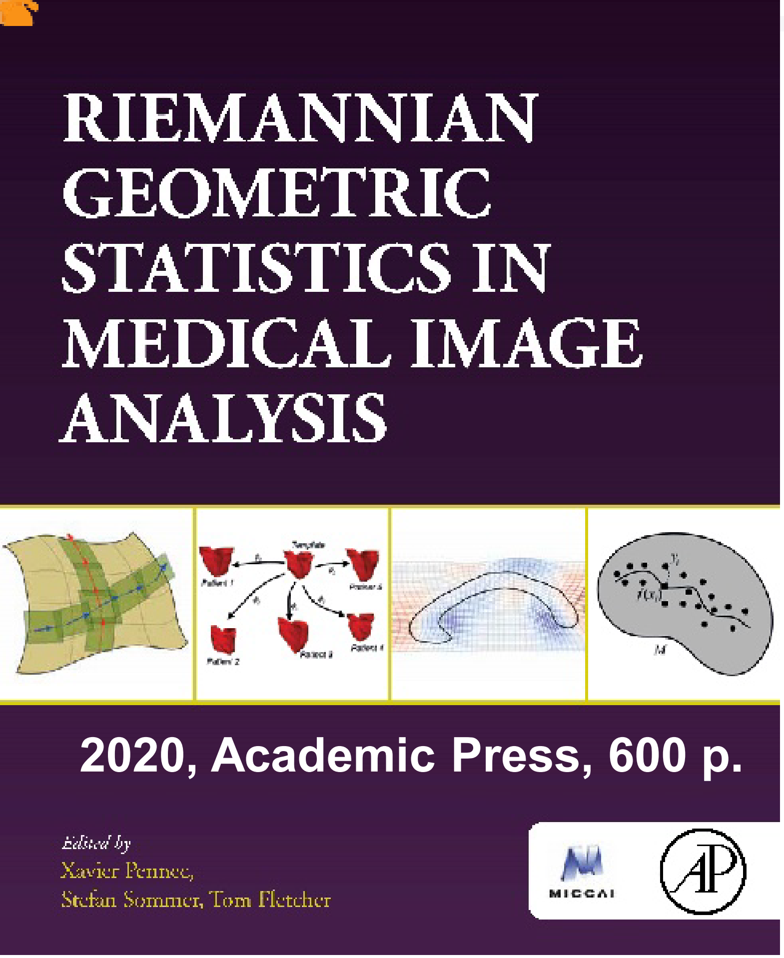


Addition specific component for AD

[Lorenzi, XP. IJCV, 2013]

[Sivera et al, Neuroimage, 2019]

Geometric statistics in 2020



RIEMANNIAN
GEOMETRIC
STATISTICS IN
MEDICAL IMAGE
ANALYSIS

2020, Academic Press, 600 p.

Edited by
Xavier Pennec,
Stefan Sommer, Tom Fletcher

Part 1: Foundations

- 1: Riemannian geometry [Sommer, Fetcher, Pennec]
- 2: Statistics on manifolds [Fletcher]
- 3: Manifold-valued image processing with SPD matrices [Pennec]
- 4: Riemannian Geometry on Shapes and Diffeomorphisms [Marsland, Sommer]
- 5: Beyond Riemannian: the affine connection setting for transformation groups [Pennec, Lorenzi]

Part 2: Statistics on Manifolds and Shape Spaces

- 6: Object Shape Representation via Skeletal Models (s-reps) and Statistical Analysis [Pizer, Maron]
- 7: Inductive Fréchet Mean Computation on S(n) and SO(n) with Applications [Chakraborty, Vemuri]
- 8: Statistics in stratified spaces [Feragen, Nye]
- 9: Bias in quotient space and its correction [Miolane, Devilier, Pennec]
- 10: Probabilistic Approaches to Statistics on Manifolds: Stochastic Processes, Transition Distributions, and Fiber Bundle Geometry [Sommer]
- 11: Elastic Shape Analysis, Square-Root Representations and Their Inverses [Zhang, Klassen, Srivastava]

Part 3: Deformations, Diffeomorphisms and their Applications

- 13: Geometric RKHS models for handling curves and surfaces in Computational Anatomy : currents, varifolds, f-shapes, normal cycles [Charlie, Charon, Glaunes, Gori, Roussillon]
- 14: A Discretize-Optimize Approach for LDDMM Registration [Polzin, Niethammer, Vialad, Modezitski]
- 15: Spatially varying metrics in the LDDMM framework [Vialard, Risser]
- 16: Low-dimensional Shape Analysis In the Space of Diffeomorphisms [Zhang, Fleche, Wells, Golland]
- 17: Diffeomorphic density matching, Bauer, Modin, Joshi]

Main questions of this talk

Statistics on manifolds based on Fréchet mean

- Uncertainty of its estimation: **confidence region?**
- Is there an **impact of curvature** on statistical tests?
- In practice: **limited number of samples (50 to 100)**
- **How large should be n** for asymptotic results?

Parallel transport algorithms

- Ladders algorithms appear to be very efficient
- Establish **numerical accuracy beyond first order**?

A common mathematical tool

- **Intrinsic Taylor expansions of geodesic triangles**

Taylor expansion of geodesic triangles in Riemannian manifolds: a central tool to study the effect of curvature in geometric statistics

Motivations

Empirical Fréchet mean concentration

[XP, Curvature effects on the empirical mean in Manifolds 2019, arXiv:1906.07418]

Numerical accuracy of parallel transport algorithms

[N. Guigui, XP, Numerical Accuracy of Ladder Schemes for Parallel Transport on Manifolds. Foundations of Computational Mathematics, 22:757-790, 2022]

Conclusions

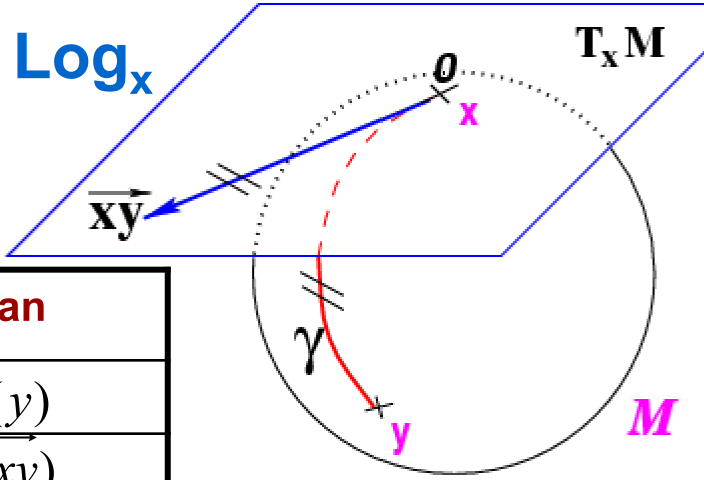
Bases of Algorithms in Riemannian Manifolds

Exponential map (Normal coordinate system):

- **Exp$_x$** = geodesic shooting parameterized by the initial tangent
- **Log$_x$** = unfolding the manifold in the tangent space along geodesics
  - Geodesics = straight lines with Euclidean distance
  - Geodesic completeness: covers M \ Cut(x)

Reformulate algorithms with Exp$_x$ and Log$_x$

Vector -> Bi-point (no more equivalence classes)



Operation	Euclidean space	Riemannian
Subtraction	$\overrightarrow{xy} = y - x$	$\overrightarrow{xy} = Log_x(y)$
Addition	$y = x + \overrightarrow{xy}$	$y = Exp_x(\overrightarrow{xy})$
Distance	$\text{dist}(x, y) = \lVert y - x \rVert$	$\text{dist}(x, y) = \lVert \overrightarrow{xy} \rVert_x$
Gradient descent	$x_{t+\varepsilon} = x_t - \varepsilon \nabla C(x_t)$	$x_{t+\varepsilon} = Exp_{x_t}(-\varepsilon \nabla C(x_t))$

Statistical tools

Maurice Fréchet
(1878-1973)

Fréchet mean set

- Integral only valid in Hilbert/Wiener spaces [Fréchet 44]
- $MSD(x) = Tr_g\left(\mathfrak{M}_2(x)\right) = \int_M dist^2(x,z)\,P(dz)$
- **Fréchet mean** [1948] = global minima of Mean Sq. Dist.
- **Exponential barycenters** [Emery & Mokobodzki 1991]

  $\mathfrak{M}_1(\bar{x}) = \int_M \log_{\bar{x}}(z) P(dz) = 0$ [critical points if P(C) =0]

Moments of a random variable: tensor fields

- $\mathfrak{M}_1(x) = \int_M \log_x(z) P(dz)$ Tangent mean: (0,1) tensor field
- $\mathfrak{M}_2(x) = \int_M \log_x(z) \otimes \log_x(z) P(dz)$ Second moment: (0,2) tensor field
  - Tangent covariance field: $Cov = \mathfrak{M}_2 - \mathfrak{M}_1 \otimes \mathfrak{M}_1$
- $\mathfrak{M}_k(x) = \int_M \log_x(z) \otimes \log_x(z) \otimes \cdots \otimes \log_x(z) P(dz)$ k-moment: (0,k) tensor field

Asymptotic behavior of the mean

Uniqueness of p-means with convex support

[Karcher 77 / Buser & Karcher 1981 / Kendall 90 / Afsari 10 / Le 11]

- Non-positively curved metric spaces (Aleksandrov): OK [Gromov, Sturm]
- Positive curvature: [Karcher 77 & Kendall 89] **concentration conditions (KKC)**:
  Support in a regular geodesic ball of radius $r < r^* = \frac{1}{2}\min(inj(M), \pi/\sqrt{\kappa})$

Bhattacharya-Patrangenaru CLT [BP 2005, B&B 2008]

- Under suitable concentration conditions [KKC], for IID n-samples:
  - $\bar{x}_n \to \bar{x}$ *(consistency of empirical mean)*
  - $\sqrt{n}\, log_{\bar{x}}(\bar{x}_n) \to N(0, \boldsymbol{\bar{H}^{-1}\, \Sigma\, \bar{H}^{-1}})$ *if* $\bar{H} = \int_M Hess_{\bar{x}}\left(\frac{1}{2}d^2(y,\bar{x})\right) \mu(dy)$ *invertible*
- Problems for larger supports [Huckemann & Eltzner, H. Le, D. Tran]

Behavior in high concentration conditions?

- **No expression for Hessian: interpretation of covariance modulation?**
- **What happens for a small sample size (non-asymptotic behavior)?**
- **Can we extend results to affine connection spaces?**

Curvature effects in Geometric statistics : empirical Fréchet mean and parallel transport accuracy

Motivations for statistics on manifolds

Empirical Fréchet mean concentration

[XP, Curvature effects on the empirical mean in Manifolds 2019, arXiv:1906.07418]

- Asymptotic BP-CLT
- Small sample & high concentration expansion

Numerical accuracy of parallel transport algorithms

[N. Guigui, XP, Numerical Accuracy of Ladder Schemes for Parallel Transport on Manifolds. Foundations of Computational Mathematics, online 04-2021]

Conclusions

Principle and difficulty

The empirical mean $\bar{x}_n$ of an IID n-sample with population mean $\bar{x}$ is a random variable on M

- Locate $\bar{x}_n$ in a normal coordinate system at x for a given empirical law
- Compute the moments of the empirical mean $\bar{x}_n$at $\bar{x}$:
  - Expectation at the population mean: $\mathbf{Bias}(\bar{x}_n) = \mathbb{E}\left(log_{\bar{x}}(\bar{x}_n)\right)$
  - Covariance matrix $\mathbf{Cov}(\bar{x}_n) = \mathbb{E}\left(log_{\bar{x}}(\bar{x}_n) \otimes log_{\bar{x}}(\bar{x}_n)\right)$
  - Compare with asymptotic BP-CLT for large n

Empirical and population means are exponential barycenters

- n-sample $\mathrm{X_n} = \frac{1}{n}\sum_i \delta_{x_i}$ → tangent mean vector field is $\mathfrak{M}_1(x) = \frac{1}{n}\sum_i \log_x(x_i)$
- Locate the zero $\bar{x}_n$ → **Taylor expansion of $\boldsymbol{log_{x_v}(y)}$ for $\boldsymbol{x_v = exp_x(v)}$?**

Riemannian distance derivatives

How does the (squared) distance (Synge's world function) vary with endpoints?

- First order derivatives is easy

$$D_v\big(\mathrm{dist}^2(x_v, y)\big) = -2\log_{x_v}(y) \text{ with } \mathrm{x_v} = \exp_x(v)$$

- Higher order derivatives begin to be quite involved:
  Taylor expansion in normal coordinates (Grey 1973, Brewin 1998, 2009)

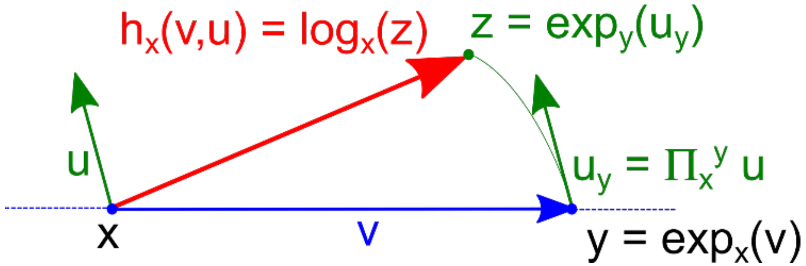
$$\begin{aligned}
D(v) =&\ \mathrm{dist}^2(\exp_x(v), y) = \|y\|_x^2 + D_{,a}v^a + D_{,ab}v^a v^b + D_{,abc}v^a v^b v^c + D_{,abcd}v^a v^b v^c v^d + O(\epsilon^5),\\
D_{,a} =&\ -2y_a\\
D_{,ab} =&\ g_{ab} - \frac{1}{3}y^c y^d R_{acbd} - \frac{1}{12}y^c y^d y^e \nabla_d R_{aebc} - \frac{1}{180}y^c y^d y^e y^f (44R^g_{eaf}R_{gcbd} - 3\nabla_{ef}R_{acbd})\\
D_{,abc} =&\ -\frac{1}{12}y^d y^e \nabla_c R_{aebd} + \frac{1}{60}y^d y^e y^f (\nabla_{da}R_{bfce} - 2\nabla_{ad}R_{bfce} + 32R^g_{dbe}R_{gacf})\\
&- \frac{1}{108}y^d y^e y^f y^g (8R^h_{eaf}\nabla_g R_{hbcd} + 9R^h_{eaf}\nabla_h R_{bgcd} + 20R^h_{eaf}\nabla_b R_{hgcd} - 6R^h_{abe}\nabla_f R_{hgcd})\\
D_{,abcd} =&\ +\frac{1}{180}y^e y^f (8R^g_{cde}R_{gabf} - 9\nabla_{cd}R_{aebf} - 8R^g_{dae}R_{gcbf} + 9\nabla_{dc}R_{afbe} - 44R^g_{daf}R_{gcbe} - 3\nabla_{db}R_{ceaf})\\
&+ \frac{1}{45}y^e y^f y^g (4R^h_{ecf}\nabla_a R_{hbdg} + 4R^h_{cae}\nabla_b R_{hfdg} + 4R^h_{cae}\nabla_f R_{hbdg} - 3\nabla_{dae}R_{bfcg})\\
&+ \frac{1}{108}y^e y^f y^g (8R^h_{eaf}\nabla_d R_{hbcg} + 8R^h_{daf}\nabla_g R_{hbce} + 9R^h_{eaf}\nabla_h R_{bdcg} + 9R^h_{daf}\nabla_h R_{bgce}\\
&\qquad + 20R^h_{eaf}\nabla_b R_{hdcg} + 20R^h_{daf}\nabla_b R_{hgce} - 6R^h_{abe}\nabla_f R_{hdcg} - 6R^h_{abe}\nabla_d R_{hgcf})
\end{aligned}$$

- Problem: $\log_{x_v}(y) \in T_{x_v}M$ and not to $\mathrm{T_x}M$: many terms due to $\mathrm{Dexp}_x(v)$

Taylor expansion of geodesic triangles

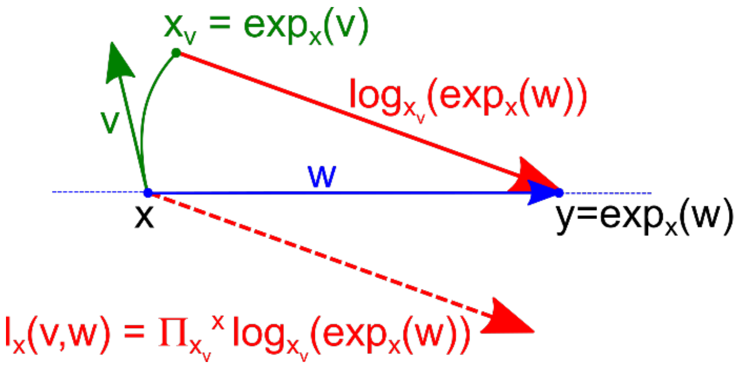
Key idea: use parallel transport rather that normal chart to relate $T_x M$ to $T_{x_v} M$

Gavrilov's double exponential is a tensorial series (2006):



$$h_x(v,u) = \log_x(\exp_{x_v}(\Pi_x^{x_v}\, u))$$
$$= v + u + \frac{1}{6}R(u,v)v + \frac{1}{3}R(u,v)u$$
$$+ \frac{1}{24}\nabla_v R(u,v)(2v + 5u) + \frac{1}{24}\nabla_u R(u,v)(v + 2u) + O(5)$$

Neighboring log expansion [XP arXiv:1906.07418, 2019]



$$l_x(v,w) = \Pi_{x_v}^x \log_{x_v}(\exp_x(w))$$
$$= w - v + \frac{1}{6}R(w,v)(v - 2w) + \frac{1}{24}\nabla_v R(w,v)(2v - 3w)$$
$$+ \frac{1}{24}\nabla_w R(w,v)(v - 2w) + O(5)$$

Torsion free affine manifolds

Taylor expansion of recentered mean map

$\mathbf{x_v} = \exp_{\boldsymbol{x}}(\boldsymbol{v})$ is an exponential barycenter if $\mathfrak{M}_1(\boldsymbol{x_v}) = \mathbf{0}$

- $\mathfrak{N}_x(v) = \Pi_{x_v}^{x} \mathfrak{M}_1(x_v) = \int_M \boldsymbol{\Pi_{x_v}^{x}} \boldsymbol{\log_{x_v}(y)}\, \mu(dy)$ has a zero at $\mathrm{v} = \log_x(\bar{x})$
- $\boldsymbol{\mathfrak{M}_1}$ **is a tensor field,** $\boldsymbol{\mathfrak{N}_x}$ **is an analytic endomorphism of** $\boldsymbol{T_x M}$

Taylor expansion with neighboring log:

$$\mathfrak{N}_x(v) = \mathfrak{M}_1 - v + \frac{1}{6} R(\mathfrak{M}_1, v)v - \frac{1}{3} R(*, v) * : \mathfrak{M}_2^{**} + \frac{1}{12} (\nabla_v R)(\mathfrak{M}_1, v)v$$
$$+ \frac{1}{24} (\nabla_* R)(*, v)v : \mathfrak{M}_2^{**} - \frac{1}{8} (\nabla_v R)(*, v) *: \mathfrak{M}_2^{**} - \frac{1}{12} (\nabla_* R)(*, v) *: \mathfrak{M}_3^{***} + O(\varepsilon^5)$$

Solve for the value $\mathrm{v} = \log_x(\bar{x})$ zeroing-out the polynomial

$$\log_x(\bar{x}) = \mathfrak{M}_1 - \frac{1}{3} R(*, \mathfrak{M}_1) * : \mathfrak{M}_2 + \frac{1}{24} (\nabla_* R)(*, \mathfrak{M}_1)\mathfrak{M}_1 : \mathfrak{M}_2^{**}$$
$$- \frac{1}{8} \left(\nabla_{\mathfrak{M}_1} R\right)(*, \mathfrak{M}_1) * : \mathfrak{M}_2^{**} - \frac{1}{12} (\nabla_* R)(*, \mathfrak{M}_1) * : \mathfrak{M}_3^{***} + O(\varepsilon^5)$$

Expectation for a random n-sample

For one empirical n-sample $\mathrm{X_n} = \frac{1}{n}\sum_i \delta_{x_i}$ with moments $\mathfrak{X}_k^n$

- $\log_x(\bar{x}_n) = \mathfrak{X}_1^n - \frac{1}{3}R(*, \mathfrak{X}_1^n) * : \mathfrak{X}_2^n + \frac{1}{24}(\nabla_* R)(*, \mathfrak{X}_1^n)\mathfrak{X}_1^n : \mathfrak{X}_2^{n\,**}$
  $- \frac{1}{8}\left(\nabla_{\mathfrak{X}_1^n} R\right)(*, \mathfrak{X}_1^n) * : \mathfrak{X}_2^{n\,**} - \frac{1}{12}(\nabla_* R)(*, \mathfrak{X}_1^n) * : \mathfrak{X}_3^{n\,***} + O(\varepsilon^5)$

Take expectation for a random IID n-sample

- $\mathbb{E}[\mathfrak{X}_k^n(x)] = \mathfrak{M}_k(x)$
- $\mathbb{E}\left[\mathfrak{X}_p^n \otimes \mathfrak{X}_q^n\right] = \frac{n-1}{n}\mathfrak{M}_{p+q} \otimes \mathfrak{M}_{p+q} + \frac{1}{n}\mathfrak{M}_{p+q}$
- Etc…

Moments of the empirical mean at the population mean:

- $\mathbf{Bias}(\bar{x}_n) = \mathbb{E}[\log_{\bar{x}}(\bar{x}_n)] = \frac{n-1}{6n^2}\,(\nabla_* R)(*,\circ) \circ : \mathfrak{M}_2^{**} : \mathfrak{M}_2^{\circ\circ} + O(\varepsilon^5)$
- $\mathbf{Cov}(\bar{x}_n) = \mathbb{E}[\log_{\bar{x}}(\bar{x}_n) \otimes \log_{\bar{x}}(\bar{x}_n)]$
  $= \frac{1}{n}\mathfrak{M}_2 - \frac{n-1}{3n^2}\mathfrak{M}_2^{**} : (\circ \otimes R(*,\circ) * + R(*,\circ) * \otimes \circ) : \mathfrak{M}_2^{\circ\circ} + O(\varepsilon^5)$

Asymptotic behavior of empirical Fréchet mean

Moments of the Fréchet mean of a n-sample

- **Surprising Bias** in 1/n on the empirical Fréchet mean (**gradient of curvature**)

$$\mathbf{Bias}(\bar{x}_n) = \mathbb{E}\big(log_{\bar{x}}(\bar{x}_n)\big) = \frac{1}{6n}\,(\mathfrak{M}_2 : \nabla R : \mathfrak{M}_2) + \; O(\epsilon^5, 1/n^2)$$

- **Concentration rate:** term in 1/n modulated by the **curvature**:

$$\mathbf{Cov}(\bar{x}_n) = \mathbb{E}\big(log_{\bar{x}}(\bar{x}_n) \otimes log_{\bar{x}}(\bar{x}_n)\big) = \frac{1}{n}\mathfrak{M}_2 \; + \frac{1}{3n}\mathfrak{M}_2 : R : \mathfrak{M}_2 + \; O(\epsilon^5, 1/n^2)$$

  - Negative curvature: faster CV than Euclidean
  - Positive curvature: slower CV than Euclidean

Central-limit theorem in manifolds [Bhattacharya & Bhattacharya 2008; Kendall & Le 2011]

- Under Kendall-Karcher concentration conditions:

$$\sqrt{n}\, log_{\bar{x}}(\bar{x}_n) \xrightarrow{D} N(0, H^{-1}\, \Sigma\, H^{-1}) \quad \textit{if} \;\; H = Hess\big(\, MSD(X, \bar{x}_n)\big) \; \textit{invertible}$$

- **Hessian of mean sq. dist:** $\frac{1}{2}\overline{H} = Id + \frac{1}{3}R : \mathfrak{M}_2 + \frac{1}{12}\nabla \mathrm{R} : \mathfrak{M}_3 + O(\epsilon^4, 1/n^2)$
- Same expansion for large n: *modulation of the CV rate by curvature (but our non asymptotic expansion is valid for small data as well)*

[XP, 2018, ARXIV : 1811.01370]

Isotropic distribution in constant curvature spaces

- Symmetric spaces: no bias at order 5
- Modulation of variance w.r.t. Euclidean: $Var(\bar{x}_n) = \alpha \frac{\sigma^2}{n}$

$$\lim_{\kappa\theta^2 = \pi^2/2^2} \alpha = +\infty$$

No CV for uniform distrib on equator

High concentration expansion

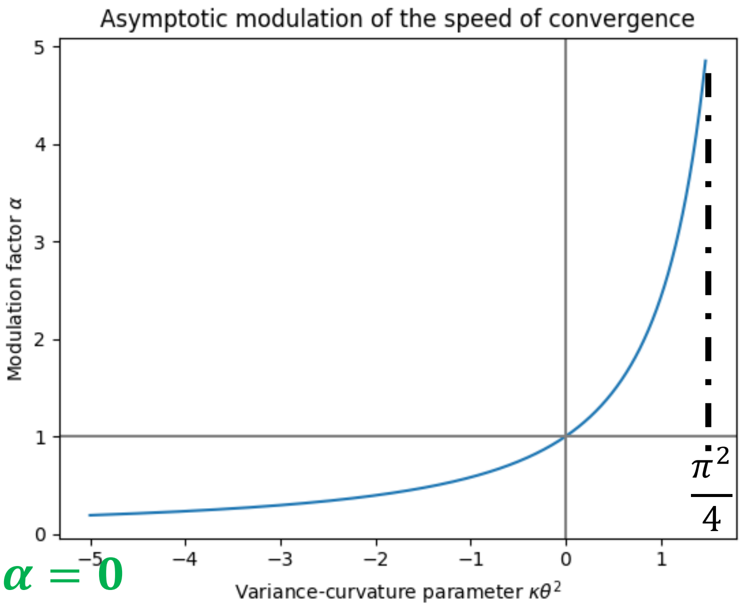
- $\alpha = 1 + \frac{2}{3}\left(1 - \frac{1}{d}\right)\left(1 - \frac{1}{n}\right)\kappa\sigma^2 + O(\epsilon^5)$

Asymptotic BP-CLT expansion

- $\alpha = \left(\frac{1}{d} + \left(1 - \frac{1}{d}\right)\bar{h}\right)^{-2} + O(n^{-2})$

Archetypal modulation factor

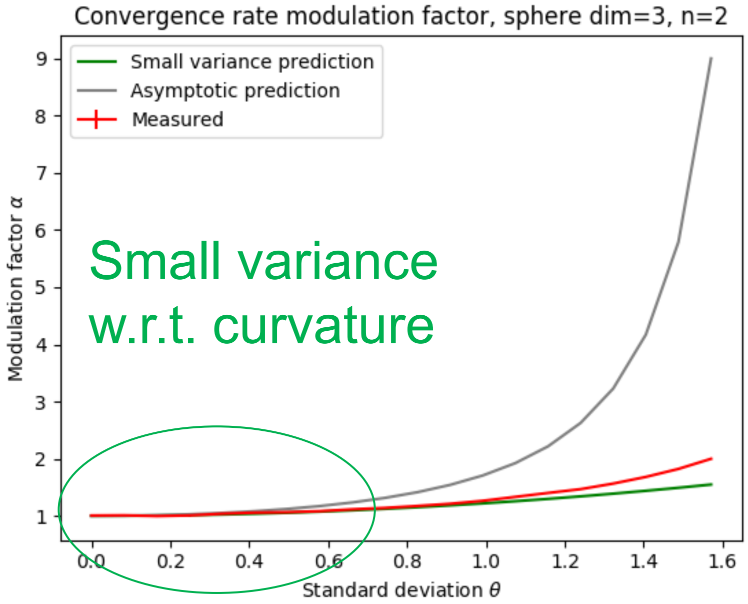
- Uniform distrib on $S(\bar{x}, \theta) \subset M$, large n, large d: $\alpha = \frac{\tan^2(\sqrt{\kappa\theta^2})}{\kappa\theta^2}$



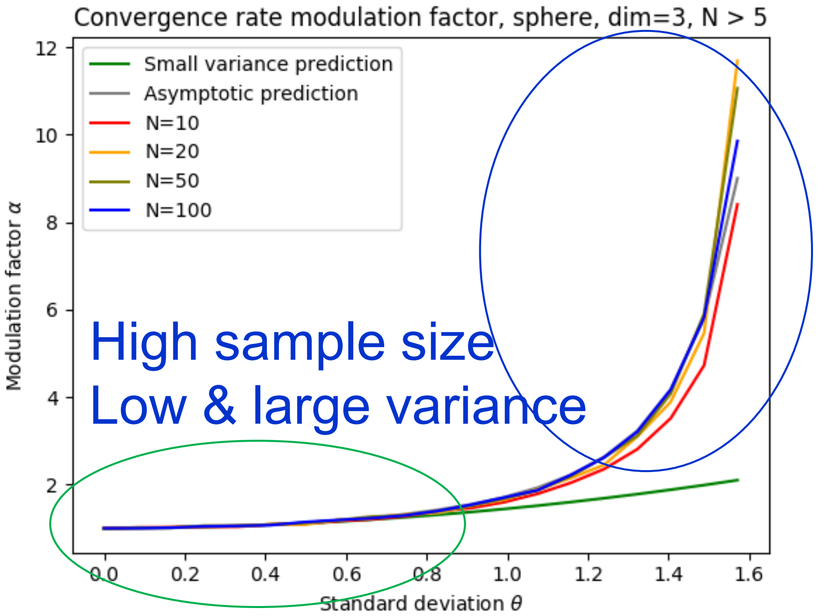
$$\lim_{\kappa\theta^2 = -\infty} \alpha = 0$$

Immediate convergence: sticky mean

Positive curvature


Convergence rate modulation factor, sphere dim=3, n=2

Small variance
w.r.t. curvature

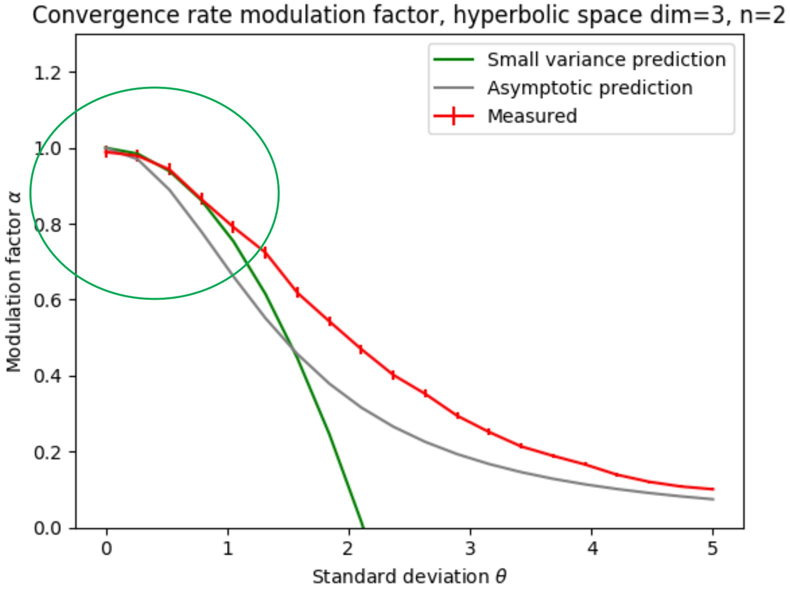

Convergence rate modulation factor, sphere, dim=3, N > 5

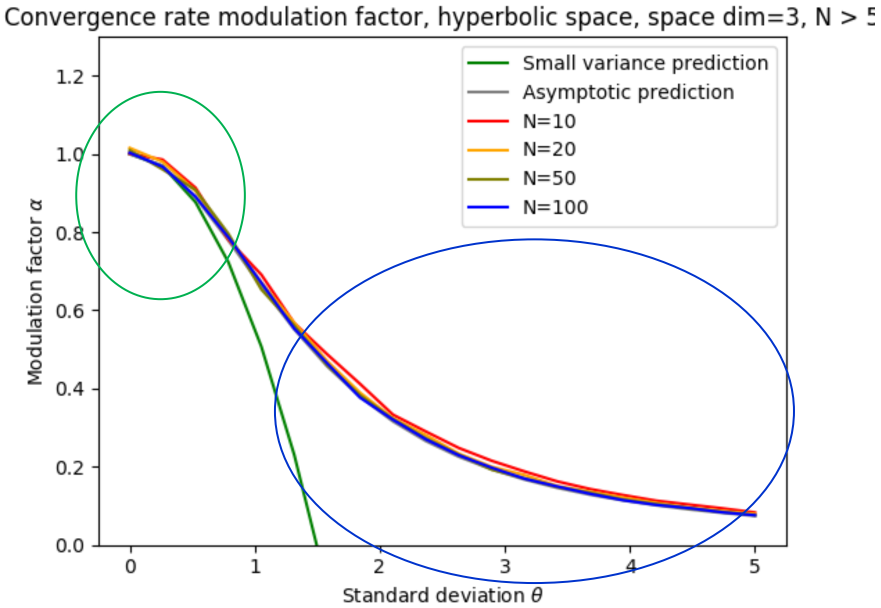
High sample size
Low & large variance

Accurate expansion even with small sample

Accurate asymptotic expansion

Negative curvature

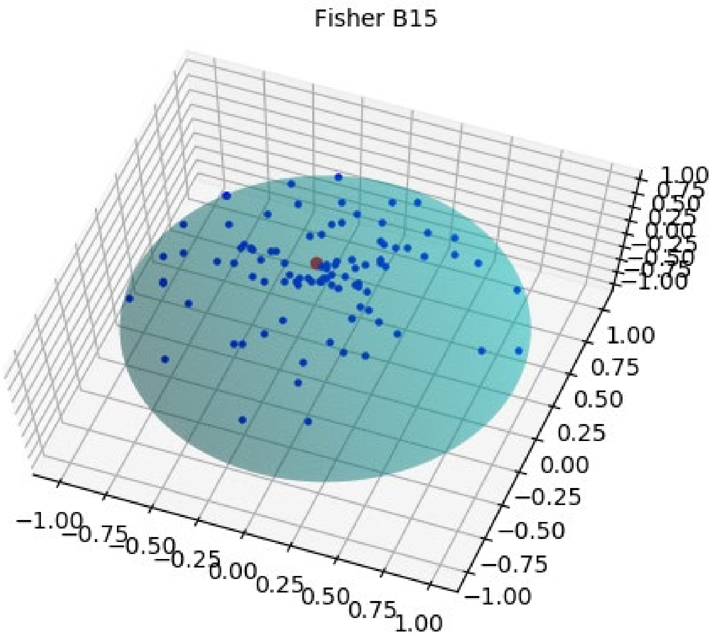

Convergence rate modulation factor, hyperbolic space dim=3, n=2

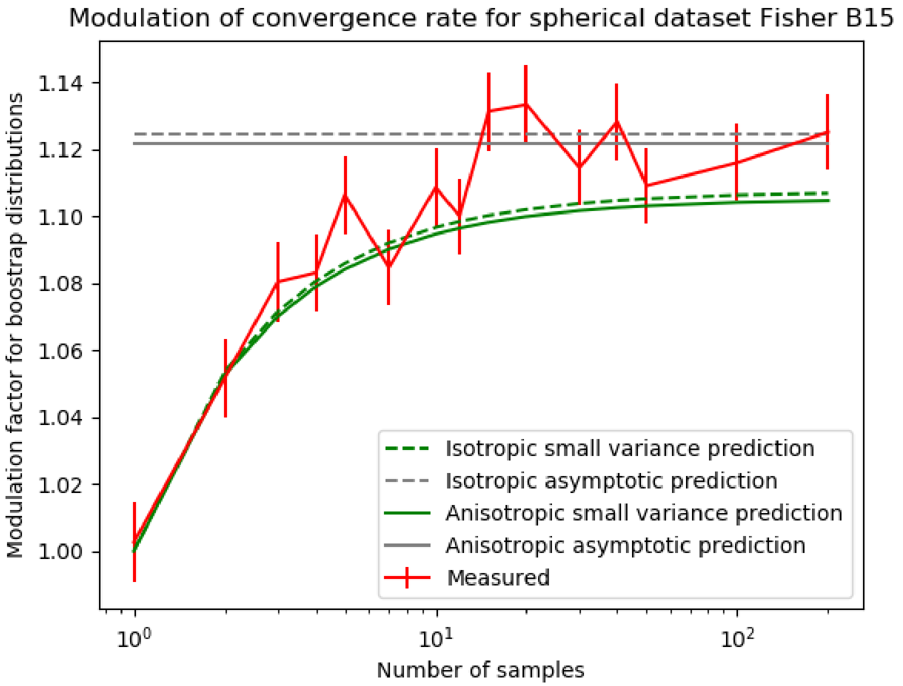

Convergence rate modulation factor, hyperbolic space, space dim=3, N >

Boostrap on real spherical data from [Fisher, Lewis, Embleton 1987]

B15: high isotropic dispersion (stddev 32º, bbox: 76ºx63º)

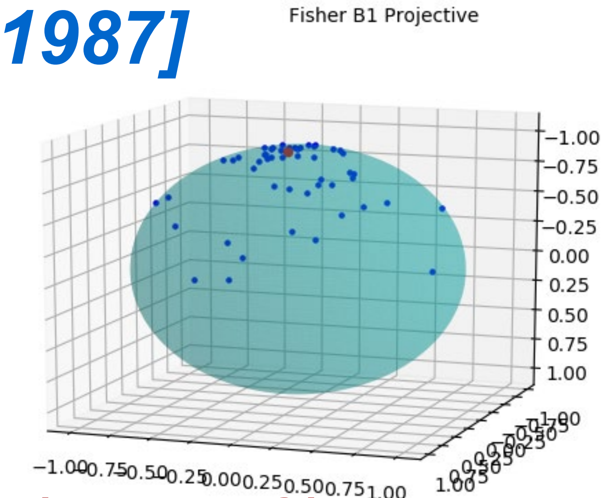
- 94 orientations of dendritic fields in cat’s retinas [Keilson et al 1983]
- High dispersion, KKC on the sphere





- Visible modulation (isotropic formulas are good)
- Small sample expansion behavior is well predicted

Boostrap on real projective data from [Fisher, Lewis, Embleton 1987]


Fisher B1 Projective

Fisher B1: high dispersion

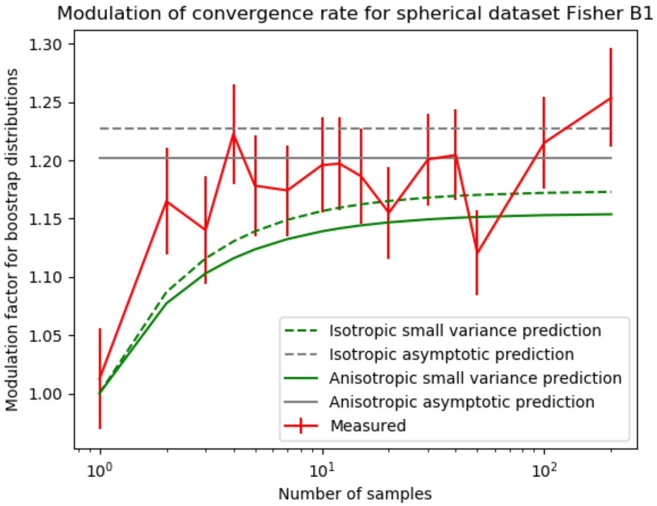
- 50 pole positions from Paleomagnetic study of new Caledonian laterites (Falvey & Mustgrave)

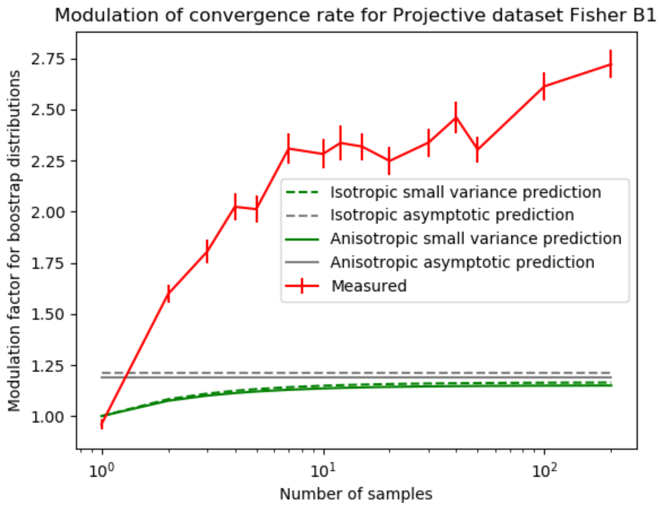
Spherical (not KKC)

- Stddev 41°, bbox: 98° x 67°
- Small var and asymptotic OK

Projective (not KKC)

- Stddev 40°, bbox: 86° x 76°
- Prediction fails: **smeary mean?**


Modulation of convergence rate for spherical dataset Fisher B1


Modulation of convergence rate for Projective dataset Fisher B1

Taylor expansion of geodesic triangles in Riemannian manifolds: a central tool to study the effect of curvature in geometric statistics

Motivations

Empirical Fréchet mean concentration

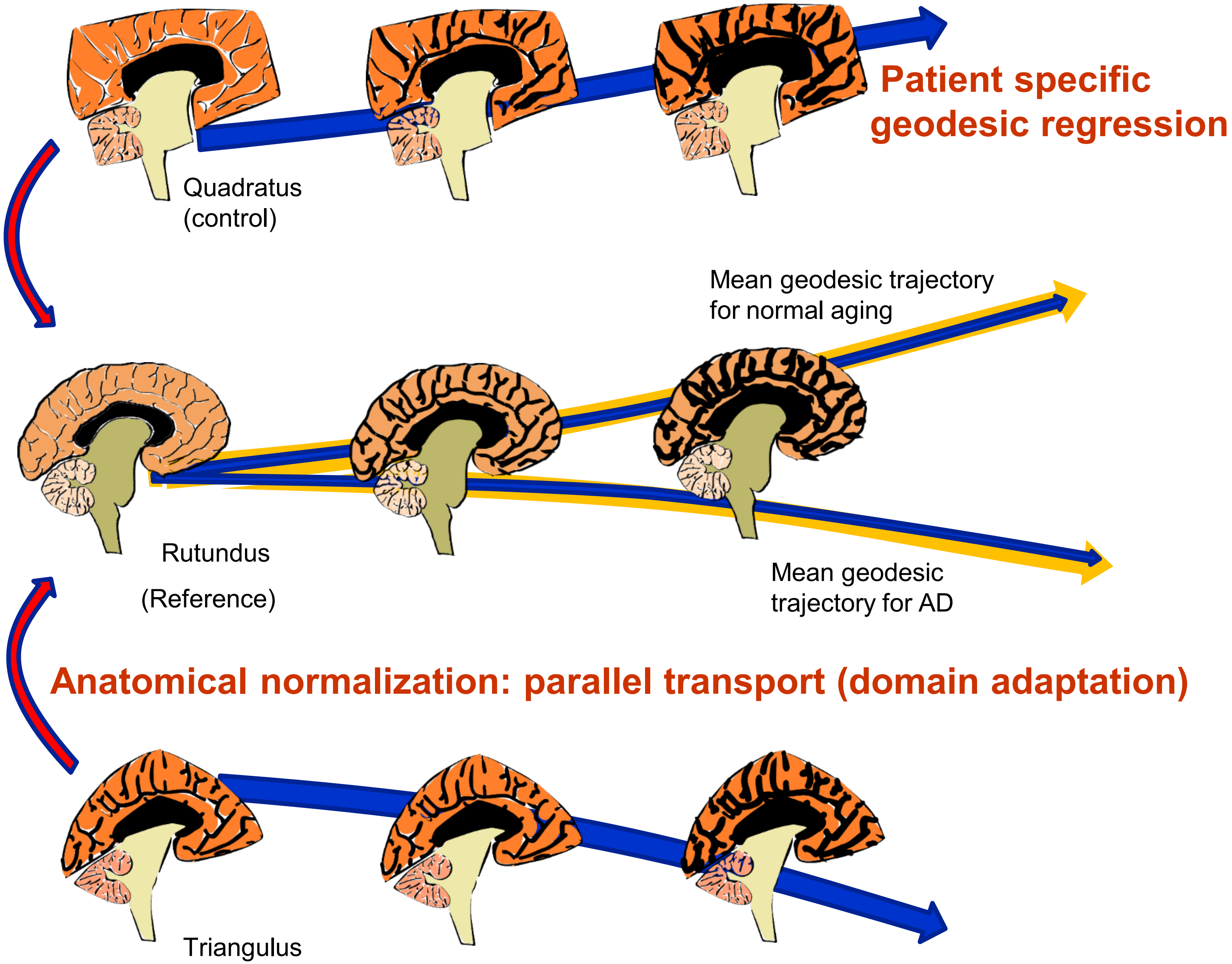
[XP, Curvature effects on the empirical mean in Manifolds 2019, arXiv:1906.07418]

Numerical accuracy of parallel transport algorithms

[N. Guigui, XP, Numerical Accuracy of Ladder Schemes for Parallel Transport on Manifolds. Foundations of Computational Mathematics, 22:757-790, 2022]

Conclusions

Normal/AD modeling: Statistics on diffeomorphisms



Patient specific geodesic regression

Quadratus (control)

Mean geodesic trajectory for normal aging

Rutundus

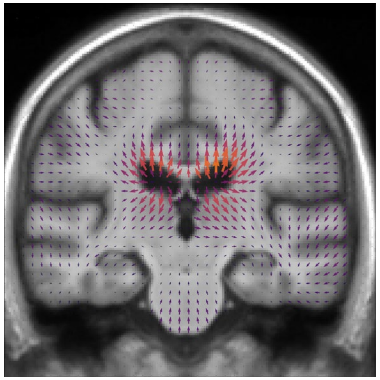
(Reference)

Mean geodesic trajectory for AD

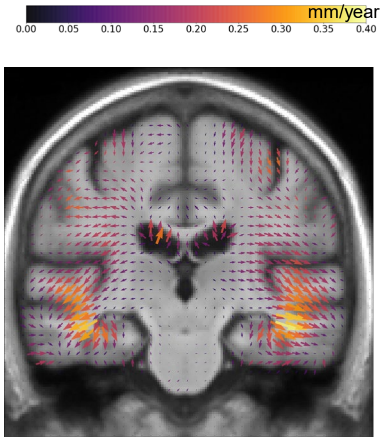
Anatomical normalization: parallel transport (domain adaptation)

Triangulus (Alzheimer)

SVF parametrizing the deformation trajectory



Normal aging



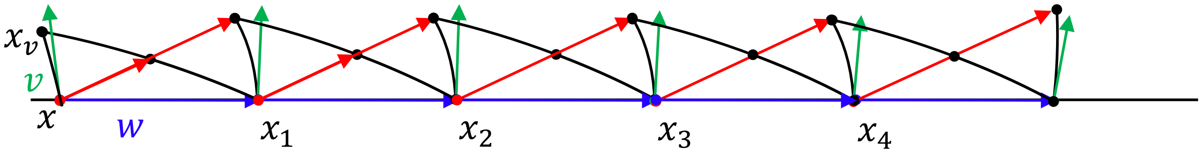
Addition specific component for AD

[Lorenzi, XP. IJCV, 2013]

[Sivera et al, Neuroimage, 2019]

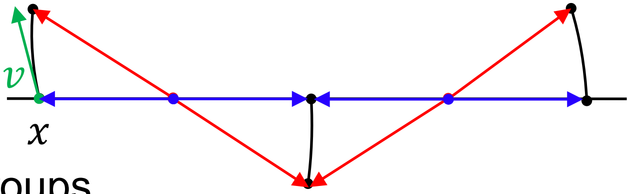
Discrete approximations of Parallel transport

Schild's Ladder [Schild's lectures at Princeton 60ies, Elhers et al 1972]



- Build geodesic parallelogramme
- Iterate along the curve
- One step is a 1st order approximation **[Kheyfets et al 2000]**

Pole ladder: [Lorenzi, XP, JMIV 50 (1-2), 2013]



- Simpler method with piecewise geodesics
  - Closed form expression for Cartan connection on Lie groups
- One step is of order 4 in general affine manifolds **[XP, Arxiv 1805.11436, 2018]**

$$\text{pole(u)} = \Pi(u) + \frac{1}{12} \nabla_v R(u,v)(5u - 2v) + \frac{1}{12} \nabla_u R(u,v)(v - 2u) + O(5)$$

  - **Exact in symmetric spaces (transvection)!**

→ No approximation formula beyond 1st order for SL
→ No results for the iterated SL and PL schemes
→ No results for approximate geodesics

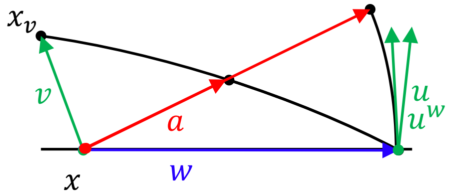
Convergence of Schild’s Ladder

Gavrilov’s Taylor expansion of one Schild’s ladder step

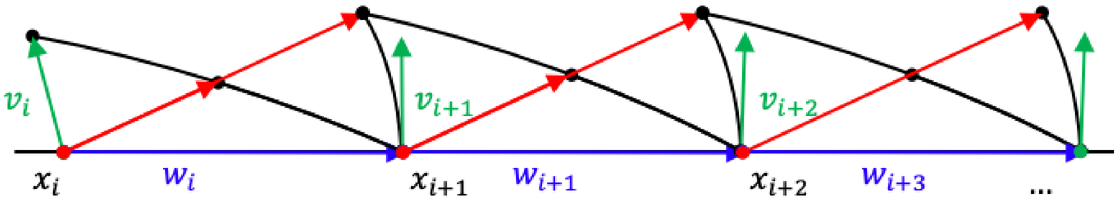
- A new Taylor series for mid-point rule

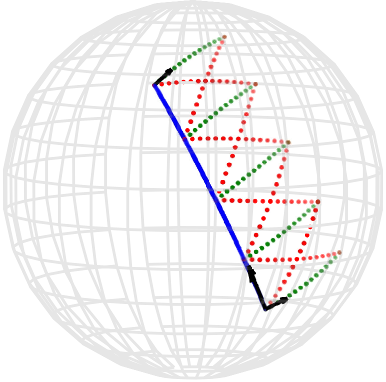
$$2a = w + v + \frac{1}{6}R(v,w)(w-v) + O(4)$$

$$u - u^w = \frac{1}{2}R(w,v)v + O(4)$$



Convergence of the iterated Schild’s ladder





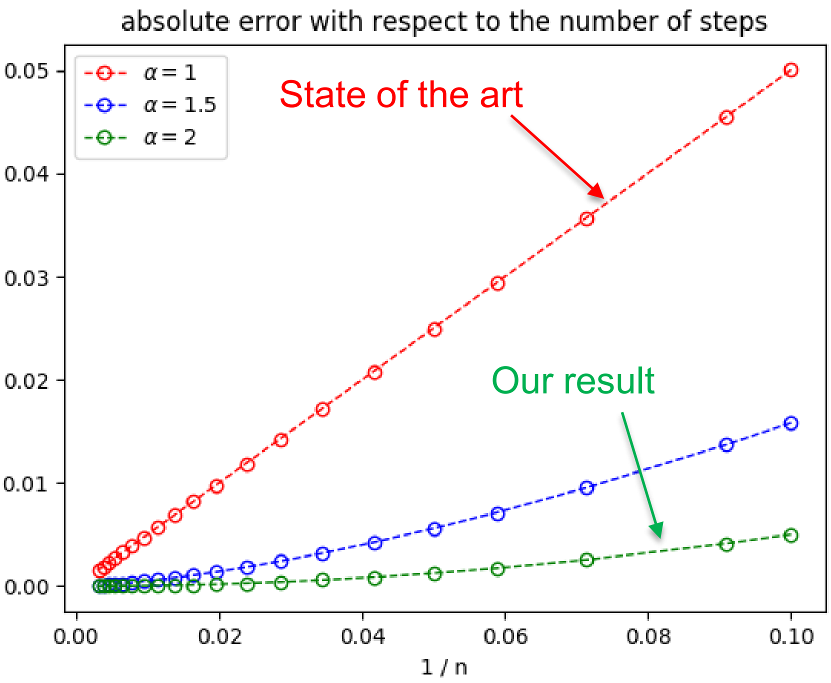
$$v_{i+1} = n^{\alpha} \cdot \text{schild}(x_i, \frac{w_i}{n}, \frac{v_i}{n^{\alpha}}),$$

Theorem: the scheme converge at speed $\|v_n - \Pi_x^{x_n} v\| \leq \frac{\tau}{n^{\alpha}} + \frac{\beta}{n^2}.$

[N. Guigui, XP, Numerical Accuracy of Ladder Schemes for Parallel Transport on Manifolds. Foundations of Computational Mathematics, 22:757-790, 06-2022. Arxiv 2007.07585.]

Convergence of Schild's Ladder

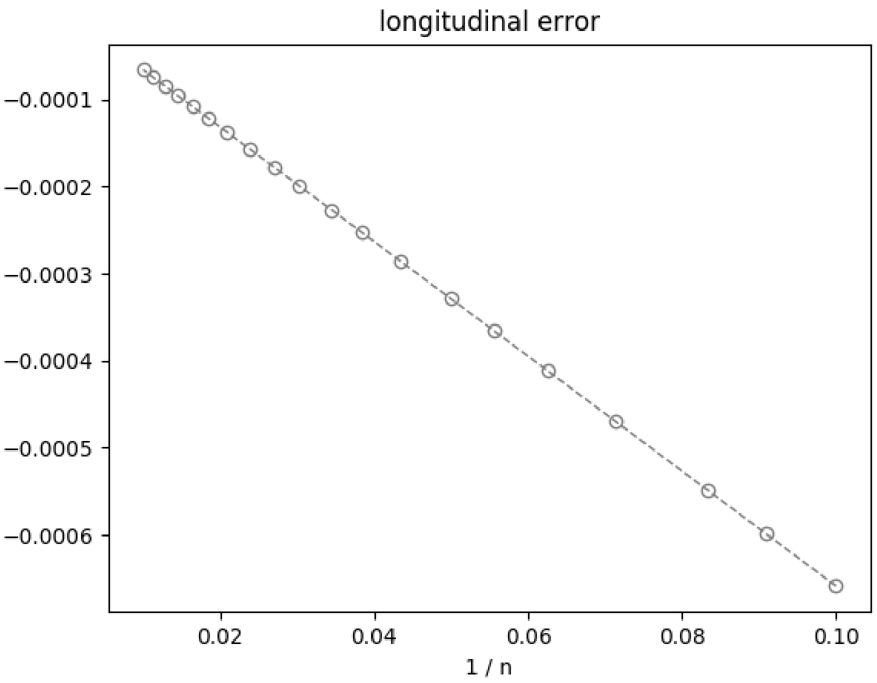
Numerical experiments in controlled spaces



State of the art

Our result

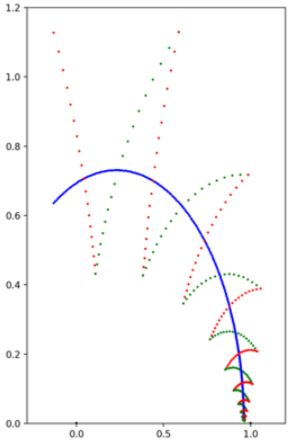
Simulations on the sphere:
constant curvature



Simulations on the space of SPD
matrices: negative curvature

[N. Guigui, XP, Numerical Accuracy of Ladder Schemes for Parallel Transport on Manifolds. Foundations of Computational Mathematics, 22:757-790, 06-2022. Arxiv 2007.07585.]

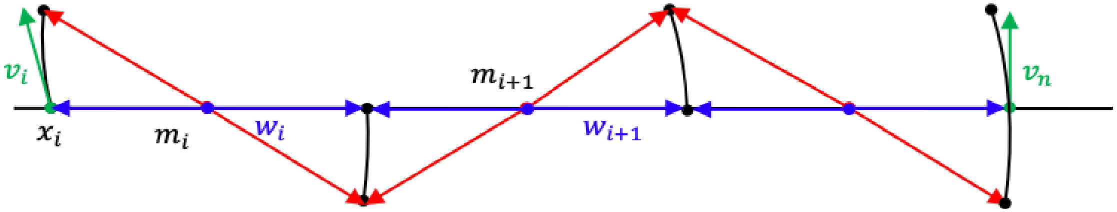
Convergence of pole Ladder

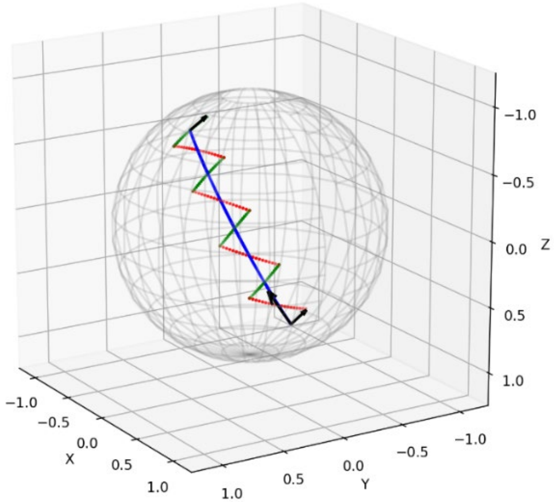


Taylor expansion of one pole ladder step

pole(u) $= \Pi(u) + \frac{1}{12}\nabla_v R(u,v)(5u-2v) + \frac{1}{12}\nabla_u R(u,v)(v-2u) + O(5)$

Convergence of the iterated pole ladder

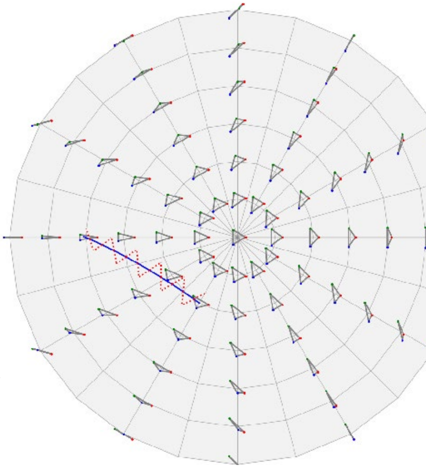




Theorem: the scheme converge at speed $\|v_n - \Pi_m^{m_n} v\| \leq \frac{\gamma}{n^2}$

[N. Guigui, XP, Numerical Accuracy of Ladder Schemes for Parallel Transport on Manifolds. Foundations of Computational Mathematics, 22:757-790, 06-2022. Arxiv 2007.07585.]

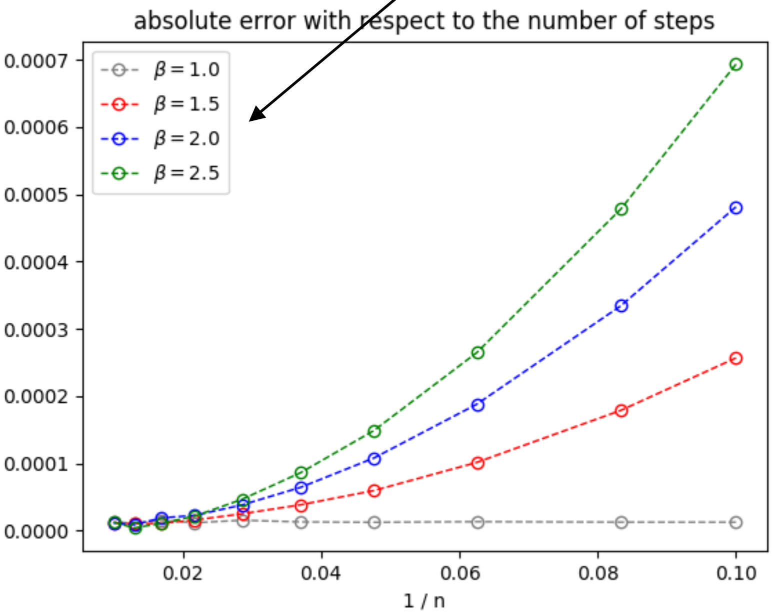
Convergence of pole Ladder

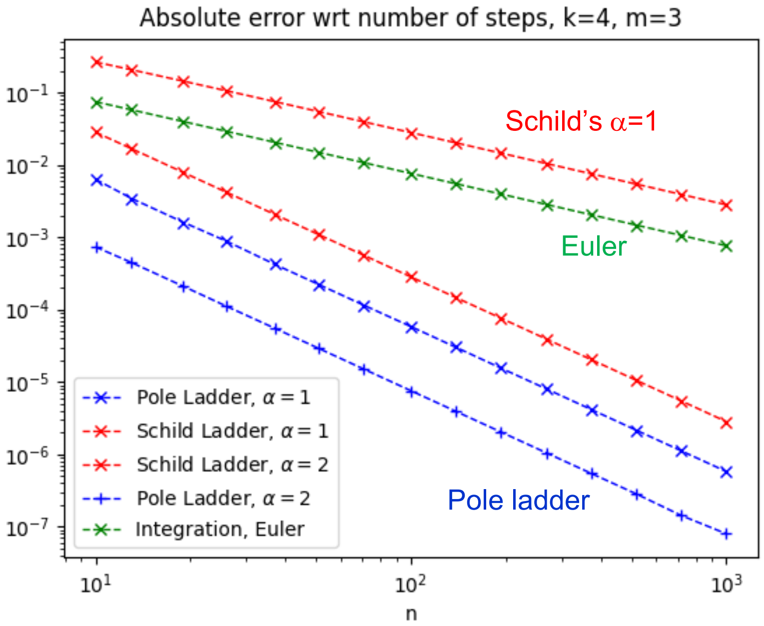


Numerical experiments in controlled spaces

Anisotropic metric on the Lie group SE(3)

Kendall shape space Σ_3^3



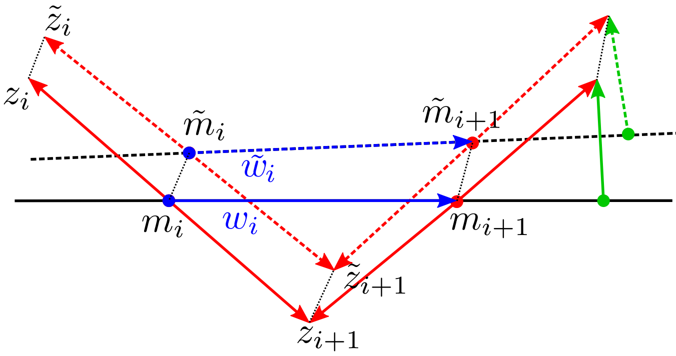


[N. Guigui, XP, Numerical Accuracy of Ladder Schemes for Parallel Transport on Manifolds. Foundations of Computational Mathematics, 22:757-790, 06-2022. Arxiv 2007.07585.]

Approximate geodesics & other schemes

Approximated geodesics

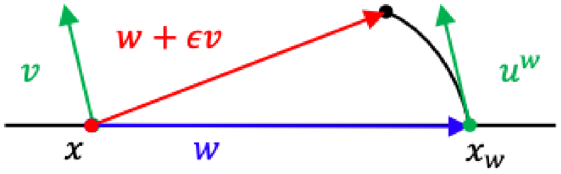
- Integration using Runge-Kutta
- Compute the log by gradient descent
- Convergence results remain valid with sufficiently accurate numerical scheme



Fanning Scheme [Louis et al 2018]

- Can be analyzed similarly
- Cannot ne made 2nd order

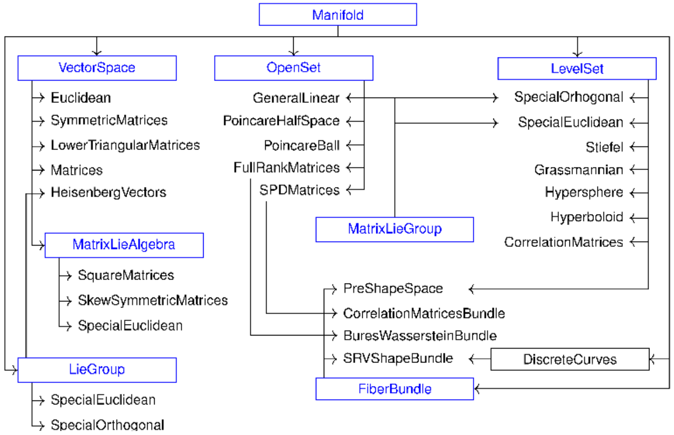
$$\|v_n - \Pi_x^{x_n} v\| \leq \frac{\beta}{n}.$$



$$\frac{u}{h\epsilon} = v + \frac{h^2}{6} R(w, v)w + O(h^3).$$

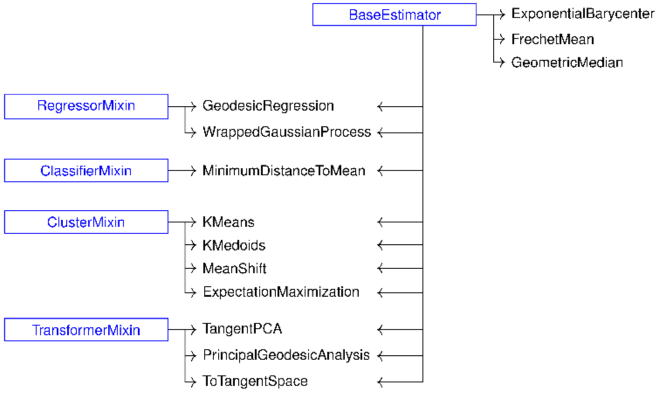
[N. Guigui, XP, Numerical Accuracy of Ladder Schemes for Parallel Transport on Manifolds. Foundations of Computational Mathematics, 22:757-790, 06-2022. Arxiv 2007.07585.]

http://geomstats.ai : a python library to implement generic algorithms on many Riemannian manifolds



Specific & generic manifolds

- Exp/Log map to generalize Euclidean tools
- **20+ specific manifolds / Lie groups** with closed-forms (SPD, H(n), SE(n), etc)
- Generic manifolds with **geodesics by integration / optimization**



Algorithms

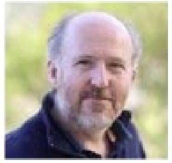
- **Fréchet mean, geodesic regression, tangent / geodesic PCA, Riemannian k-means, mean-shift, parallel transport**
- scikit-learn API (GPU & learning tools)
- Collaboration with pyriemann for BCI

N. Miolane

N. Guigui.

A. Le Brigant

X. Pennec



L. Pereira.

J. Deschamps.

A. Myers.

J. Mathe

A. Calissano

and many more collaborators

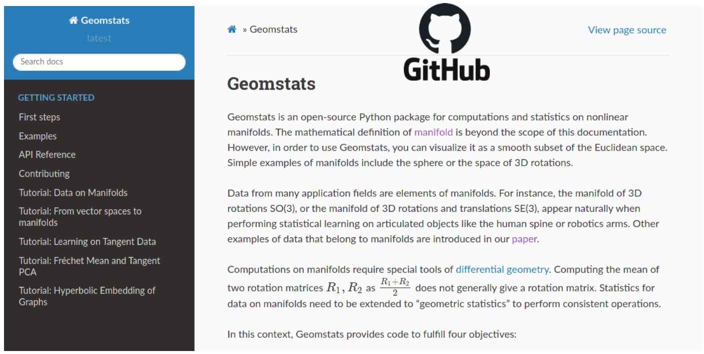
http://geomstats.ai : a python library to implement generic algorithms on many Riemannian manifolds

Collaborative development

- 10 introductory tutorials
- ~ 35000 lines of code
- ~20 academic developers
- 8 hackathons in 2020-2022, 1 Inria ADT

Semestre thématique IHP Geometry and Statistics in Data Science
Hackathon IHP Oct 17-21+ Journée Math & entreprises Nov 08, 2022

pypi package 2.5.0 Downloads 93k DOI 10.5281/zenodo.6478729



Geomstats
latest
Search docs
GETTING STARTED
First steps
Examples
API Reference
Contributing
Tutorial: Data on Manifolds
Tutorial: From vector spaces to manifolds
Tutorial: Learning on Tangent Data
Tutorial: Fréchet Mean and Tangent PCA
Tutorial: Hyperbolic Embedding of Graphs

» Geomstats
View page source
GitHub

Geomstats

Geomstats is an open-source Python package for computations and statistics on nonlinear manifolds. The mathematical definition of manifold is beyond the scope of this documentation. However, in order to use Geomstats, you can visualize it as a smooth subset of the Euclidean space. Simple examples of manifolds include the sphere or the space of 3D rotations.

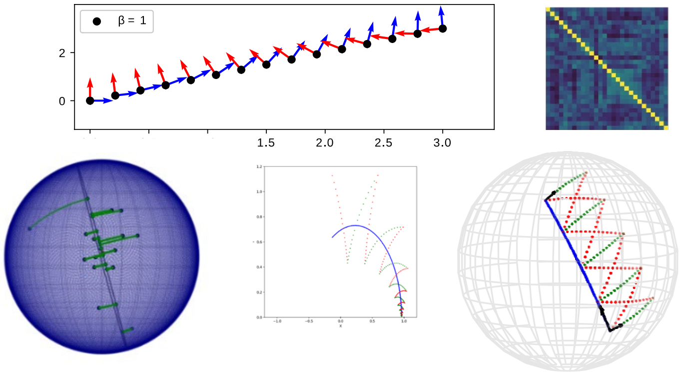
Data from many application fields are elements of manifolds. For instance, the manifold of 3D rotations SO(3), or the manifold of 3D rotations and translations SE(3), appear naturally when performing statistical learning on articulated objects like the human spine or robotics arms. Other examples of data that belong to manifolds are introduced in our paper.

Computations on manifolds require special tools of differential geometry. Computing the mean of two rotation matrices R_1, R_2 as $\frac{R_1+R_2}{2}$ does not generally give a rotation matrix. Statistics for data on manifolds need to be extended to "geometric statistics" to perform consistent operations.

In this context, Geomstats provides code to fulfill four objectives:

Interest in Machine Learning

- Miolane, Guigui, et al. SciPy Int. Conf. (2020).
- Miolane et al. **Journal of Machine Learning Research (2020)**
- Guigui, Miolane, Pennec. Intro. to Riem. Geom. and Geom. Stats: from basic theory to implementation with Geomstats. Monography of 164 p. **Foundations and Trends in Machine Learning (2023, 16 (3):329-493).**



and many more collaborators

Taylor expansion of geodesic triangles in Riemannian manifolds: a central tool to study the effect of curvature in geometric statistics

Motivations

Empirical Fréchet mean concentration

[XP, Curvature effects on the empirical mean in Manifolds 2019, arXiv:1906.07418]

Numerical accuracy of parallel transport algorithms

[N. Guigui, XP, Numerical Accuracy of Ladder Schemes for Parallel Transport
on Manifolds. Foundations of Computational Mathematics, 22:757-790, 2022]

Conclusions

Intrinsic Taylor expansions of geodesic triangles in manifolds

A new tool for the analysis of algorithms on manifolds

- Double exponential (Gavrilov) and neighboring log are **simple tensor series!**
- Also **valid in affine connection spaces** (Lie groups with CS connection)

Numerical accuracy of discrete parallel transport methods

- Jacobi field/fanning scheme is limited to order 1
- Schild's ladder can be made of order 2
- **Simpler pole ladder is order 2 + exact in one step in symmetric spaces**

Riemannian manifolds with no closed-form geodesics

- Computing geodesics by integration and log by gradient descent
- **Theorems continue to hold, implementation available in** http://geomstats.ai
- ~~log by gradient descent~~: natural schemes for mid-point/doubling rule?

Numerical accuracy of other geodesics-based algorithms?

Empirical and population means: curvature?

Curvature-covariance controls bias and concentration modulation

- **Bias** on empirical mean (**gradient of curvature-covariance**)

$$\mathbf{Bias}(\bar{x}_n) = \boldsymbol{E}\left(log_{\bar{x}}(\bar{x}_n)\right) = \frac{\mathbf{1}}{\boldsymbol{6n}}\left(\mathfrak{M}_2 : \nabla R : \mathfrak{M}_2\right) + O(\epsilon^5, 1/n^2)$$

- **Concentration rate** modulated by the **curvature-covariance:**

$$\mathbf{Cov}(\bar{x}_n) = \boldsymbol{E}\left(log_{\bar{x}}(\bar{x}_n) \otimes log_{\bar{x}}(\bar{x}_n)\right) = \frac{\mathbf{1}}{\boldsymbol{n}}\mathfrak{M}_2 + \frac{\mathbf{1}}{\boldsymbol{3n}}\mathfrak{M}_2 : R : \mathfrak{M}_2 + O(\epsilon^5, 1/n^2)$$

  - **Faster convergence (asymptotically infinitely)** for negative curvature
  - **Slower convergence (up to no convergence at KKC limit)** in positive curvature

Lesson for AI: high curvature has drastic impact with small data!

- High concentration and asymptotic predictions are confirmed by real data
- Lower concentration: **prelude to stickiness / smeariness**
  [Hotz et al 2013] [Huckemann & Eltzner 2019, 2020]

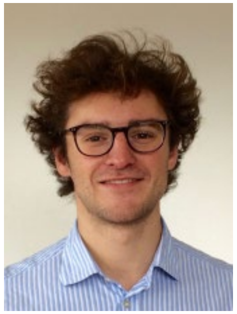
Curvature at a ~~point~~ distribution: deviation from Euclidean CLT?

- **Distributional torsion:** $\lim_{n\to\infty} \mathrm{n}\ \mathrm{Bias}(\bar{x}_n) \cong \frac{1}{6}\mathfrak{M}_2 : \nabla R : \mathfrak{M}_2 + O(\epsilon^5)$
- **Distributional curvature:** $\lim_{n\to\infty} \mathrm{n}\ \mathrm{Cov}(\bar{x}_n) - \mathrm{Cov}(x) \cong \frac{1}{3}\mathfrak{M}_2 : R : \mathfrak{M}_2 + O(\epsilon^5)$
  - Differs from Efron's "statistical curvature" of a family of distributions [Efron, AoS 1975]
  - Relation to coarse [Ollivier 07,09] & synthetic Ricci curvatures [Sturm 06 Lott-Villani 09]?

The G-Statistics group

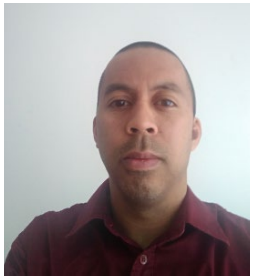

Yann Thanwerdas

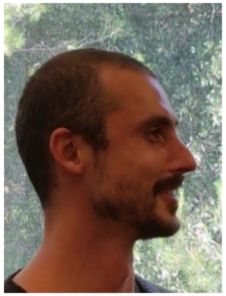
Geometric Statistics


Nicolas Guigui


Morten Pedersen

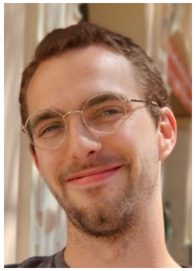

Elodie Maignant


Dimbihery Rabenoro

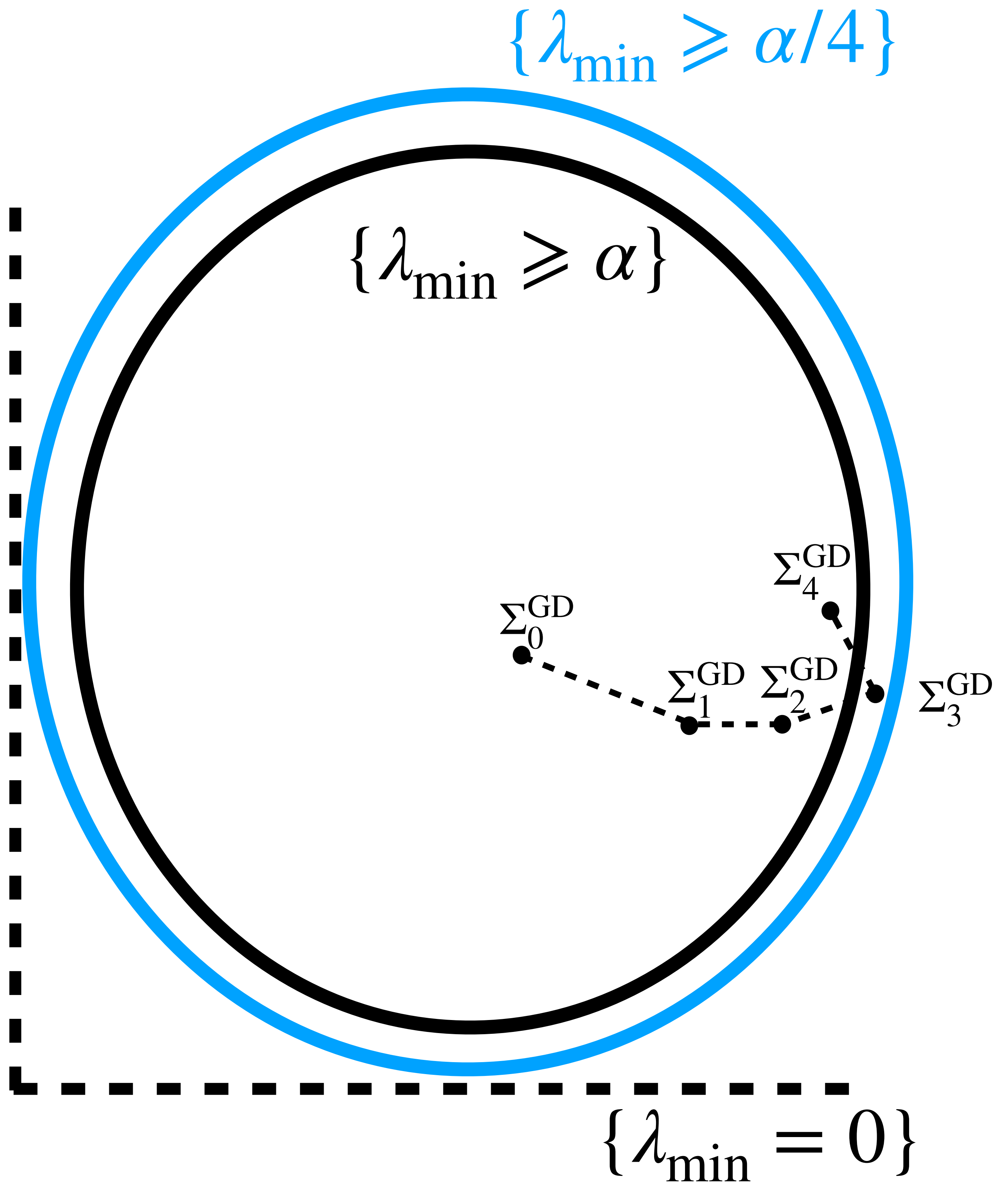

James Benn


Anna Calissano


Luis G. Pereira


Tom Szwagier

Global, dimension-free convergence of first-order methods for Bures-Wasserstein barycenters

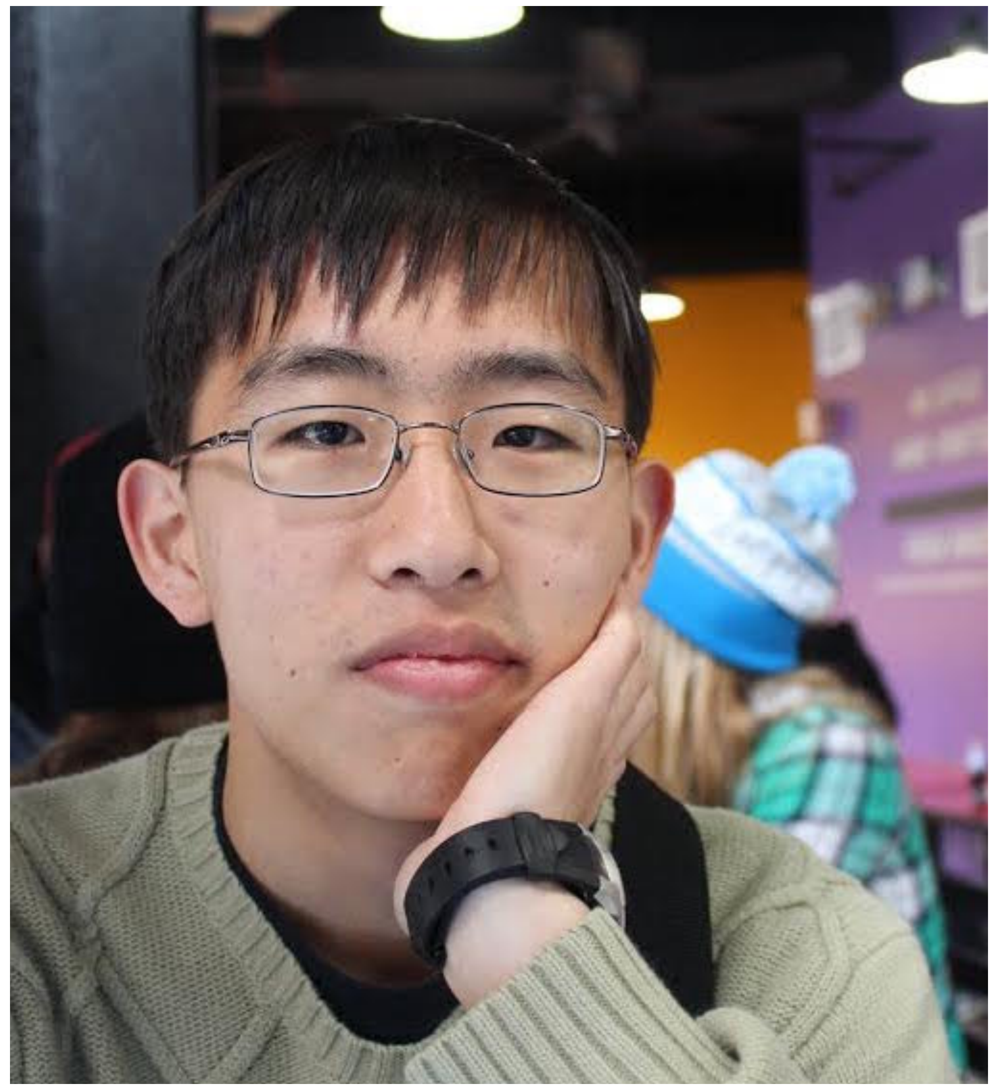


Austin J. Stromme
12/10/2023

ENSAE | CREST
CENTER FOR RESEARCH IN ECONOMICS AND STATISTICS

Joint work

Jason Altschuler
(UPenn)



Sinho Chewi
(IAS)

Patrik Gerber
(MIT)

Tyler Maunu
(Brandeis)

Philippe Rigollet
(MIT)

Averaging on non-Euclidean spaces

For a metric-measure space (X, d_X, P), the **barycenter problem** asks

$$\min_{b \in X} \int d_X^2(b, x) \mathrm{d}P(x)$$

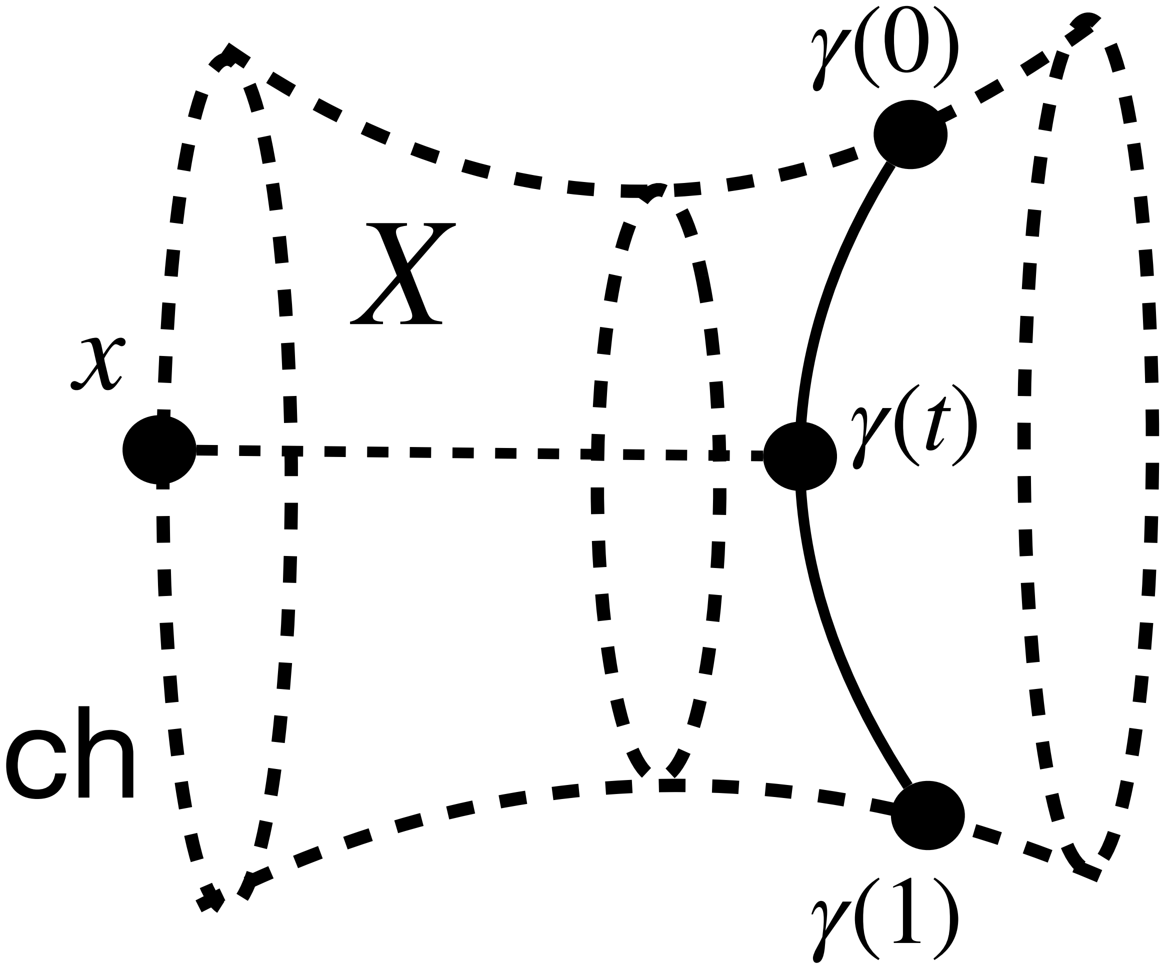
Existence and uniqueness?

Statistical convergence?

Algorithms?

[Fréchet ’48, Karcher ’77]

Barycenters and geometry



Suppose (X, d_X) is complete, geodesic metric space. For each $P \in \mathscr{P}_2(X, d_X)$, let

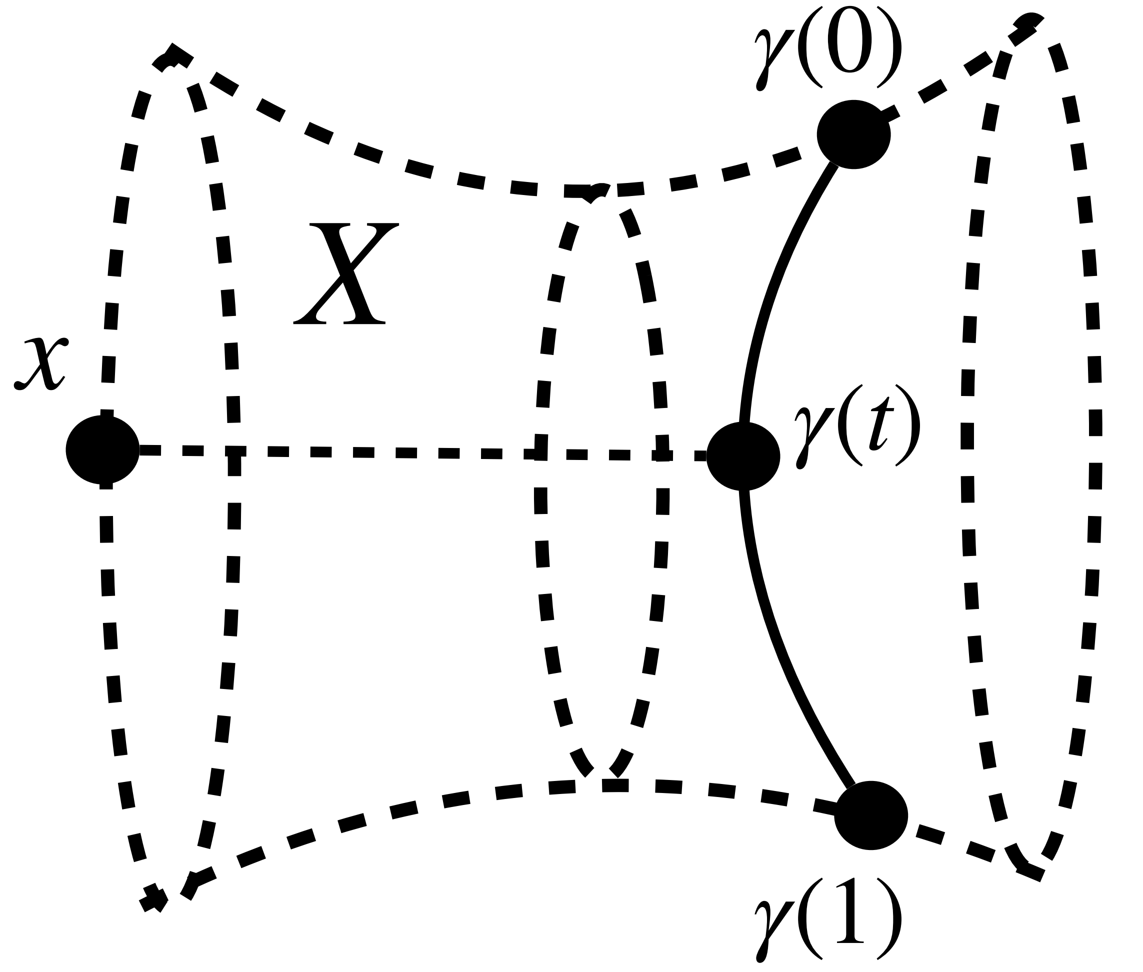
$$F_P(b) := \int d_X^2(b, x)\mathrm{d}P(x)$$

Then F_P is 1-geodesically convex for all $P \in \mathscr{P}_2(X, d_X)$ **if and only if** (X, d_X) is non-positively curved (in sense of Alexandrov)

$$F_P(\gamma(t)) \leqslant (1-t)F_P(\gamma(0)) + tF_P(\gamma(1)) - \frac{1}{2}t(1-t)d_X^2(\gamma(0), \gamma(1))$$

[Sturm ’03]

Barycenters in NPC spaces



Suppose (X, d_X) is an NPC space. For each $P \in \mathscr{P}_2(X, d_X)$, let

$$F_P(b) := \int d_X^2(b, x)\mathrm{d}P(x)$$

Then:

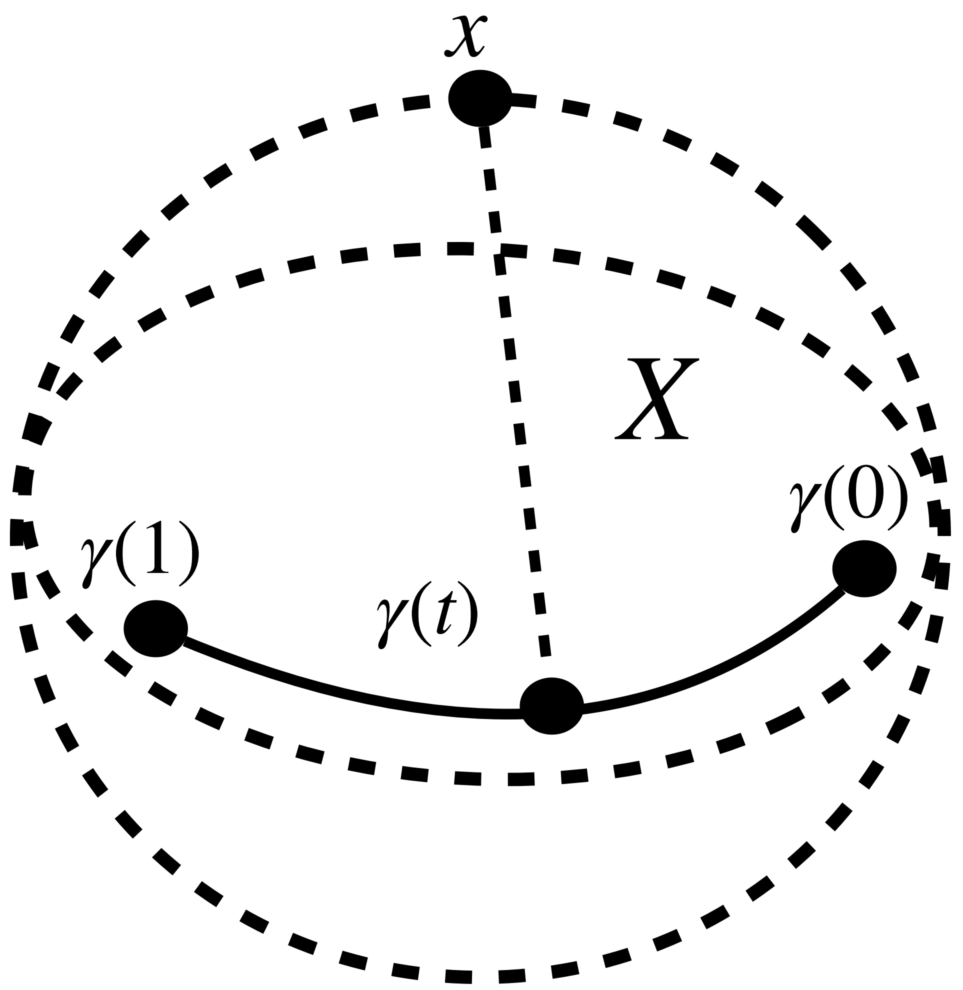
Existence and uniqueness

Statistical convergence

Algorithms

[Sturm '03, Bhattacharya, Patrangenaru '06, Le Gouic, Paris, Rigollet, and Stromme '22, Brunel, Serres '23]

Barycenters in NNC spaces



Suppose (X, d_X) is a NNC space, so

$$d_X^2(\gamma(t), x) \geqslant (1-t)d_X^2(\gamma(0), x) + td_X^2(\gamma(1), x) - \frac{1}{2}t(1-t)d_X^2(\gamma(0), \gamma(1))\,.$$

Then:

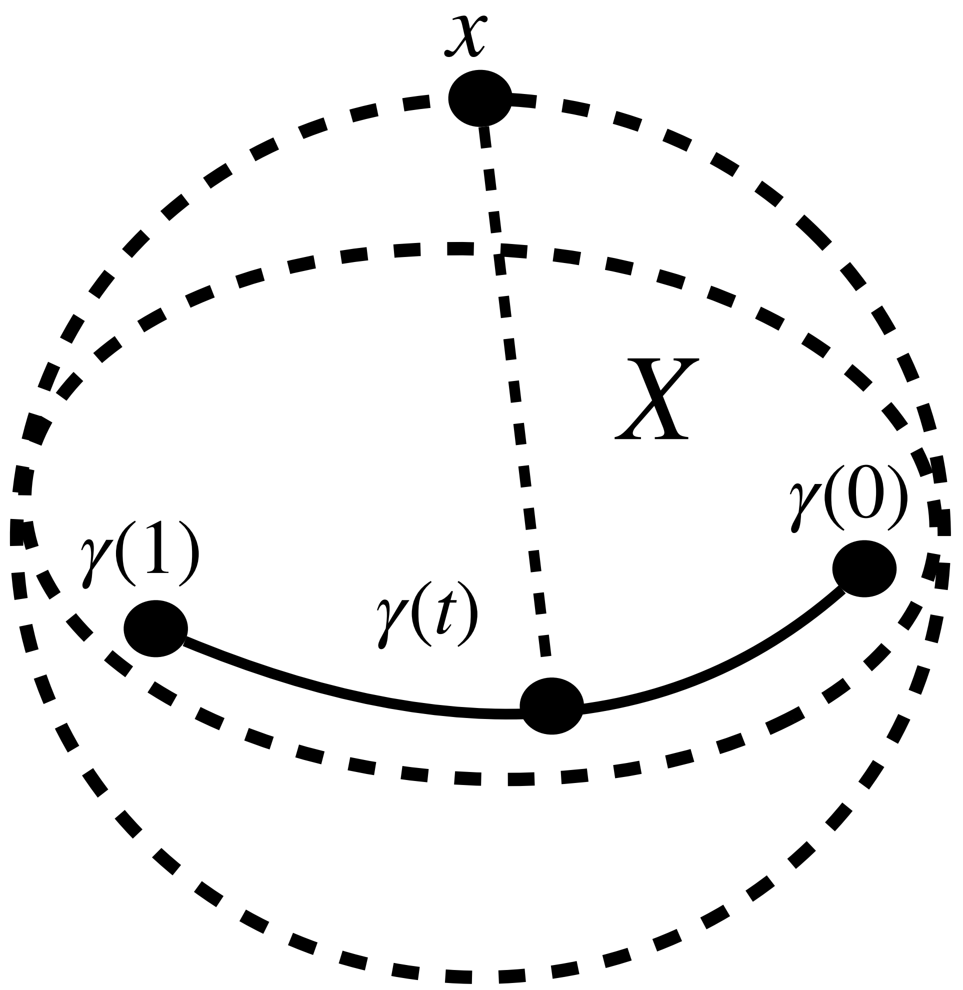
Existence and uniqueness?

Statistical convergence?

Algorithms?

[Bhattacharya, Patrangenaru '06, Afsari '11]

Barycenters in NNC spaces



Suppose (X, d_X) is a NNC space, so

$$d_X^2(\gamma(t), x) \geqslant (1-t)d_X^2(\gamma(0), x) + t d_X^2(\gamma(1), x) - \frac{1}{2}t(1-t)d_X^2(\gamma(0), \gamma(1))\,.$$

Then:

Existence and uniqueness?

Statistical convergence?

Algorithms?

Can restrict to a small ball, but that isn't completely satisfying

[Bhattacharya, Patrangenaru '06, Afsari '11]

The Wasserstein space

$$W_2(\mu,\nu) := \min_{\pi \in \Pi(\mu,\nu)} \left(\int \|x-y\|^2 \mathrm{d}\pi(\mathrm{x},\mathrm{y}) \right)^{1/2} \quad W_2(\mathbb{R}^d) := (\mathscr{P}_2(\mathbb{R}^d), W_2)$$

Endows space of probability distributions with a Riemannian-like geometry:

- Tangent spaces, exponential maps, geodesics, **calculus**
- Fundamental in PDEs, functional inequalities, geometry of non-smooth spaces
- Fundamental for sampling algorithms

[McCann '97, Jordan, Kindelehrer, and Otto '98, Otto and Villani '00, Otto '01, Sturm '06, Lott and Villani '06, Ambrosio, Gigli, and Savare '08]

The Wasserstein space is NNC

$$W_2(\mu,\nu) := \min_{\pi\in\Pi(\mu,\nu)} \left(\int \|x-y\|^2 \mathrm{d}\pi(\mathrm{x},\mathrm{y})\right)^{1/2} \quad W_2(\mathbb{R}^d) := (\mathscr{P}_2(\mathbb{R}^d), W_2)$$

The Wasserstein space $W_2(\mathbb{R}^d)$ is non-negatively curved:

$$W_2^2(\gamma(t),\mu) \geqslant (1-t)d_X^2(\gamma(0),\mu) + tW_2^2(\gamma(1),\mu) - \frac{1}{2}t(1-t)W_2^2(\gamma(0),\gamma(1))\,.$$

[Otto ’01, Ambrosio, Gigli, and Savare ’08]

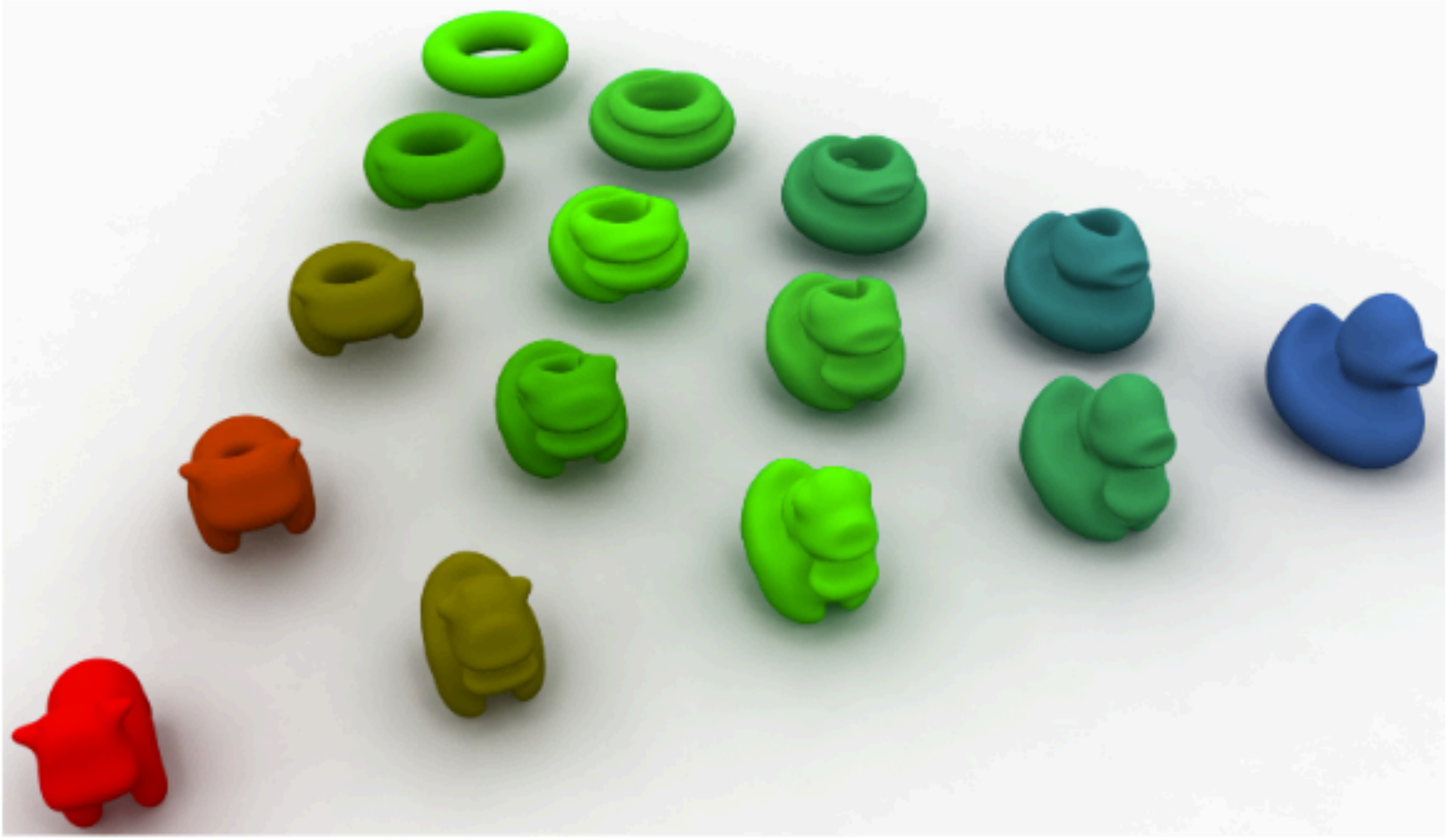
Wasserstein barycenters

Given $P \in \mathscr{P}(\mathscr{P}_2(\mathbb{R}^d))$

$$\min_{b \in \mathscr{P}_2(\mathbb{R}^d)} \int W_2^2(b, \mu) \, \mathrm{dP}(\mu)$$

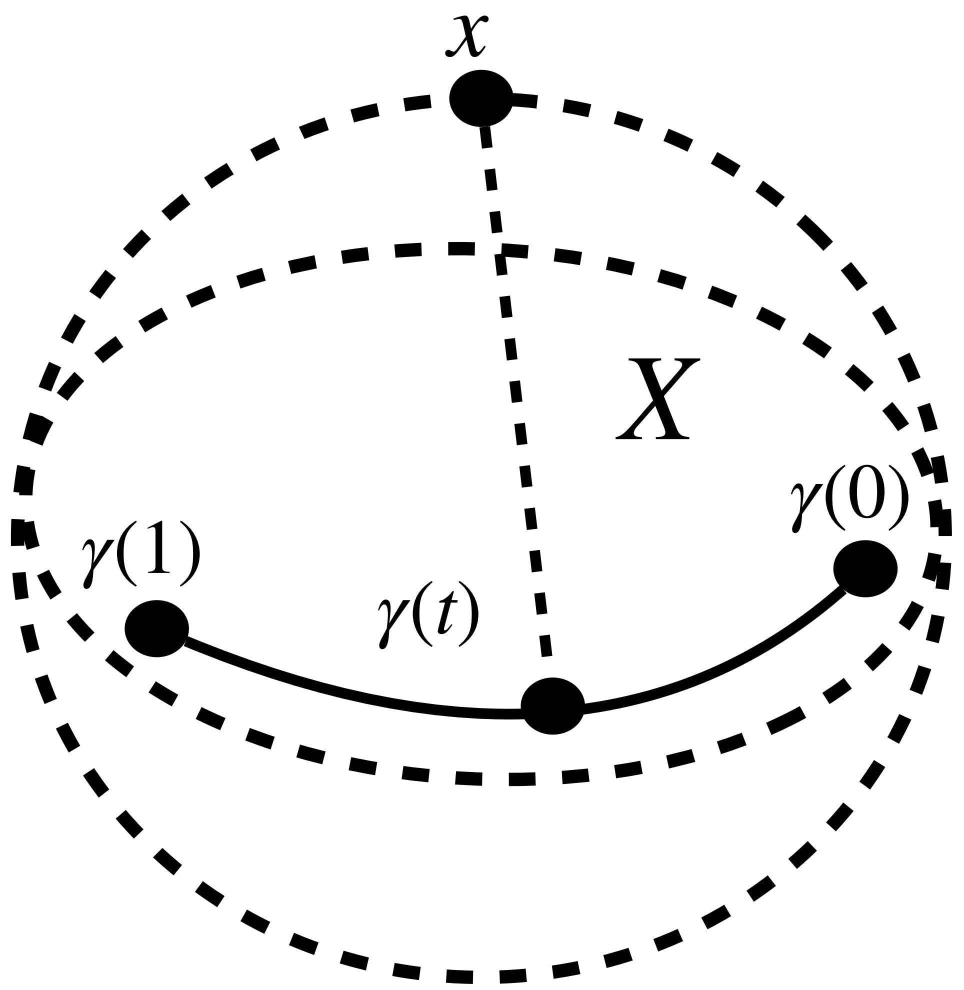
- Graphics
- Bayesian statistics
- Transfer learning
- Trajectory reconstruction
- …

[Solomon et al ’15]



AC’11, CD’14, CFTR’16, AC’17, LGL’17, ZP’19, KSS’19, ALP’18, S’03, O’12, Y’16, S’18, CCS’19, CAD’19, ABA’21, ABA’21, BVFR’22, ABA’22, CDM’22, JRE’23, +

Wasserstein barycenters



Given $P \in \mathscr{P}(\mathscr{P}_2(\mathbb{R}^d))$, solve

$$\min_{b \in \mathscr{P}_2(\mathbb{R}^d)} \int W_2^2(b, \mu)\, \mathrm{dP}(\mu)$$

Surprisingly, the NNC is rather benign:

Existence and uniqueness (under mild conditions)

Statistical convergence (under various conditions)

Algorithms (this talk)

[Agueh and Carlier ’11, Kroshnin, Spokoiny, Suvorikova ’19, Ahidar-Coutrix, Le Gouic, Paris ’20, Carlier, Delalande, Merigot ’22]

First-order methods for Wasserstein barycenters

How to solve

$$\min_{b \in \mathscr{P}_2(\mathbb{R}^d)} F_P(b) := \int W_2^2(b, \mu)\, \mathrm{dP}(\mu)\,?$$

(Cuturi and Doucet ’14): **gradient descent using the Wasserstein geometry!**

$$\nabla_{W_2} F_P(b) = \int (\nabla \varphi_{b \to \mu} - \mathrm{id})\, \mathrm{d}P(\mu), \qquad b_{t+1} = ((1 - \eta_t)\mathrm{id} + \eta_t \nabla_{W_2} F_P(b_t))_{\#} b_t$$

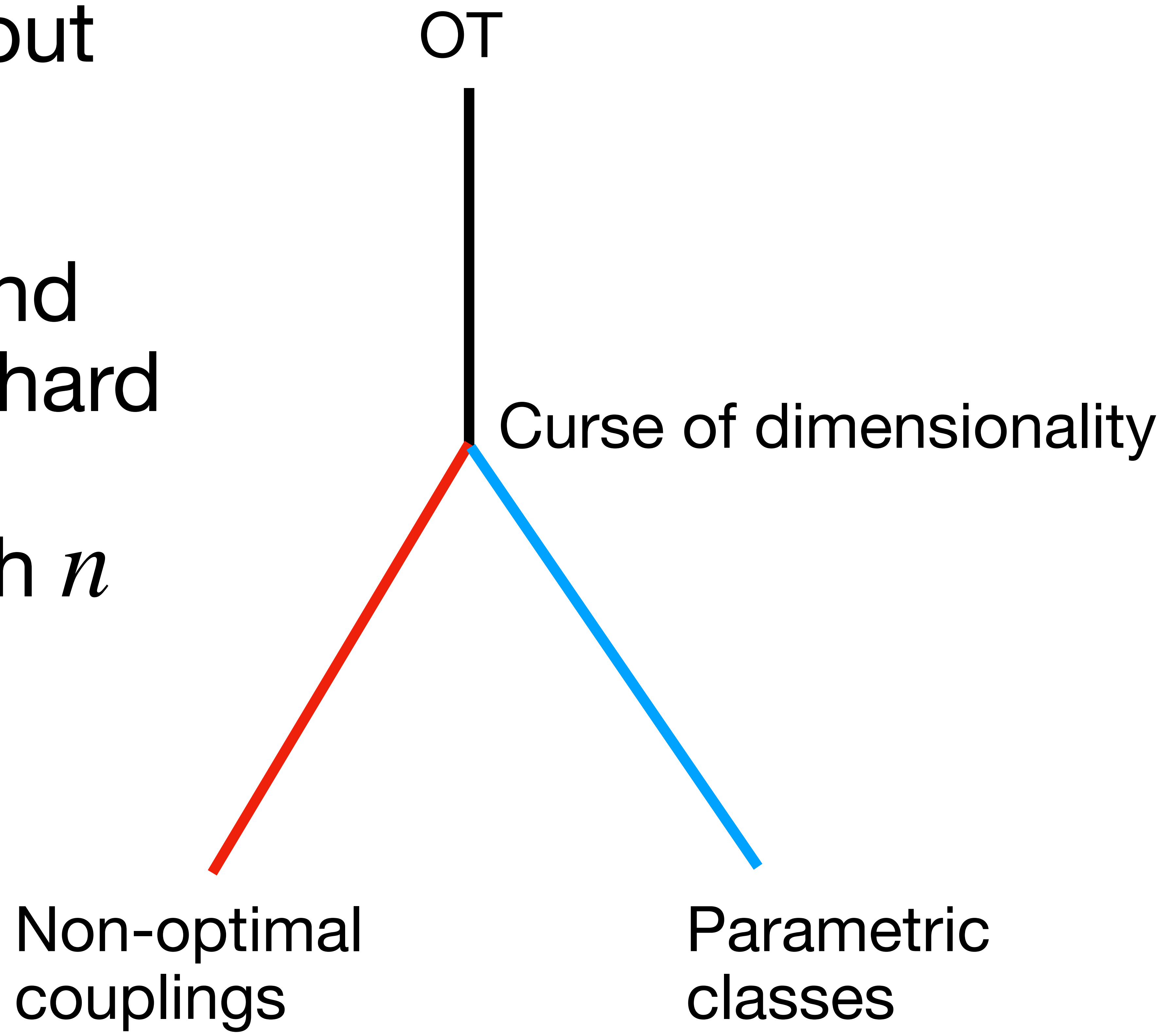
[Cuturi and Doucet ’14]

Curse of dimensionality for Wasserstein space

Unfortunately, this won't work in high dimensions without further assumptions:

Computational curse of dimensionality: Altschuler and Boix-Adsera showed Wasserstein barycenters are NP-hard

Statistical curse of dimensionality: Discretization with n samples entails unavoidable statistical error $n^{-1/d}$



[Dudley '69, Niles-Weed and Rigollet '21, Altschuler, Boix-Adsera '22]

Restricting to Gaussians

Multivariate Gaussians form an especially well-behaved subset of $W_2(\mathbb{R}^d)$:

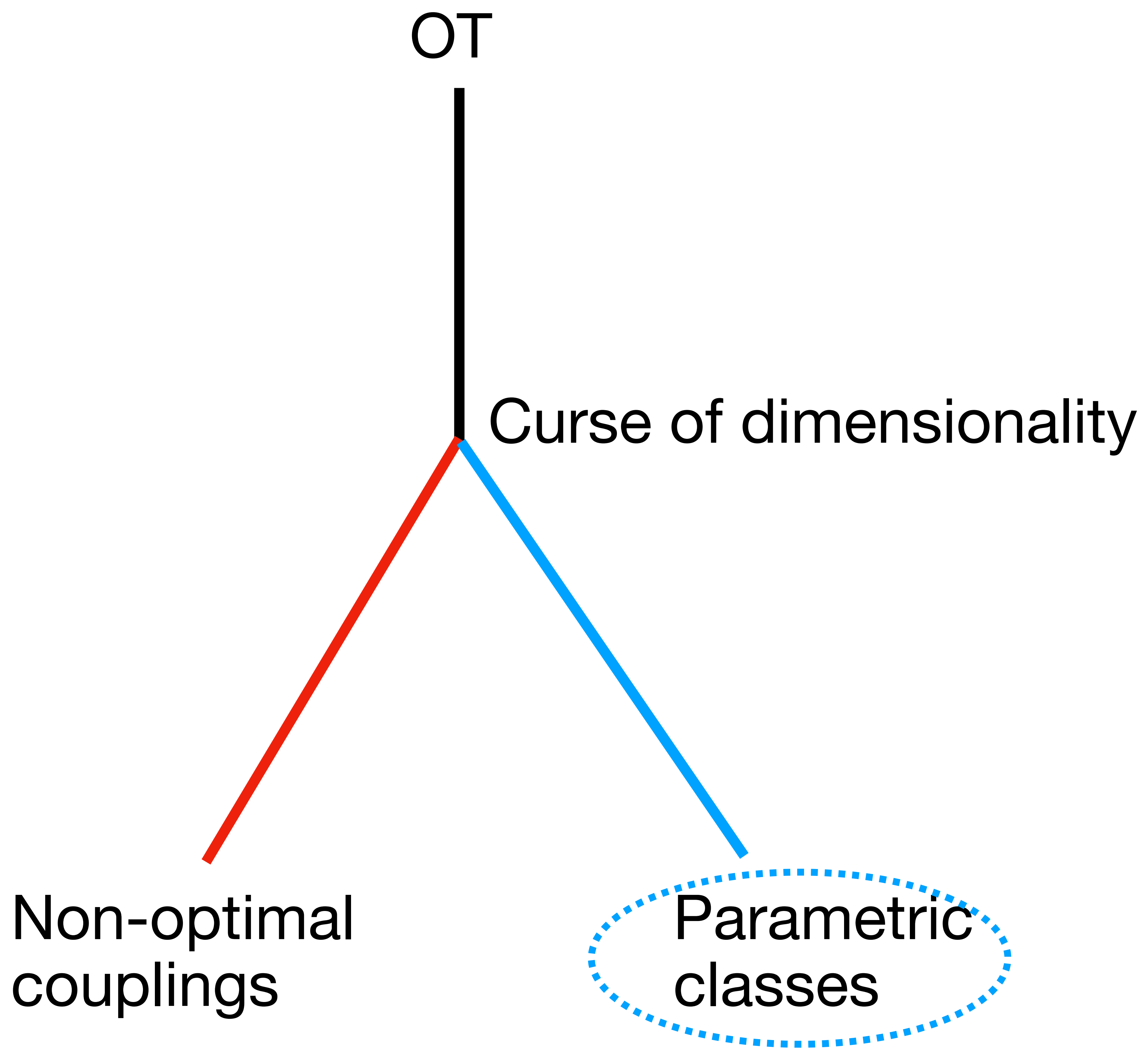
- Totally geodesic subset (i.e. convex)
- Closed form for distances

$$W_2^2(\Sigma_0, \Sigma_1) = \mathrm{tr}(\Sigma_0) + \mathrm{tr}(\Sigma_1) - 2\mathrm{tr}((\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2})^{1/2})\,.$$

- Closed form for geodesics

$$\Sigma_{0\to 1} := \Sigma_0^{-1/2}(\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2})^{1/2}\Sigma_0^{-1/2}$$

$$\Sigma_t = (1-t)^2\Sigma_0 + t^2\Sigma_1 + t(1-t)(\Sigma_0\Sigma_{0\to 1} + \Sigma_{0\to 1}\Sigma_0)$$



OT

Curse of dimensionality

Non-optimal couplings

Parametric classes

The Bures-Wasserstein manifold

A **non-negatively curved** Riemannian manifold on the set of positive-definite matrices

$$\mathbb{B}_d := (\{\Sigma \in \mathbb{R}^{d \times d} \colon \Sigma = \Sigma^T,\ \Sigma \succ 0\}, W_2)$$

Many connections and uses:

- theory of deep learning/implicit regularization
- Pre-conditioner for OT in applications
- SDP solvers

[Bures ’69, Knott, Smith ’94, Burer, Monteiro ’03, Burer, Monteiro ’05, Alvarez-Esteban et al ‘16, Bhatia, Jain and Lim ’19, Kroshnin, Spokoiny, Suvorikova ’19]

Riemannian gradient descent for BW barycenters

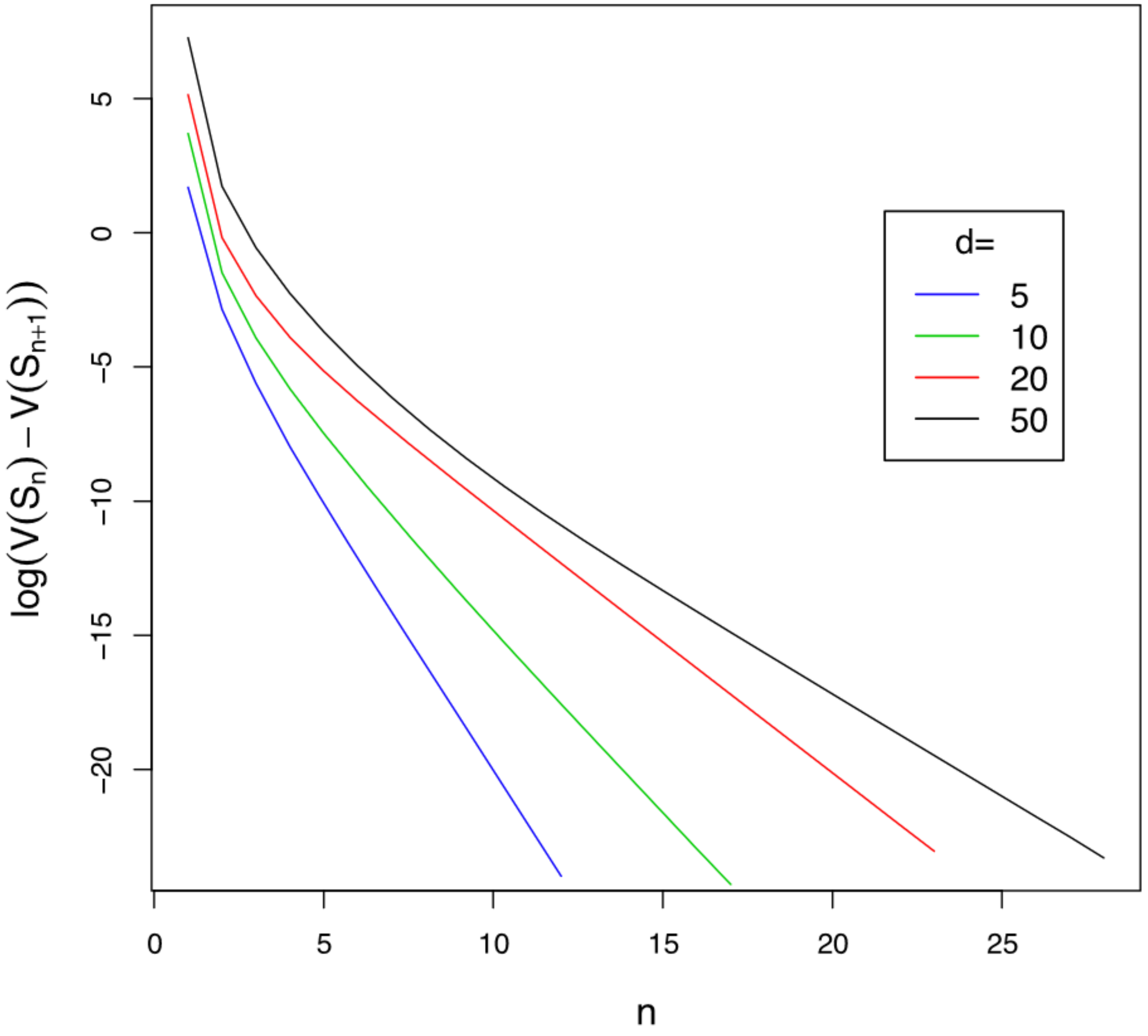
Can explicitly compute the gradient of the BW barycenter functional

$$\nabla_{W_2} F_P(\Sigma_t) = \int \Sigma_t^{-1/2}(\Sigma_t^{1/2}\Sigma\Sigma_t^{1/2})^{1/2}\Sigma_t^{-1/2}\mathrm{dP}(\Sigma)\,.$$

And the GD update with step-size η_t

$$\Sigma_{t+1} = (I - \eta_t \nabla_{W_2} F_P(\Sigma_t))\Sigma_t(I - \eta_t \nabla_{W_2} F_P(\Sigma_t))$$

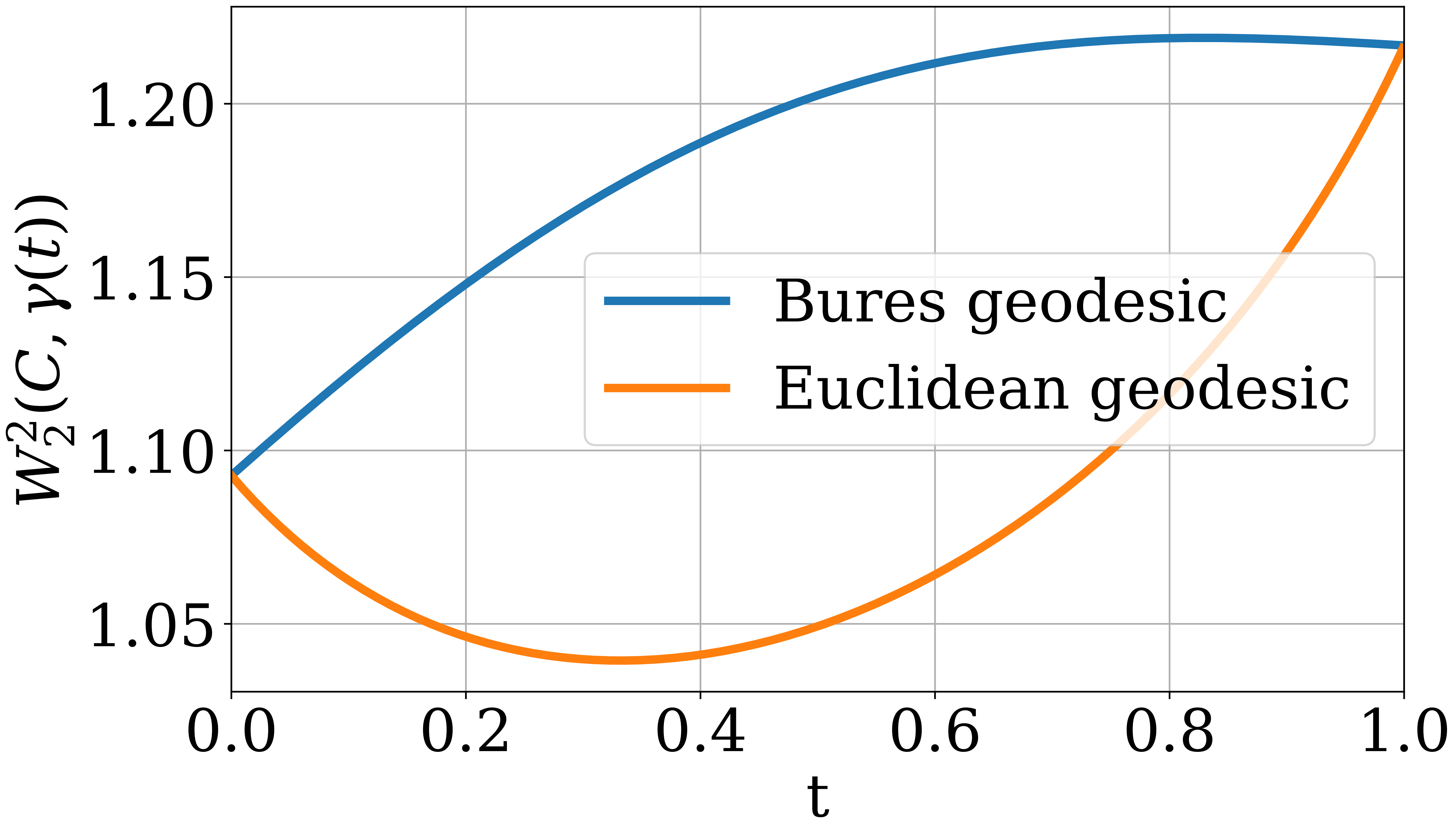
Converges quickly in practice



Plot of convergence vs iterations from Alvarez-Esteban et al ’16

[Alvarez-Esteban et al ‘16]

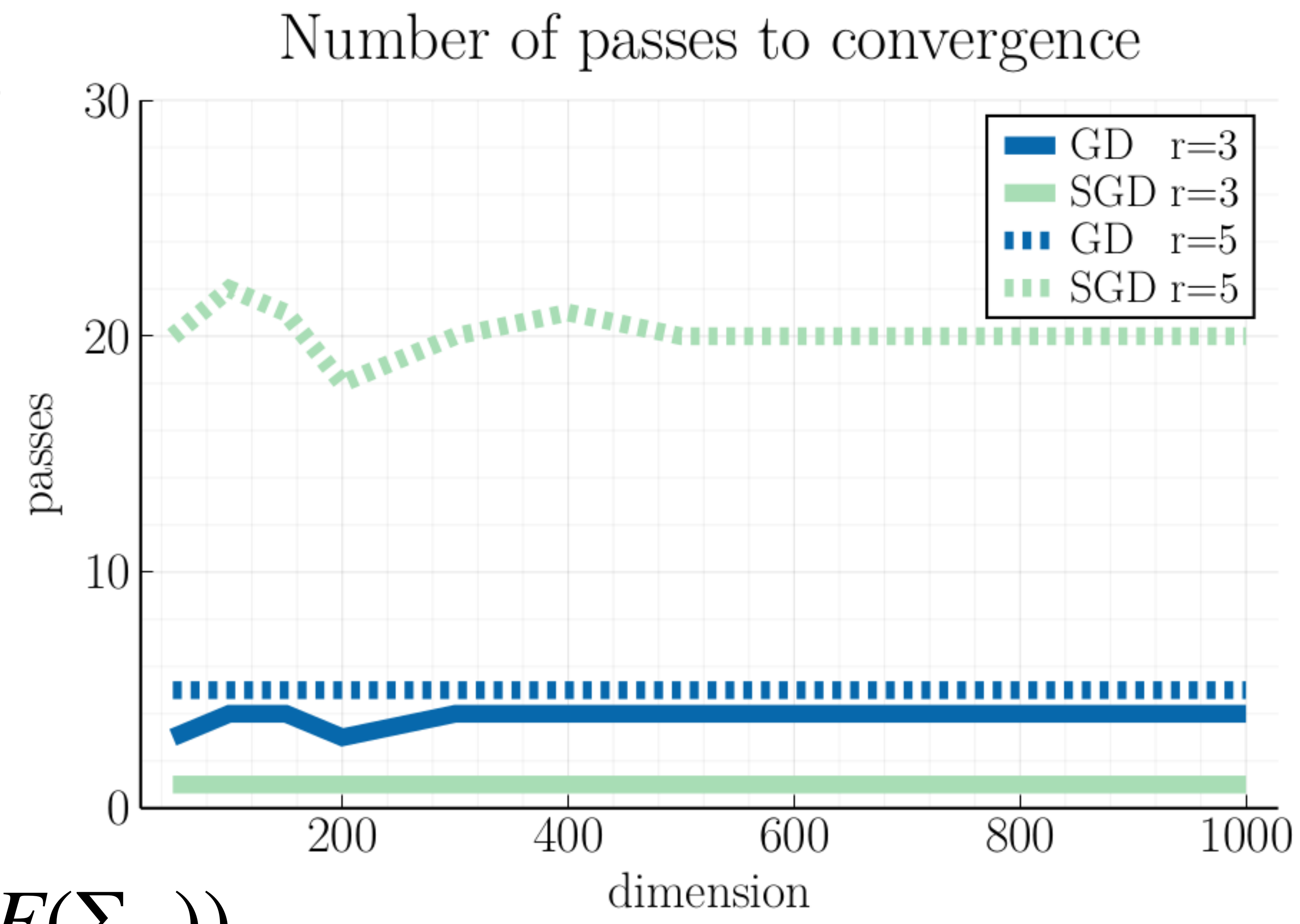
Non-convexity of BW barycenter functional



Dimension-free, global, linear rates for GD

$$\Sigma_{t+1} = (I - \eta_t \nabla_{W_2} F_P(\Sigma_t))\Sigma_t(I - \eta_t \nabla_{W_2} F_P(\Sigma_t))$$

Theorem. (CMR**S**'20, ACG**S**'21)
Suppose P is supported on centered Gaussians with eigenvalues in the range $[\alpha, \beta]$. Then GD with step-size $\eta_t := \alpha/2\beta$ converges as



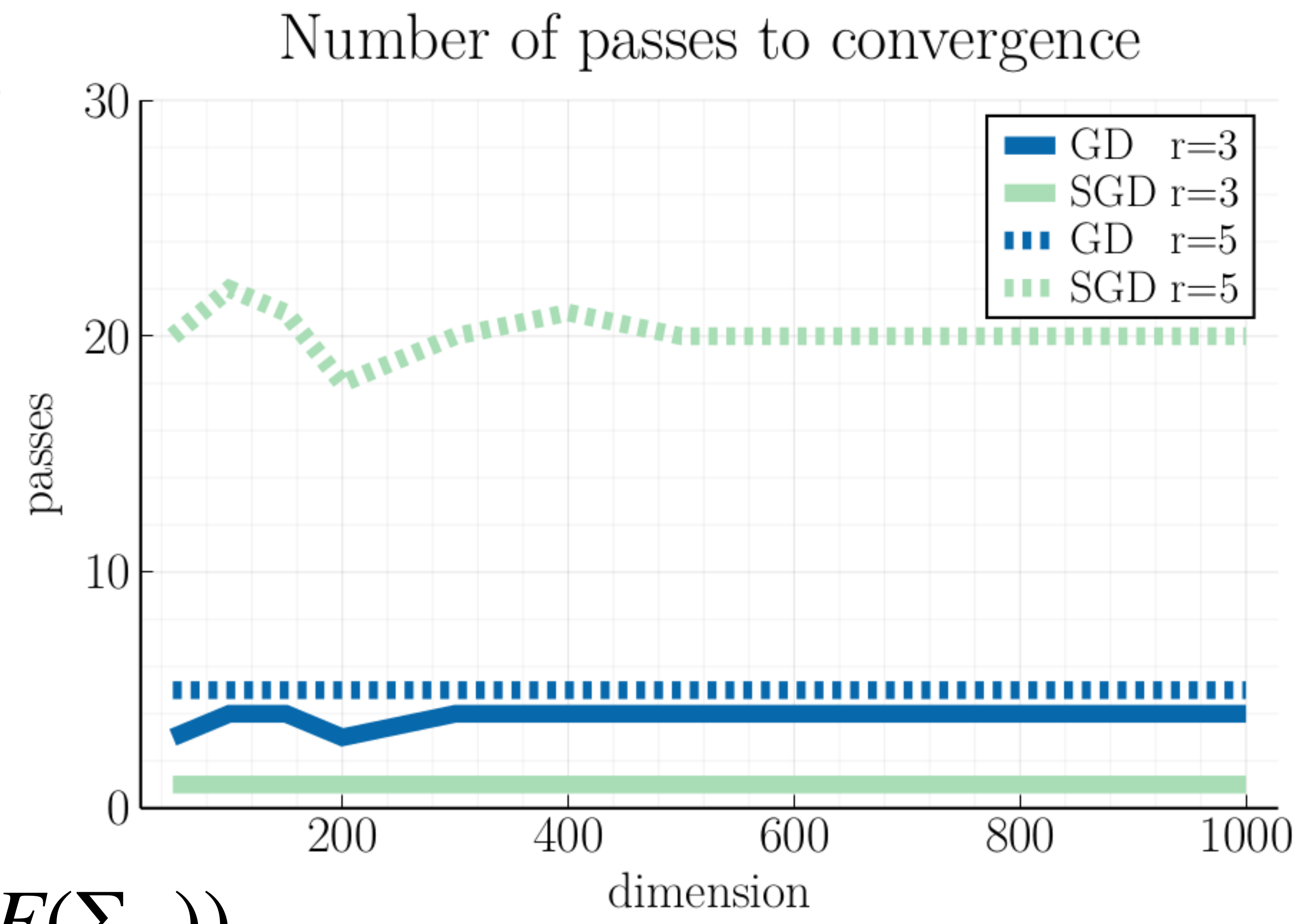
$$F(\Sigma_T) - F(\Sigma_\star) \leqslant \exp\Big(- \frac{3T}{64} \cdot \big(\frac{\alpha}{\beta}\big)^{5/2}\Big) \cdot (F(\Sigma_0) - F(\Sigma_\star))\,.$$

[Chewi, Maunu, Rigollet, Stromme '20,
Altschuler, Chewi, Gerber, Stromme '21]

Dimension-free, global, linear rates for GD

$$\Sigma_{t+1} = (I - \eta_t \nabla_{W_2} F_P(\Sigma_t))\Sigma_t(I - \eta_t \nabla_{W_2} F_P(\Sigma_t))$$

Theorem. (CMR**S**'20, ACG**S**'21)
Suppose P is supported on centered Gaussians with eigenvalues in the range $[\alpha, \beta]$. Then GD with step-size $\eta_t := \alpha/2\beta$ converges as



$$F(\Sigma_T) - F(\Sigma_\star) \leqslant \exp\Big(- \frac{3T}{64} \cdot \big(\frac{\alpha}{\beta}\big)^{5/2}\Big) \cdot (F(\Sigma_0) - F(\Sigma_\star))\,.$$

In fact, this holds for the *average* condition numbers

$$\alpha := \Big(\int \sqrt{\lambda_{\min}(\Sigma)}\mathrm{d}P(\Sigma)\Big)^2 \quad \beta := \Big(\int \sqrt{\lambda_{\max}(\Sigma)}\mathrm{d}P(\Sigma)\Big)^2\,.$$

[Chewi, Maunu, Rigollet, Stromme '20,
Altschuler, Chewi, Gerber, Stromme '21]

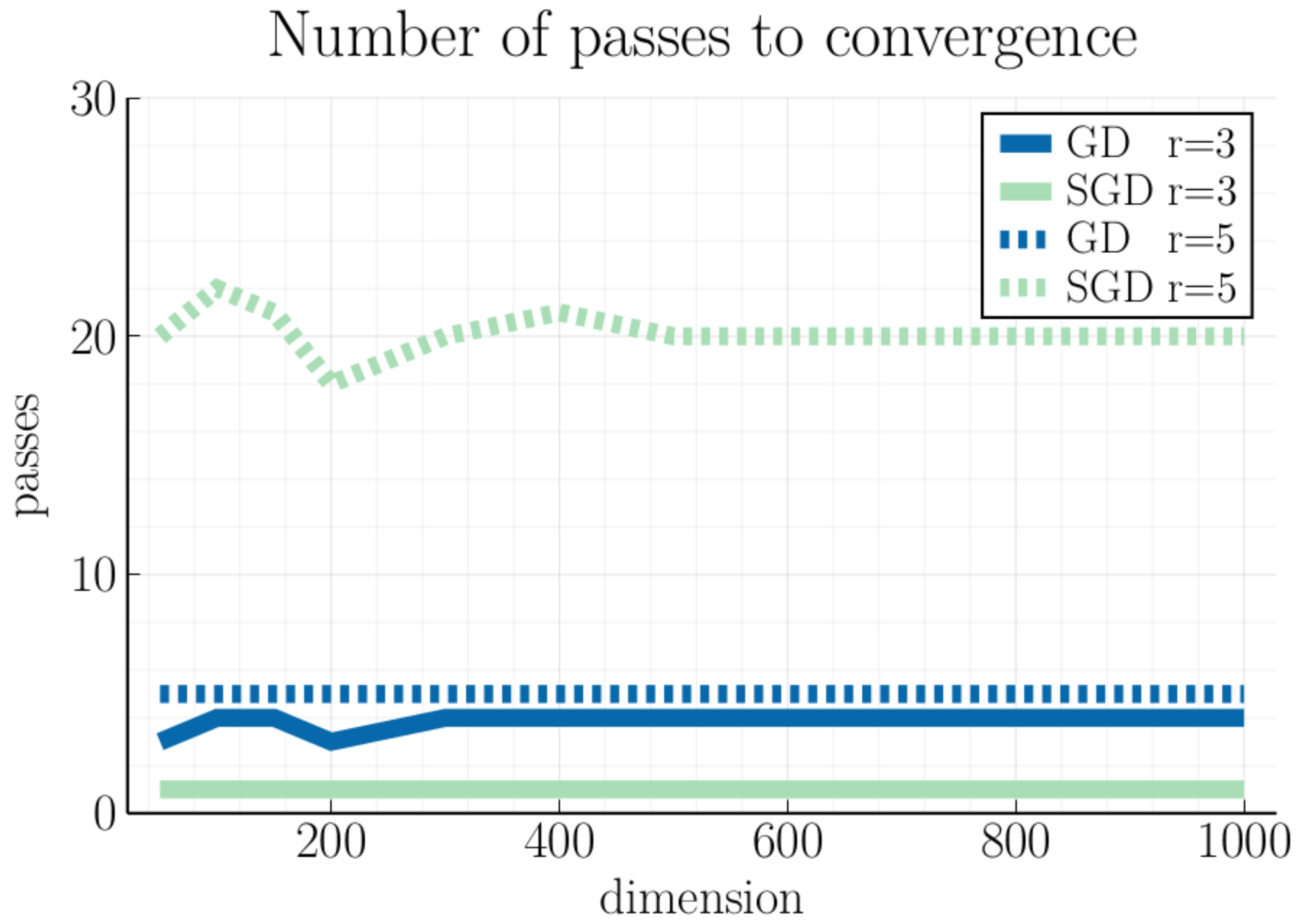
Dimension-free, global rates for SGD

$$S_t = \Sigma_t^{-1/2}(\Sigma_t^{1/2}X_t\Sigma_t^{1/2})^{1/2}\Sigma_t^{-1/2}$$

$$\Sigma_{t+1} = (I - \eta_t S_t)\Sigma_t(I - \eta_t S_t)$$

Theorem. (CMR**S**'20, ACG**S**'21)
Suppose P is supported on centered Gaussians with eigenvalues in the range $[\alpha, \beta]$. Then

$$\mathbb{E}[W_2^2(\Sigma_T, \Sigma_\star)] \leqslant \left(\frac{4\beta}{\alpha}\right)^{\frac{7}{2}} \cdot \frac{\sigma^2}{T}\,.$$

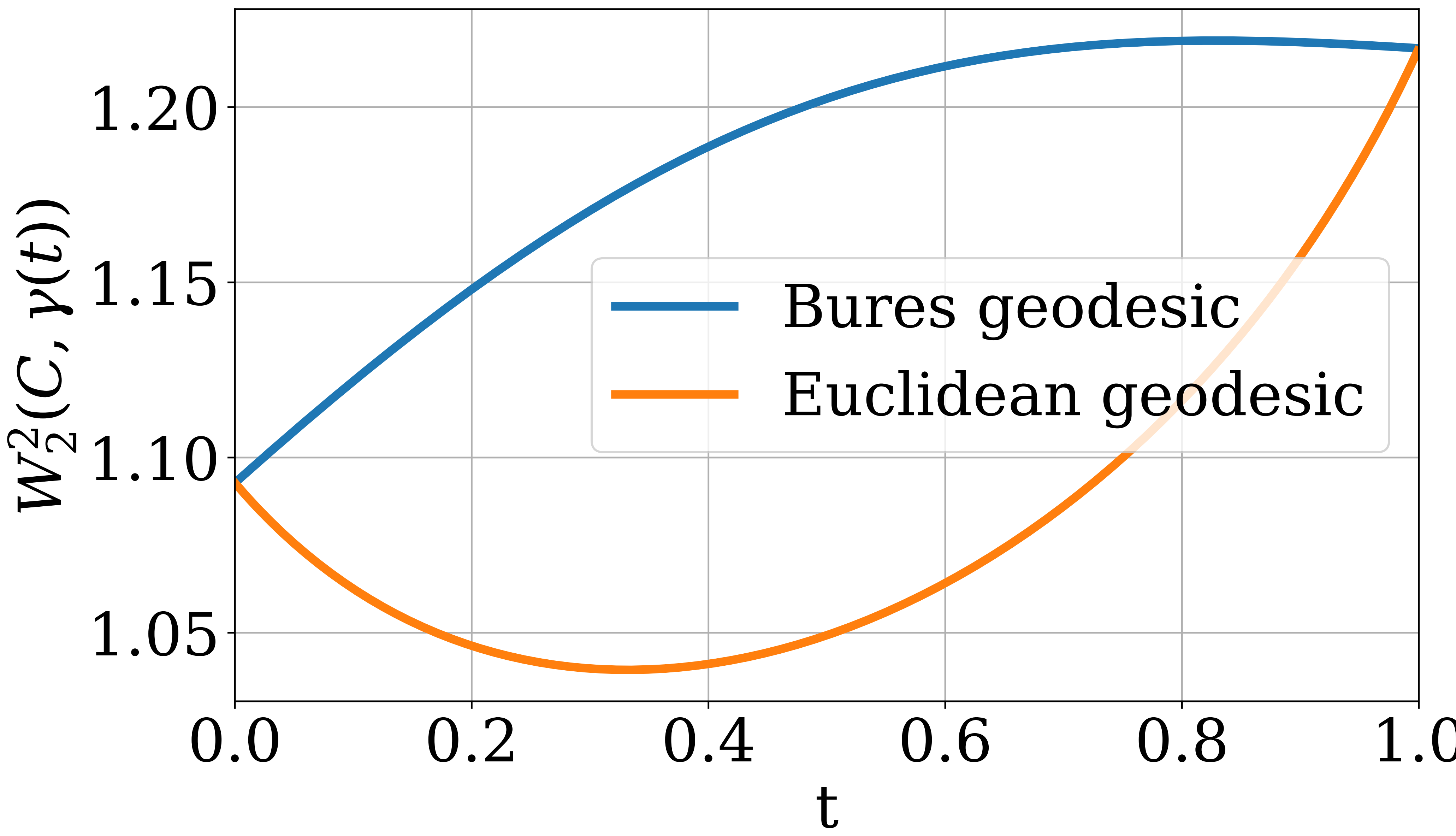


[Chewi, Maunu, Rigollet, Stromme '20,
Altschuler, Chewi, Gerber, Stromme '21]

Proof strategy

The NNC of Bures-Wasserstein space makes the barycenter functional non-convex, but is also automatically makes it **1-smooth**:

$$F_P(\Sigma_1) \leqslant F_P(\Sigma_0) + \langle \nabla F_P(\Sigma_1), \log_{\Sigma_0}(\Sigma_1) \rangle_{\Sigma_0} + \frac{1}{2} W_2^2(\Sigma_0, \Sigma_1)$$



It is known from convex optimization that under smoothness, strong convexity can be weakened to a quantitative condition known as a **Polyak-Łojasiewicz inequality**

[Otto, Villani ’00, Karimi, Nutini, Schmidt ’16]

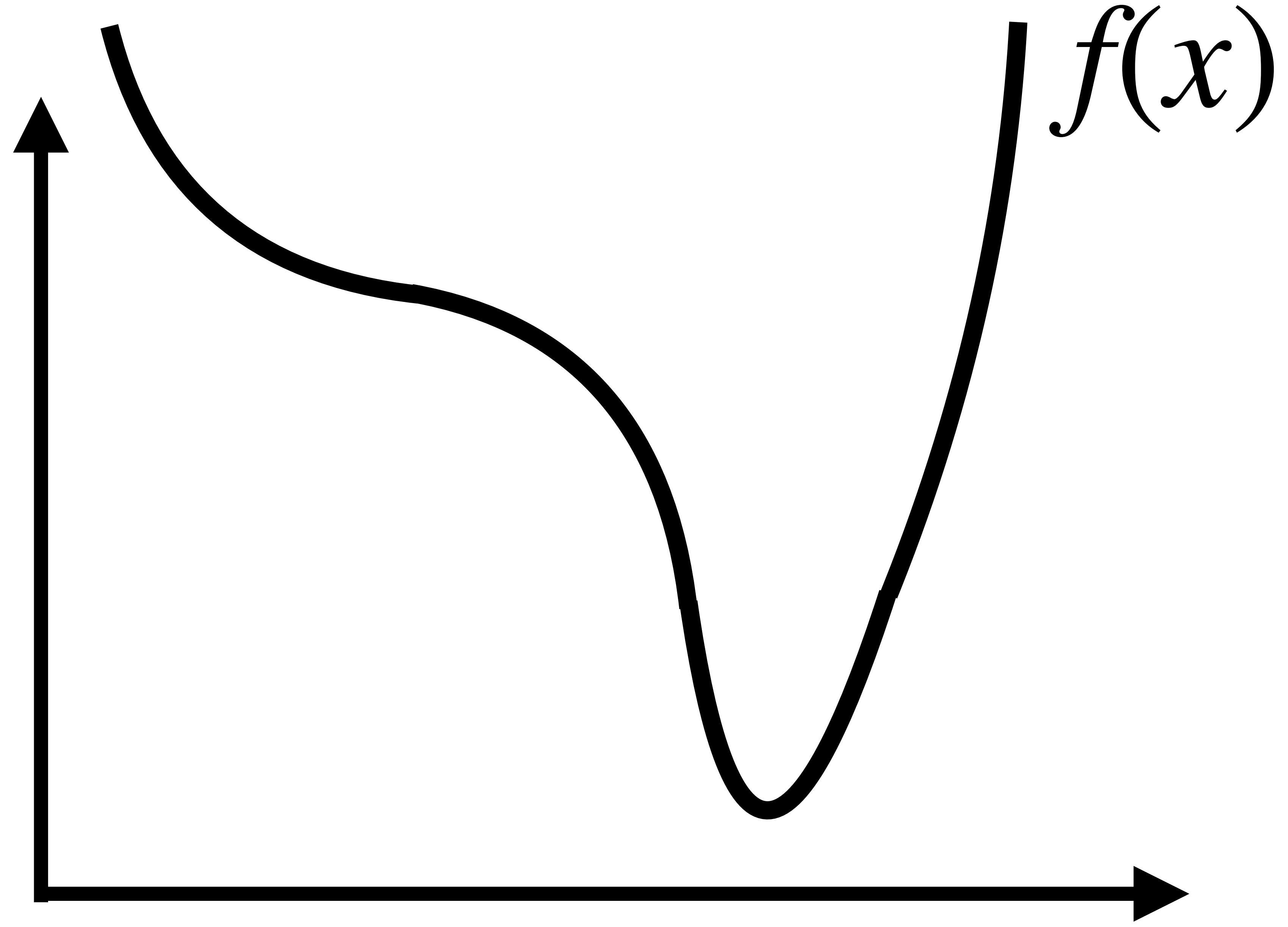
A Polyak-Łojasiewicz (PL) inequality

PL inequalities are a very useful tool to make the following statement quantitative:

“First-order critical points are global optima”

We say a function $f\colon X \to \mathbb{R}$ satisfies a PL inequality with constant C_{PL}

$$f(x) - \inf_{x\in X} f(x) \leqslant C_{\mathrm{PL}} \|\nabla_X f(x)\|_x^2$$



PL inequalities are a weak form of strong convexity that still imply similar optimization results

[Otto, Villani ’00, Karimi, Nutini, Schmidt ’16]

A Polyak-Łojasiewicz (PL) inequality

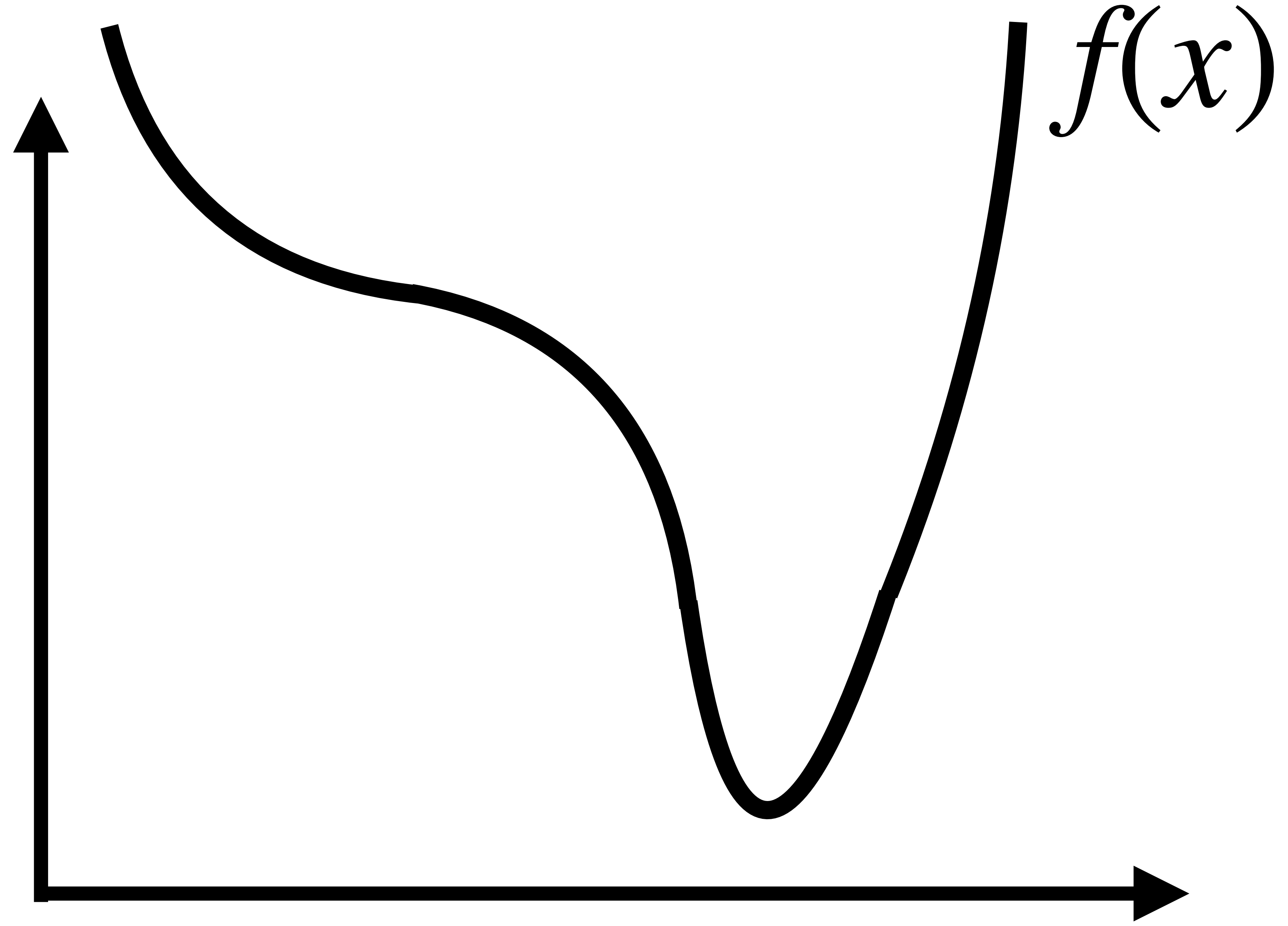
PL inequalities are a very useful tool to make the following statement quantitative:

“First-order critical points are global optima”

We say a function $f: X \to \mathbb{R}$ satisfies a PL inequality with constant C_{PL}

$$f(x) - \inf_{x \in X} f(x) \leqslant C_{\mathrm{PL}} \|\nabla_X f(x)\|_x^2$$

PL inequalities are a weak form of strong convexity that still imply similar optimization results



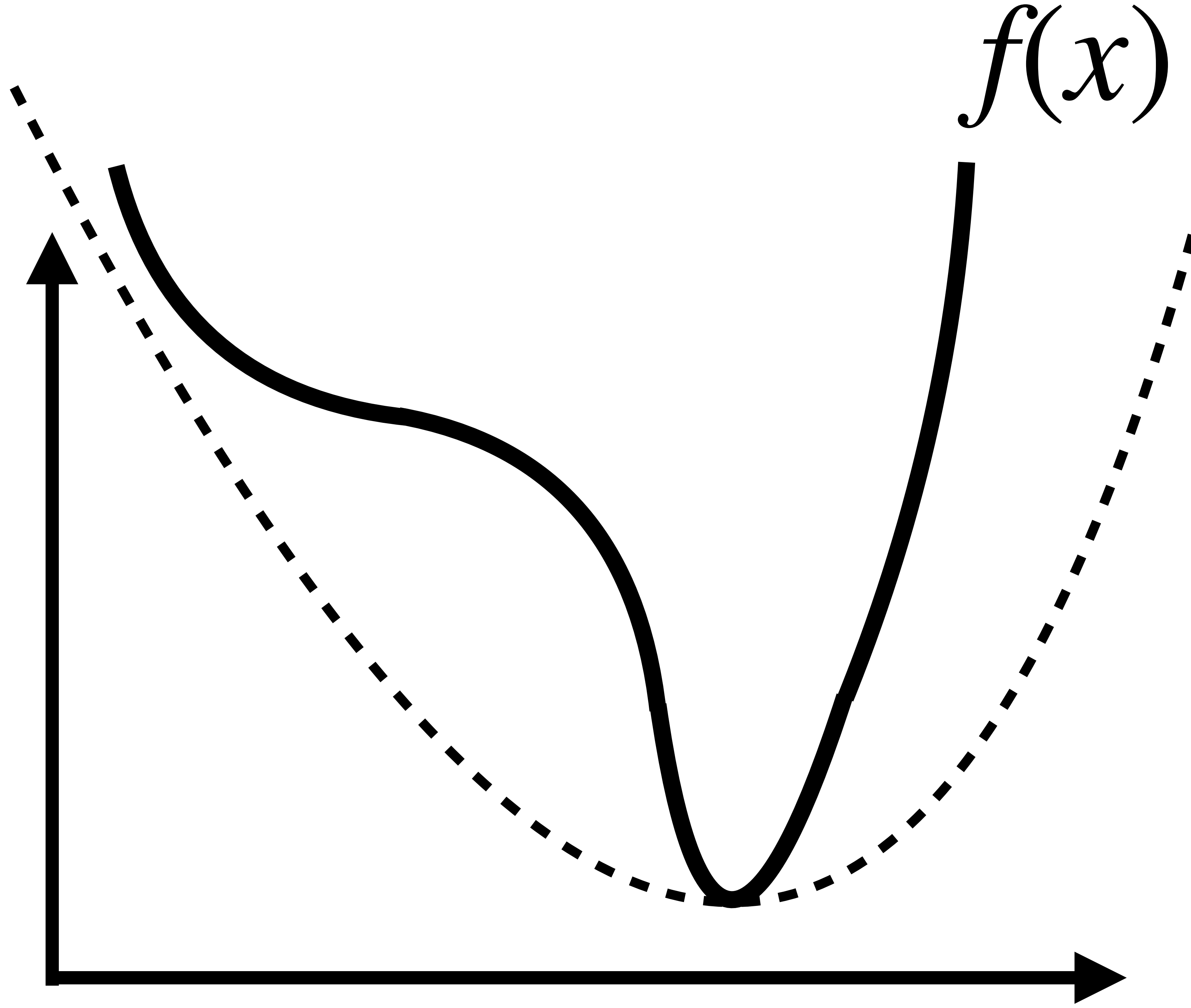
For the OT crowd: log-Sobolev is a PL inequality while displacement convexity is strong convexity

[Otto, Villani ’00, Karimi, Nutini, Schmidt ’16]

A variance (or quadratic growth) inequality

Suppose P is supported on centered Gaussians with eigenvalues in the range $[\alpha, \beta]$, and

$$F_P(\Sigma_\star) = \min_{\Sigma \succ 0} F_P(\Sigma) := \int W_2^2(\Sigma, \Sigma') \mathrm{d}P(\Sigma') \,.$$



Proposition. (CMR**S**'20) Then we have a **variance inequality** for all $\Sigma \succ 0$

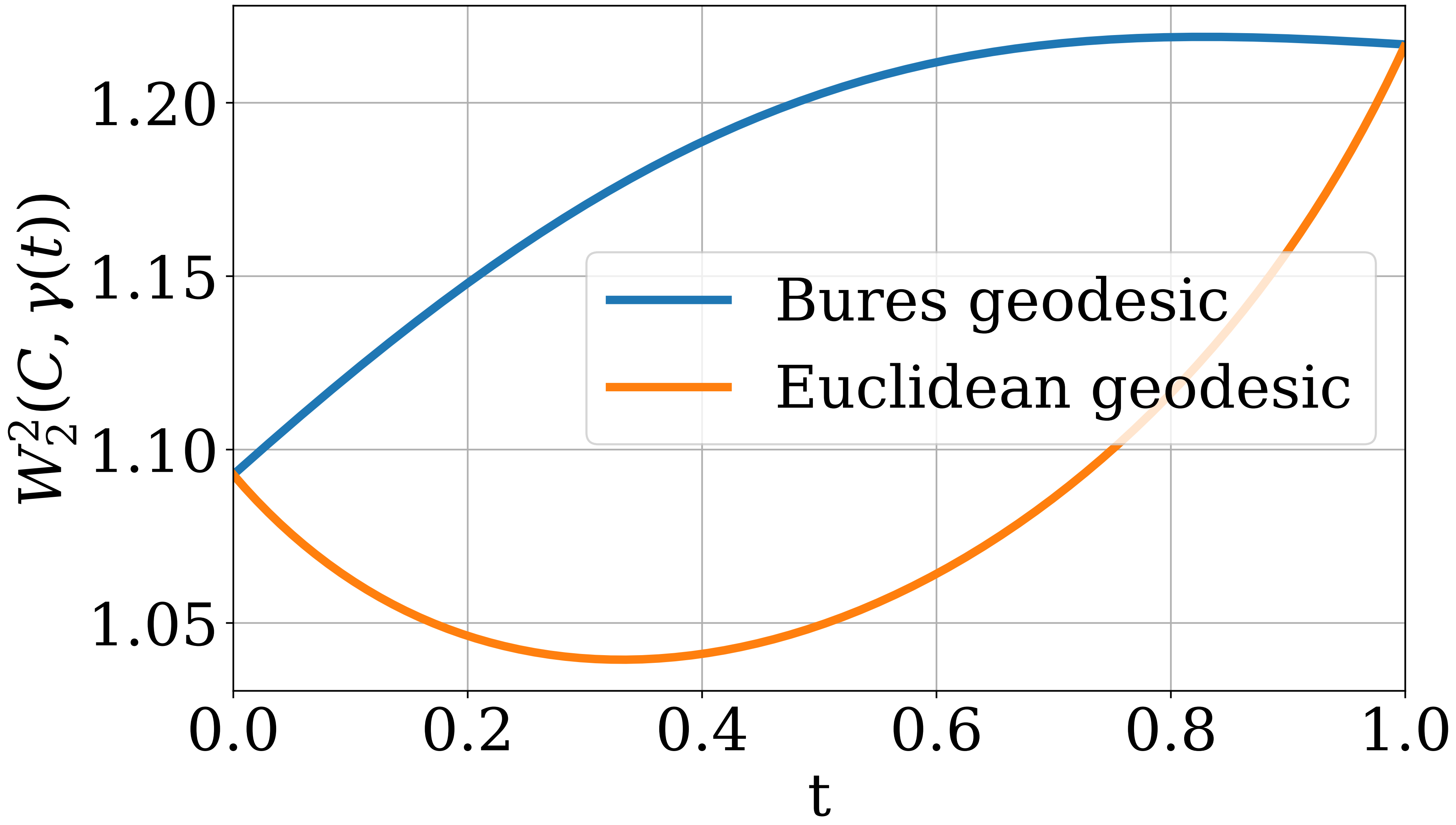
$$\frac{1}{2} W_2^2(\Sigma, \Sigma_\star) \leqslant \frac{\beta}{\alpha}(F_P(\Sigma) - F_P(\Sigma_\star))$$

[Otto, Villani '00, Sturm '03, Karimi, Nutini, Schmidt '16]

PL inequality for the BW barycenter functional

Suppose P is supported on centered Gaussians with eigenvalues in the range $[\alpha, \beta]$, and

$$F_P(\Sigma_\star) = \min_{\Sigma \succ 0} F_P(\Sigma) := \int W_2^2(\Sigma, \Sigma') \mathrm{d}P(\Sigma')\,.$$



Proposition. (CMR**S**'20) If Σ also has eigenvalues in the range $[\alpha, \beta]$ then

$$F_P(\Sigma) - F_P(\Sigma_\star) \leqslant 2\Big(\frac{\beta}{\alpha}\Big)^2 \|\nabla_{W_2} F_P(\Sigma)\|_\Sigma^2$$

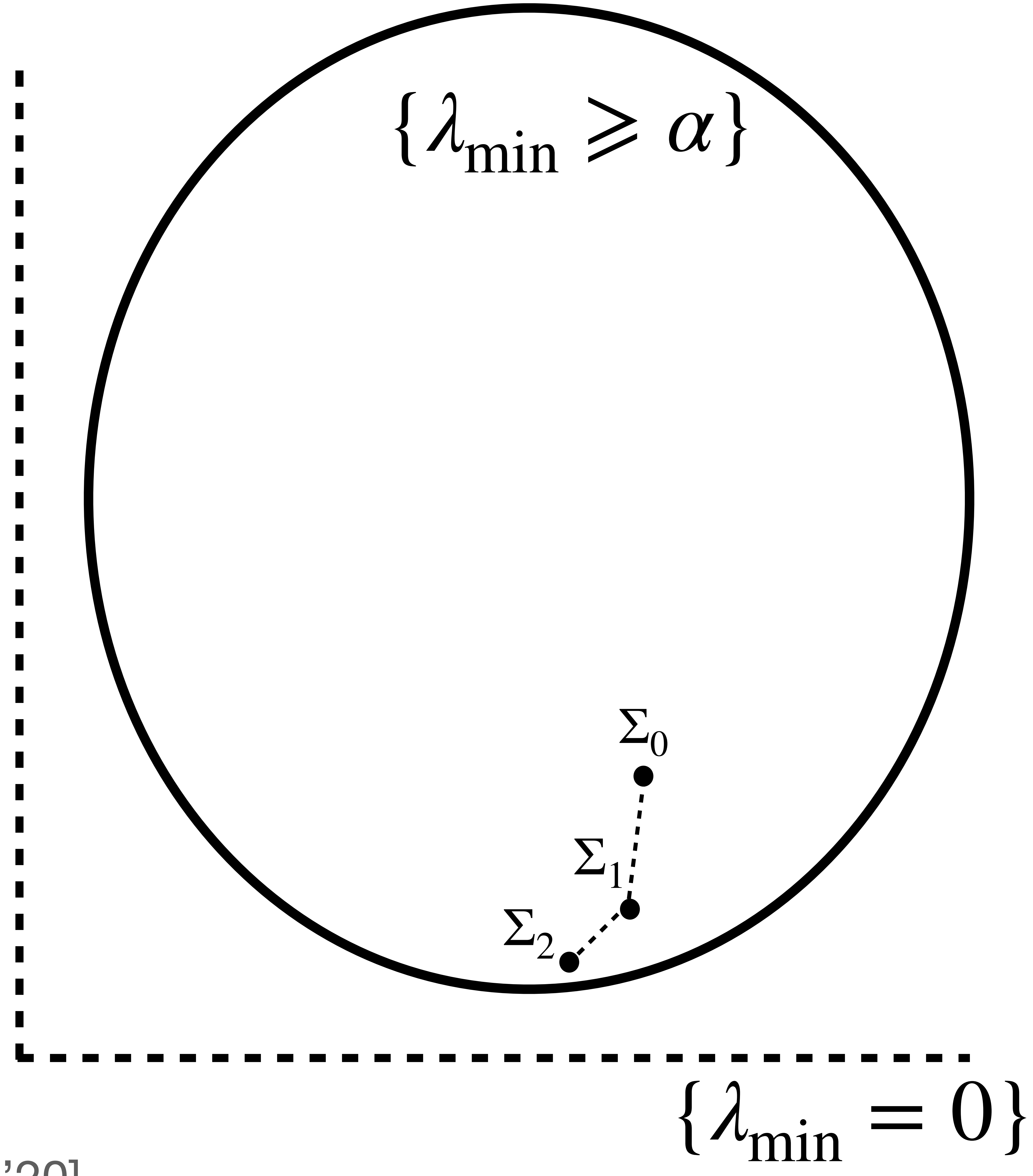
[Agueh and Carlier '11, Chewi, Maunu, Rigollet, Stromme '20]

Trapping iterates

Proposition. (ACG**S**'20) If Σ also has eigenvalues in the range $[\alpha, \beta]$ then

$$F_P(\Sigma) - F_P(\Sigma_\star) \leqslant 2\Big(\frac{\beta}{\alpha}\Big)^2 \|\nabla_{W_2} F_P(\Sigma)\|_\Sigma^2$$

Want this to hold along the optimization trajectory, **else the PL constant will blow up**



[Massart, Hendricx, and Absil '19, Chewi, Maunu, Rigollet, Stromme '20]

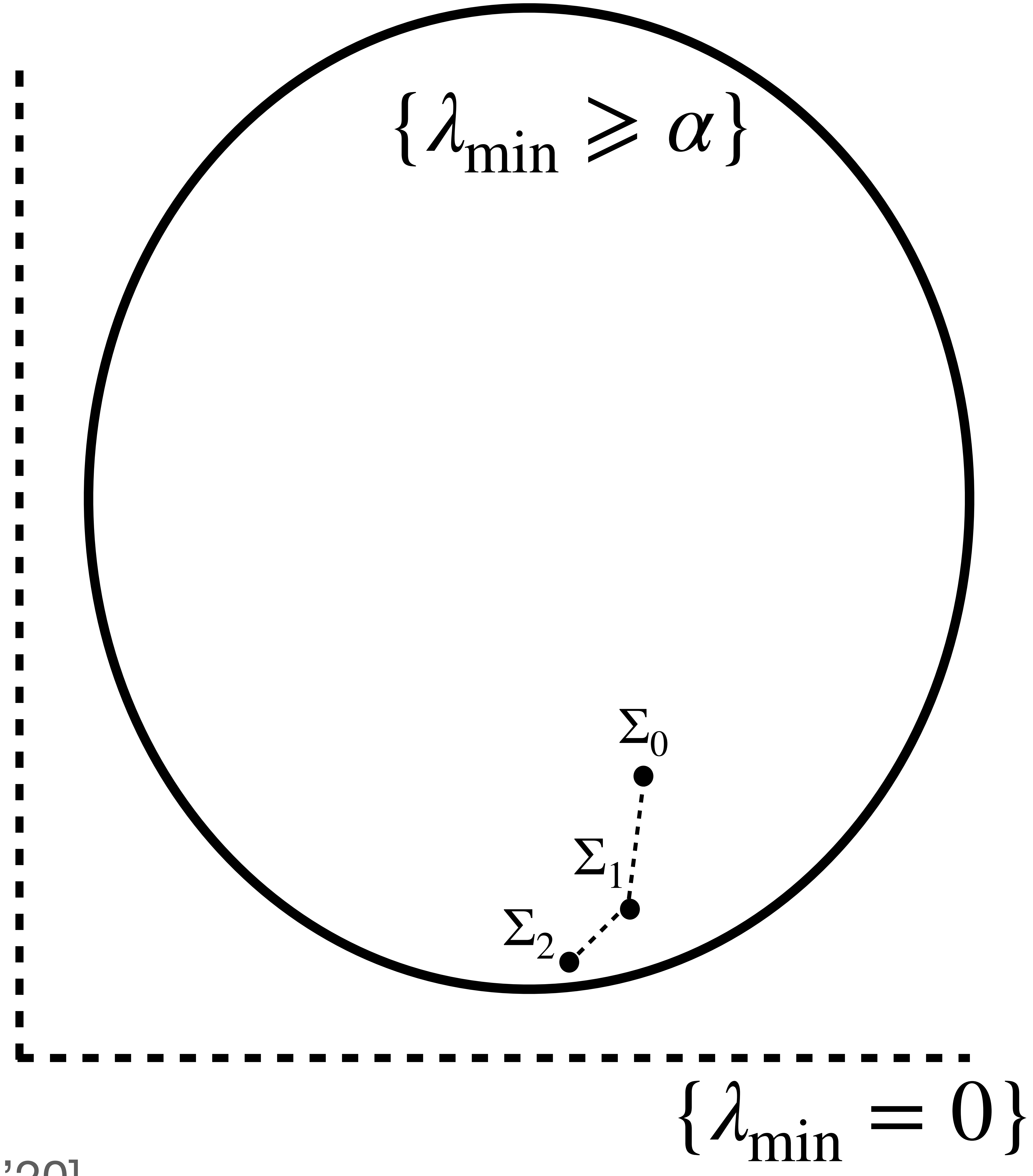
Trapping iterates

Proposition. (ACG**S**'20) If Σ also has eigenvalues in the range $[\alpha, \beta]$ then

$$F_P(\Sigma) - F_P(\Sigma_\star) \leqslant 2\left(\frac{\beta}{\alpha}\right)^2 \|\nabla_{W_2} F_P(\Sigma)\|_\Sigma^2$$

Want this to hold along the optimization trajectory, **else the PL constant will blow up**

Intuitively, we want to keep the iterates in a part of the manifold with bounded curvature



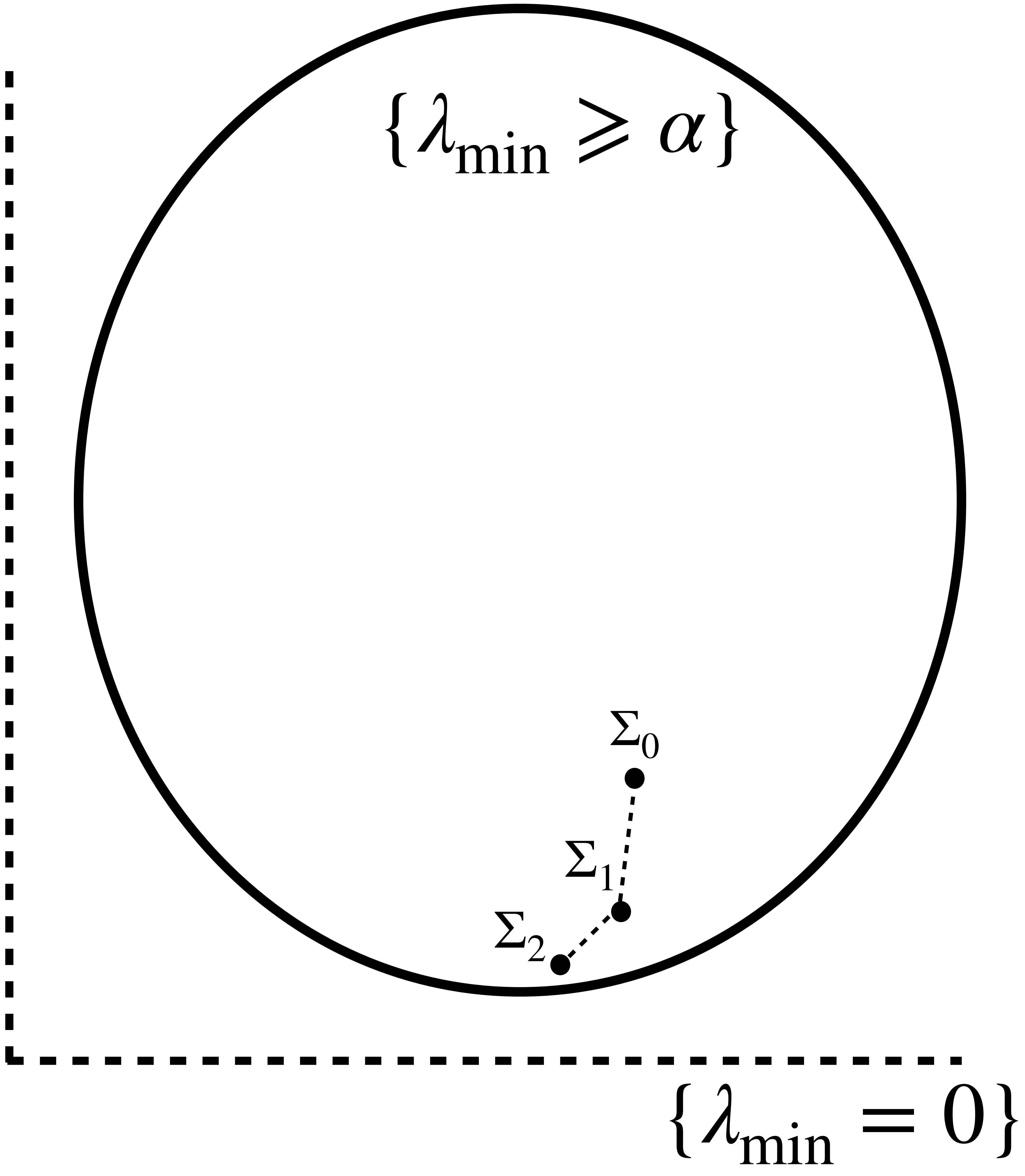
[Massart, Hendricx, and Absil '19, Chewi, Maunu, Rigollet, Stromme '20]

Trapping iterates: SGD

In fact, the functionals $-\sqrt{\lambda_{\min}}$ and $\sqrt{\lambda_{\max}}$ are geodesically convex

Enough to analyze SGD, since each new iteration moves along a geodesic to a point in $\operatorname{supp}(P)$

$$\Sigma_{t+1}^{\mathrm{SGD}} = \exp_{\Sigma_t^{\mathrm{SGD}}}(2\eta_t \log_{\Sigma_t^{\mathrm{SGD}}}(X_t))$$



[Agueh, Carlier '11, Bhatia, Jain, and Lim '19]

Trapping iterates: SGD

In fact, the functionals $-\sqrt{\lambda_{\min}}$ and $\sqrt{\lambda_{\max}}$ are geodesically convex

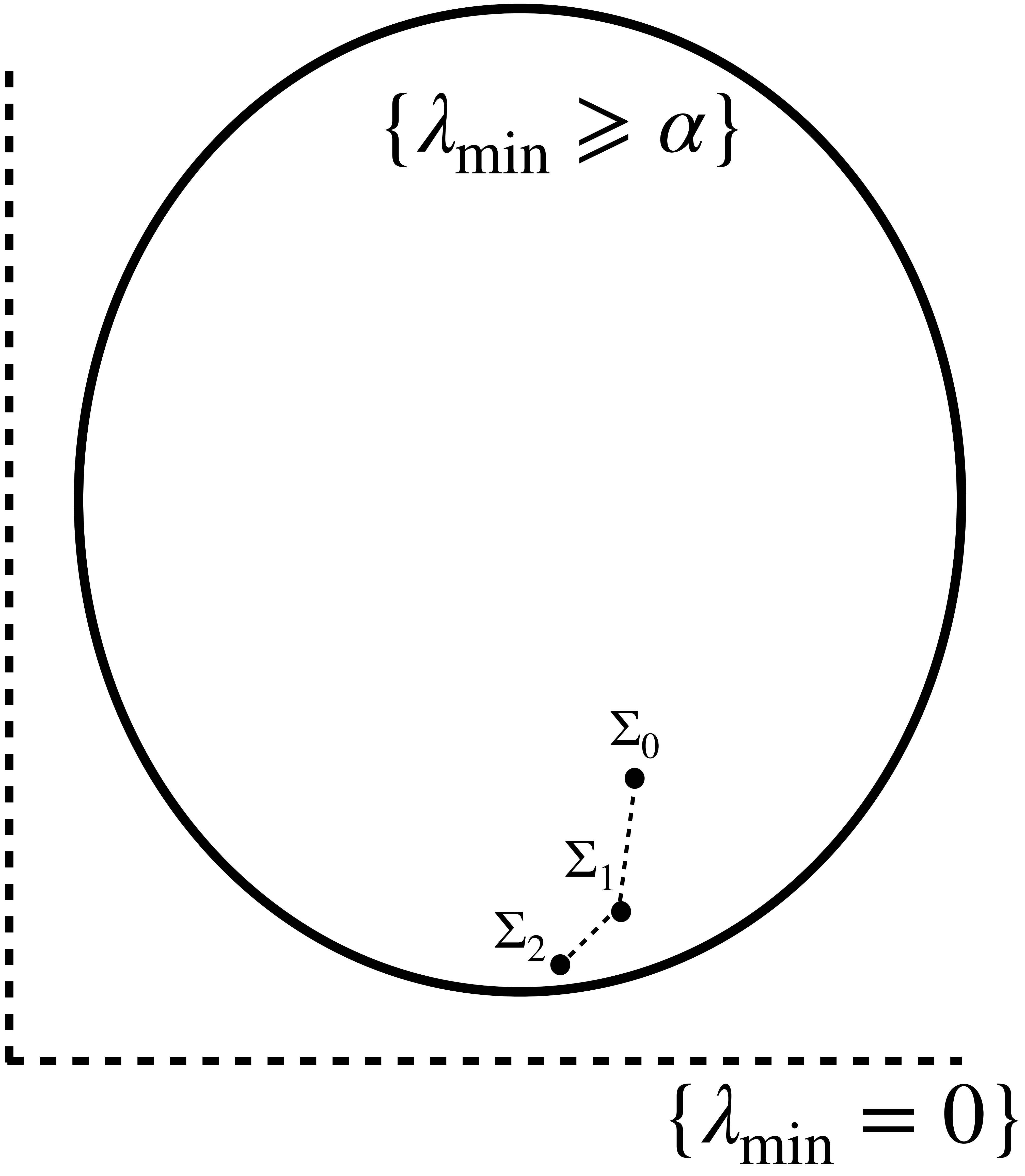
Enough to analyze SGD, since each new iteration moves along a geodesic to a point in $\mathrm{supp}(P)$

$$\Sigma_{t+1}^{\mathrm{SGD}} = \exp_{\Sigma_t^{\mathrm{SGD}}}(2\eta_t \log_{\Sigma_t^{\mathrm{SGD}}}(X_t))$$

However, this isn't enough for GD, since it moves along **generalized geodesics**:

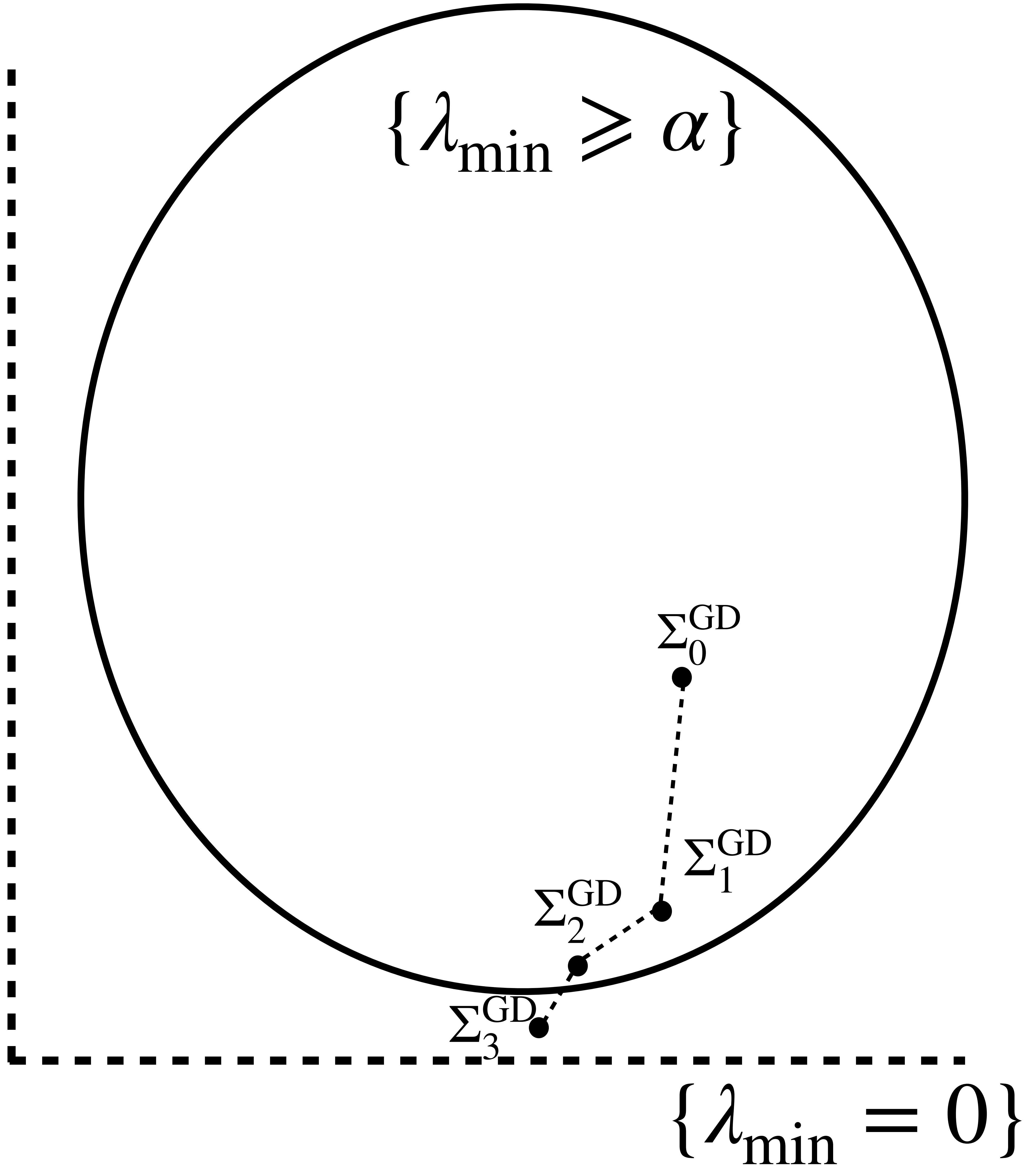
$$\Sigma_{t+1}^{\mathrm{GD}} = \exp_{\Sigma_t^{\mathrm{GD}}}\left(2\eta_t \int \log_{\Sigma_t^{\mathrm{GD}}}(X)\mathrm{d}P(x)\right)$$

[Agueh, Carlier '11, Bhatia, Jain, and Lim '19]



Trapping iterates: GD

Surprisingly, $-\sqrt{\lambda_{\min}}$ is **not** convex along generalized geodesics!



[Agueh, Carlier ’11, Altschuler, Chewi, Gerber, Stromme ’21]

Trapping iterates: GD

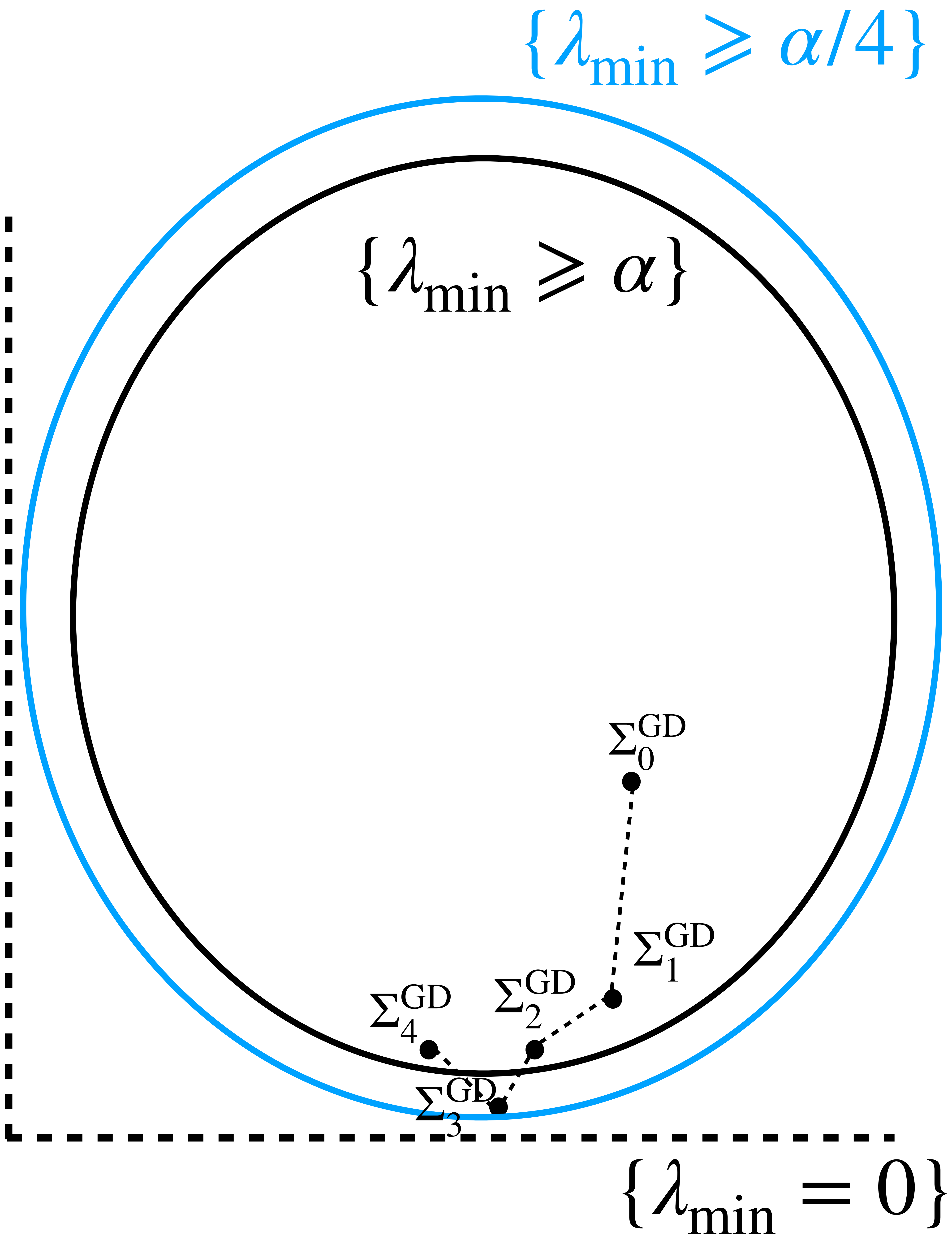
Surprisingly, $-\sqrt{\lambda_{\min}}$ is **not** convex along generalized geodesics!

We show this is an artifact of continuous vs. discrete time plus non-smoothness of $\sqrt{\lambda_{\min}}$

Ultimately show a weaker statement: for all times

$$\lambda_{\min}(\Sigma_t^{\mathrm{GD}}) \geqslant \alpha/4$$

[Agueh, Carlier ’11, Altschuler, Chewi, Gerber, Stromme ’21]



Open problems

(Ahidar-Coutrix, Le Gouic, Paris ’20): a geodesic $\gamma \colon [0,1] \to (X, d_X)$ is $(\lambda_{\text{in}}, \lambda_{\text{out}})$**-extendible** if there exists a constant speed extension $\tilde{\gamma} \colon [-\lambda_{\text{in}}, 1 + \lambda_{\text{out}}] \to (X, d_X)$ such that $\tilde{\gamma}|_{[0,1]} = \gamma$.

Suppose that $P \in \mathscr{P}_2((X, d_X))$ and $\lambda_{\text{in}}, \lambda_{\text{out}} > 0$. If P has a barycenter $b_\star$ such that for all $x \in \text{supp}(P)$, the geodesic $\gamma_{b_\star \to x}$ is $(\lambda_{\text{in}}, \lambda_{\text{out}})$-extendible, then does F_P obey a PL inequality with $C_{\text{PL}} = C_{\text{PL}}(\lambda_{\text{in}}, \lambda_{\text{out}})$?

Does this imply fast rates for the empirical barycenter?

[Ahidar-Coutrix, Le Gouic, Paris ’20, Le Gouic, Paris, Rigollet, Stromme ’22]