On optimal matching of Gaussian samples

M. Ledoux

University of Toulouse, France

& Institut Universitaire de France

optimal matching: minimal transportation cost

between two sets of (random) points

$X_1, \ldots, X_n, \quad Y_1, \ldots, Y_n$ independent random points in $\mathbb{R}^d$

$$\min_\sigma \frac{1}{n} \sum_{i=1}^{n} c(X_i, Y_{\sigma(i)})$$

σ permutation of $\{1, \ldots, n\}$

$c(\cdot, \cdot)$ cost function

$$c(x,y) = |x-y|^p, \quad 1 \leq p < \infty$$

transportation cost between an empirical and a reference measure

$X_1, \ldots, X_n, \quad Y_1, \ldots, Y_n$ independent random points in $\mathbb{R}^d$

$$\mathbb{E}\bigg(\min_\sigma \frac{1}{n} \sum_{i=1}^{n} |X_i - Y_{\sigma(i)}|^p \bigg)$$

σ permutation of $\{1, \ldots, n\}$

dependence

- dimension d
- $1 \leq p < \infty$ (mostly $p = 1$ and $p = 2$)
- common distribution of X_i, Y_i

$X_1, \ldots, X_n, \quad Y_1, \ldots, Y_n$ independent random points in $\mathbb{R}^d$

$$\mathbb{E}\left(\min_\sigma \frac{1}{n} \sum_{i=1}^{n} |X_i - Y_{\sigma(i)}|^p \right)$$

σ permutation of $\{1, \ldots, n\}$

X_i, Y_i uniform on $[0,1]^d$

typical distance between n uniform points in $[0,1]^d \quad \approx \frac{1}{n^{1/d}}$

expected: $$\mathbb{E}\left(\min_\sigma \frac{1}{n} \sum_{i=1}^{n} |X_i - Y_{\sigma(i)}|^p \right) \approx \frac{1}{n^{p/d}}$$

only true when $d \geq 3$

Ajtai-Komlós-Tusnády (1984)

$X_1, \dots, X_n, \quad Y_1, \dots, Y_n$ independent, uniform on $[0,1]^2$

$$\mathbb{E}\left(\min_\sigma \frac{1}{n} \sum_{i=1}^{n} |X_i - Y_{\sigma(i)}|^p \right) \approx \left(\frac{\log n}{n} \right)^{p/2}$$

$$1 \leq p < \infty$$

$$A \approx B \qquad \Longleftrightarrow \qquad \frac{1}{C} B \leq A \leq C B$$

combinatorial arguments (dyadic partitions)

Ajtai-Komlós-Tusnády (1984)

$X_1, \ldots, X_n, \quad Y_1, \ldots, Y_n$ independent, uniform on $[0,1]^2$

$$\mathbb{E}\bigg(\min_{\sigma} \frac{1}{n} \sum_{i=1}^{n} |X_i - Y_{\sigma(i)}|^p \bigg) \approx \bigg(\frac{\log n}{n} \bigg)^{p/2}$$

$$1 \leq p < \infty$$

alternate generic chaining ideas

Shor, Leighton (1989-91), Talagrand (1992-94)

Ajtai-Komlós-Tusnády (1984)

$X_1, \ldots, X_n, \quad Y_1, \ldots, Y_n$ independent, uniform on $[0,1]^2$

$$\mathbb{E}\bigg(\min_{\sigma} \frac{1}{n} \sum_{i=1}^{n} |X_i - Y_{\sigma(i)}|^p \bigg) \approx \bigg(\frac{\log n}{n} \bigg)^{p/2}$$

$$1 \leq p < \infty$$

other distributions?

Gaussian on $\mathbb{R}^2$ $\quad (p = 2)$

motivation

"Il y a les questions qui se posent
et les questions que l'on se pose"

Henri Poincaré

(incomplete) result

$X_1, \ldots, X_n, \quad Y_1, \ldots, Y_n$ independent, standard normal on $\mathbb{R}^2$

$$\frac{1}{C}\,\frac{\log n}{n} \le \mathbb{E}\left(\min_{\sigma} \frac{1}{n} \sum_{i=1}^{n} |X_i - Y_{\sigma(i)}|^2\right) \le C\,\frac{(\log n)^2}{n}$$

possible conjecture $\quad \dfrac{\log n}{n}$?

$X_1, \ldots, X_n, \quad Y_1, \ldots, Y_n$ independent random points in $\mathbb{R}^d$

$$\mathbb{E}\left(\min_\sigma \frac{1}{n} \sum_{i=1}^{n} |X_i - Y_{\sigma(i)}|^p \right)$$

σ permutation of $\{1, \ldots, n\}$

dependence

- dimension d
- $1 \leq p < \infty$ (mostly $p = 1$ and $p = 2$)
- common distribution of X_i, Y_i

$$c(x,y) = |x-y|^p, \quad 1 \leq p < \infty$$

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}, \quad \nu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{Y_i}$$

$$\min_{\sigma} \frac{1}{n}\sum_{i=1}^{n} |X_i - Y_{\sigma(i)}|^p = \mathrm{W}_p^p(\mu_n, \nu_n)$$

Kantorovich metric

$$\mathrm{W}_p^p(\mu,\nu) = \inf_{\pi} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |x-y|^p \, d\pi(x,y)$$

π with respective marginals μ and ν

what is the cost of optimal matching

$$\mathbb{E}\big(W_p^p(\mu_n, \nu_n)\big)$$

between two independent samples $X_1, \ldots, X_n, \quad Y_1, \ldots, Y_n$?

if the samples are iid with common law μ

what is the cost of optimal matching

$$\mathbb{E}\big(W_p^p(\mu_n, \mu)\big) \,?$$

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

$$p = 1$$

$$\mathrm{W}_1(\nu, \mu) \;=\; \sup \left(\int_{\mathbb{R}^d} \varphi \, d\nu - \int_{\mathbb{R}^d} \varphi \, d\mu \right)$$

$\varphi : \mathbb{R}^d \to \mathbb{R}$ $\quad 1$-Lipschitz

$$\mathrm{W}_1(\mu_n, \mu)) \;=\; \sup \left(\frac{1}{n} \sum_{i=1}^{n} \left[\varphi(X_i) - \mathbb{E}\big(\varphi(X_i)\big) \right] \right)$$

supremum of empirical processes

generic chaining

$$X_1, \ldots, X_n \quad \text{iid in } \mathbb{R}^d \text{ with law } \mu, \qquad \mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$$

first order study of

$$\mathbb{E}\big(\mathrm{W}_p(\mu_n, \mu)\big) \qquad \text{or} \qquad \mathbb{E}\big(\mathrm{W}_p^p(\mu_n, \mu)\big)$$

- dimension d
- mostly $p = 1$ and $p = 2$ $\qquad (\mathrm{W}_1 \leq \mathrm{W}_2)$
- distribution μ

upper bounds on $\mathbb{E}\big(\mathrm{W}_p^p(\mu_n, \mu)\big)$

standard rates:

comparison with the known uniform example

unusual or stronger rates for irregular distributions $(d \geq 3)$

specific representations of Kantorovich metrics

$$\mathrm{W}_1(\nu, \mu) = \int_{-\infty}^{+\infty} \big|G(x) - F(x)\big| dx$$

G, F distribution functions of ν, μ on $\mathbb{R}$

quantile representation for $\mathrm{W}_p(\nu, \mu), \quad p \geq 1$

$$\mathrm{W}_p^p(\nu, \mu) = \int_0^1 \big|G^{-1}(t) - F^{-1}(t)\big|^p dt$$

μ on $\mathbb{R}$ with distribution function F

$$\mathbb{E}\big(W_1(\mu_n, \mu)\big) \approx \frac{1}{\sqrt{n}}$$

if and only if

$$\int_{-\infty}^{+\infty} \sqrt{F(x)(1-F(x))}\, dx < \infty$$

(for example $\int_{\mathbb{R}} |x|^q \, d\mu < \infty, \quad q > 2$)

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}, \quad \nu_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{Y_i}$$

$$W_p^p(\mu_n, \nu_n) = \min_{\sigma}\frac{1}{n}\sum_{i=1}^{n}|X_i - Y_{\sigma(i)}|^p = \frac{1}{n}\sum_{i=1}^{n}|X_i^* - Y_i^*|^p$$

order statistics $\quad X_1^* \leq \cdots \leq X_n^*, \quad Y_1^* \leq \cdots \leq Y_n^*$

$$\mathbb{E}\big(W_p^p(\mu_n, \mu)\big) \approx \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\big(\big|X_i^* - \mathbb{E}(X_i^*)\big|^p\big)$$

$$1 \leq p < \infty$$

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n,\mu)\big) \;\approx\; \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\big(\big|X_i^* - \mathbb{E}(X_i^*)\big|^p\big)$$

μ uniform on $[0,1]$

X_i^* beta $(i, n-i+1)$

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n,\mu)\big) \;\approx\; \frac{1}{n^{p/2}}$$

$$p = 2$$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \;=\; \frac{6}{n}$$

bipartite $\quad \mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\nu_n)\big) \;=\; \frac{1}{3(n+1)}$

μ on $\mathbb{R}$ with distribution function F, density f

$$\mathbb{E}\big(W_2^2(\mu_n, \mu)\big) \approx \frac{1}{n}$$

if and only if

$$\int_{-\infty}^{+\infty} \frac{F(x)(1 - F(x))}{f(x)}\, dx < \infty$$

μ on $\mathbb{R}$ with distribution function F

$$\mathbb{E}\big(W_1(\mu_n, \mu)\big) \approx \frac{1}{\sqrt{n}}$$

if and only if

$$\int_{-\infty}^{+\infty} \sqrt{F(x)(1 - F(x))}\, dx < \infty$$

(for example $\int_{\mathbb{R}} |x|^q \, d\mu < \infty, \quad q > 2$)

μ on $\mathbb{R}$ with distribution function F, density f

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \approx \frac{1}{n}$$

if and only if

$$\int_{-\infty}^{+\infty} \frac{F(x)(1-F(x))}{f(x)}\,dx < \infty$$

$$\int_{-\infty}^{+\infty} \frac{F(x)(1-F(x))}{f(x)}\,dx = \int_0^1 \frac{t(1-t)}{I(t)^2}\,dt$$

$I(t) = f \circ F^{-1}(t)$ (isoperimetric profile)

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) \approx \frac{1}{n}$$

if and only if

$$\int_0^1 \frac{t(1-t)}{I(t)^2}\, dt < \infty$$

μ log-concave

accurate two-sided bounds

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) \approx \frac{1}{n} \int_{1/(n+1)}^{n/(n+1)} \frac{t(1-t)}{I(t)^2}\, dt$$

order statistics $X_1^* \leq \cdots \leq X_n^*$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) \approx \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}(X_i^*)$$

distribution of X_i^* log-concave, density

$$f_i(x) = n \binom{n-1}{i-1} F(x)^i \big(1 - F(x)\big)^{n-i} f(x), \qquad x \in \mathbb{R}$$

$$\mathrm{Var}(X_i^*) \approx \frac{1}{n} \, \frac{t(1-t)}{I(t)^2}, \quad t = \frac{i}{n}$$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) \approx \frac{1}{n} \int_{1/(n+1)}^{n/(n+1)} \frac{t(1-t)}{I(t)^2} \, dt$$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \approx \frac{1}{n}\int_{1/(n+1)}^{n/(n+1)} \frac{t(1-t)}{I(t)^2}\,dt$$

- μ standard normal, $\quad I(t) \approx t\sqrt{\log\frac{1}{t}}\,,\ t \to 0$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \approx \frac{\log\log n}{n}$$

- μ exponential, $\quad I(t) \approx t,\ t \to 0$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \approx \frac{\log n}{n}$$

general 1-d investigation Bobkov-L (2016)

μ uniform on $[0, 1]$

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n, \mu)\big) \approx \frac{1}{n^{p/2}} \qquad 1 \leq p < \infty$$

μ standard normal on $\mathbb{R}$

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n, \mu)\big) \approx \begin{cases} \frac{1}{n^{p/2}} & \text{if} \quad 1 \leq p < 2 \\ \frac{\log\log n}{n} & \text{if} \quad p = 2 \\ \frac{1}{n(\log n)^{p/2}} & \text{if} \quad p > 2 \end{cases}$$

W_2 more sensitive to distribution than W_1

μ uniform on $[0,1]^d$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \approx \frac{1}{n} \qquad d = 1$$

AKT $$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \approx \frac{\log n}{n} \qquad d = 2$$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \approx \frac{1}{n^{2/d}} \qquad d \geq 3$$

(of the order of the uniform spacings $\frac{1}{n^{1/d}}$)

Dereich-Scheutzow-Schottstedt (2013)

Fournier-Guillin (2015)

general μ, enough moments (dyadic partitions)

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) = O\left(\frac{1}{\sqrt{n}}\right) \qquad d = 2, 3$$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) = O\left(\frac{\log n}{\sqrt{n}}\right) \qquad d = 4$$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) = O\left(\frac{1}{n^{2/d}}\right) \qquad d \geq 5$$

some limit may exist

$$\lim_{n\to\infty} n^{1/d}\, \mathbb{E}\big(W_1(\mu_n,\mu)\big) \;=\; \gamma_d \qquad d \,\geq\, 3$$

$$\lim_{n\to\infty} n^{2/d}\, \mathbb{E}\big(W_2^2(\mu_n,\mu)\big) \;=\; \tau_d \qquad d \,\geq\, 5$$

modified subadditivity arguments

Dobric-Yukich (1995), Boutet de Monvel-Martin (2002) W_1

Barthe-Bordenave (2013),

Dereich-Scheutzow-Schottstedt (2013) $W_p, \quad p \,<\, \frac{d}{2}$

(stronger rates for singular distributions)

Ajtai-Komlós-Tusnády theorem

μ uniform on $[0,1]^2$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \approx \frac{\log n}{n}$$

general bounds

μ enough moments

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) = O\bigg(\frac{1}{\sqrt{n}}\bigg)$$

μ enough moments

$$\mathbb{E}\big(\mathrm{W}_1(\mu_n,\mu)\big) \,=\, O\left(\sqrt{\frac{\log n}{n}}\,\right)$$

(for example $\int_{\mathbb{R}^2} |x|^q \, d\mu \,<\, \infty, \quad q \,>\, 1$)

generic chaining methodology

Talagrand (1992), Yukich (1992)

Ajtai-Komlós-Tusnády theorem

μ uniform on $[0,1]^2$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \approx \frac{\log n}{n}$$

general bounds

μ enough moments

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) = O\Big(\frac{1}{\sqrt{n}}\Big)$$

μ Gaussian?

Ambrosio-Stra-Trevisan (2016)

μ uniform on $[0,1]^2$

$$\lim_{n\to\infty} \frac{n}{\log n}\, \mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) = \frac{1}{4\pi}$$

bipartite

$$\lim_{n\to\infty} \frac{n}{\log n}\, \mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \nu_n)\big) = \frac{1}{2\pi}$$

pde ansatz by Caracciolo, Lucibello, Parisi, Sicuro (2014)

Ambrosio-Stra-Trevisan (2016)

μ normalized Riemannian volume element

(M, g) compact Riemannian manifold, dimension 2

$$\lim_{n \to \infty} \frac{n}{\log n} \, \mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) \, = \, \frac{1}{4\pi}$$

2-sphere/torus: Holden-Peres-Zhai (2017) gravitational allocation upper bound

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) \;=\; O\bigg(\frac{\log n}{n}\bigg)$$

- regularization

- energy estimate

(Sobolev-type inequality, heat kernel estimates)

(M, g) compact Riemannian manifold, dimension d

(weighted manifold, Markov triple)

$p_t(x,y),\ t>0,\ x,y\in M$ heat kernel $(t=t(n)\to 0)$

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i} \ \to\ d\mu_n^t = \frac{1}{n}\sum_{i=1}^{n} p_t(X_i,\cdot)d\mu$$

standard convexity of W_2

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \le D_t + \mathbb{E}\big(\mathrm{W}_2^2(\mu_n^t,\mu)\big)$$

dispersion contribution

$$D_t = \int_M\int_M d(x,y)^2\, p_t(x,y)\, d\mu(x)d\mu(y), \quad t>0$$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n^t, \mu)\big)$$

density $\qquad d\mu_n^t \,=\, \frac{1}{n}\sum_{i=1}^{n} p_t(X_i, \cdot)d\mu$

central limit theorem heuristics

$$f \,= f(y) \,=\, \frac{1}{n}\sum_{i=1}^{n} p_t(X_i, y) \,\sim\, 1 + \frac{1}{\sqrt{n}}\,G(y)$$

infinitemisals

$$f = 1 + \varepsilon g, \quad \int_M g\, d\mu = 0$$

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2}\,\mathrm{W}_2^2(f\mu, \mu) \,=\, \int_M \big|\nabla((-\Delta)^{-1} g)\big|^2\, d\mu$$

new W_2 bound

dual Sobolev norm $\int_M g \, d\mu = 0$

$$\|g\|_{H^{-1}(\mu)} = \left(\int_M \left| \nabla ((-\Delta)^{-1} g) \right|^2 d\mu \right)^{1/2}$$

$$d\nu = f \, d\mu$$

$$W_2^2(\nu, \mu) \leq 4 \, \|f - 1\|_{H^{-1}(\mu)}^2$$

$$d\nu = f\, d\mu$$

$$W_2^2(\nu, \mu) \leq 4\, \|f - 1\|^2_{H^{-1}(\mu)}$$

$$\frac{1}{2}\, W_2(\nu, \mu)^2 = \sup \left(\int_M Q_1 \varphi\, d\nu - \int_M \varphi\, d\mu \right)$$

$\varphi : M \to \mathbb{R}$ bounded continuous

$$Q_u \varphi(x) = \inf_{y \in M} \left[\varphi(y) + \frac{d(x,y)^2}{2u} \right], \quad x \in M,\ u > 0$$

Hamilton-Jacobi $\quad \frac{d}{du}\, Q_u \varphi = -\frac{1}{2}\, |\nabla Q_u \varphi|^2$

$$\frac{1}{2}\,W_2(\nu,\mu)^2 \;=\; \sup\left(\int_M Q_1\varphi\,d\nu - \int_M \varphi\,d\mu\right)$$

$$\theta : [0,1] \;\to\; [0,1] \;\; \text{increasing}$$

$$\int_M Q_1\varphi\,d\nu - \int_M \varphi\,d\mu$$

$$= \int_0^1 \frac{d}{du}\int_M \big(1+\theta(u)g\big)Q_u\varphi\,d\mu\,du$$

$$= \int_0^1\int_M \Big[\theta'(u)g\,Q_u\varphi - \big(1+\theta(u)g\big)\frac{1}{2}|\nabla Q_u\varphi|^2\Big]d\mu\,du$$

$$= \int_0^1\int_M \Big[-\theta'(u)\nabla\big((-\Delta)^{-1}g\big)\cdot\nabla Q_u\varphi - \big(1+\theta(u)g\big)\frac{1}{2}|\nabla Q_u\varphi|^2\Big]d\mu\,du$$

$$\leq \int_0^1\int_M \frac{1}{2}\,\frac{\theta'(u)^2}{1+\theta(u)g}\,\big|\nabla\big((-\Delta)^{-1}g\big)\big|^2 d\mu\,du$$

dual Sobolev norm $\int_M g \, d\mu = 0$

$$\|g\|_{\mathrm{H}^{-1}(\mu)} = \left(\int_M \left|\nabla((-\Delta)^{-1} g)\right|^2 d\mu \right)^{1/2}$$

$$d\nu = f \, d\mu$$

$$\mathrm{W}_2^2(\nu, \mu) \leq 4 \, \|f - 1\|^2_{\mathrm{H}^{-1}(\mu)}$$

trace formula

$$(-\Delta)^{-1} = \int_0^\infty P_s \, ds$$

$$\int_M \left|\nabla((-\Delta)^{-1} g)\right|^2 d\mu = \int_M g(-\Delta)^{-1} g \, d\mu = 2 \int_0^\infty \int_M (P_s \, g)^2 d\mu \, ds$$

$$\mathrm{W}_2^2(\nu,\mu) \;\leq\; 8 \int_0^\infty \int_M (P_s\, g)^2 \, d\mu \, ds$$

density $\quad d\nu \;=\; f \, d\mu$

$$g \;=\; f - 1 \;=\; \frac{1}{n} \sum_{i=1}^{n} \big[p_t(X_i, y) - 1 \big]$$

variance

$$\mathbb{E}\big((P_s\, g)^2\big) \;=\; \frac{1}{n}\, \mathbb{E}\big(\big[p_{t+s}(X_1, y) - 1\big]^2\big) \;=\; \frac{1}{n} \big[p_{2(t+s)}(y,y) - 1\big]$$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n^t, \mu)\big) \;\leq\; \frac{4}{n} \int_{2t}^\infty \int_M \big[p_s(y,y) - 1\big] \, d\mu(y) \, ds$$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \;\leq\; D_t \,+\, \mathbb{E}\big(\mathrm{W}_2^2(\mu_n^t,\mu)\big), \quad t>0$$

dispersion

$$D_t \;=\; \int_M \int_M d(x,y)^2\, p_t(x,y)\, d\mu(x) d\mu(y)$$

energy (trace) estimate

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n^t,\mu)\big) \;\leq\; \frac{4}{n} \int_{2t}^{\infty} \int_M \big[p_s(y,y) - 1\big]\, d\mu(y)\, ds$$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \;\le\; D_t + \frac{4}{n}\int_{2t}^{\infty}\int_M \big[p_s(y,y)-1\big]\,d\mu(y)\,ds$$

(M,g) compact Riemannian manifold, dimension d

$$D_t \;=\; \int_M\int_M d(x,y)^2\, p_t(x,y)d\mu(x)d\mu(y) \;\le\; Ct$$

$$p_s(y,y) \;\le\; \frac{C}{s^{d/2}} \qquad 0<s\le 1$$

optimization $\quad t\sim\dfrac{1}{n^{2/d}}:\qquad \mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \;\le\; \begin{cases} \frac{C}{n} & \text{if } \; d=1 \\ \frac{C\log n}{n} & \text{if } \; d=2 \\ \frac{C}{n^{2/d}} & \text{if } \; d\ge 3 \end{cases}$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \;=\; \begin{cases} O\big(\frac{1}{n}\big) & \text{if} \quad d \;=\; 1 \\[2ex] O\Big(\frac{\log n}{n}\Big) & \text{if} \quad d \;=\; 2 \\[2ex] O\Big(\frac{1}{n^{2/d}}\Big) & \text{if} \quad d \;\geq\; 3 \end{cases}$$

heat kernel bounds $p_s(y,y) \;\leq\; \frac{C}{s^{d/2}}$ reflect the dimensional rates

similar (optimal) conclusions for $\mathbb{E}\big(\mathrm{W}_p^p(\mu_n,\mu)\big)$

Ambrosio-Stra-Trevisan (2016)

μ normalized Riemannian volume element

(M, g) compact Riemannian manifold, dimension 2

$$\lim_{n\to\infty} \frac{n}{\log n}\, \mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) = \frac{1}{4\pi}$$

more precise arguments

only for $p = 2, d = 2$

limit open for example for $\mathbb{E}\big(\mathrm{W}_1(\mu_n, \mu)\big)$

$X_1, \ldots, X_n$ independent

with standard Gaussian law μ in $\mathbb{R}^2$

order of

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) \, ?$$

pde-transportation technology

extended to the Gaussian setting

λ uniform on $[0,1]^2$

$U_1, \ldots, U_n$ independent with law λ, $\qquad \nu_n = \frac{1}{n}\sum_{i=1}^n \delta_{U_i}$

$$\lambda = \Phi^{\otimes 2}(\mu)$$

$$\Phi(x) = \int_{-\infty}^x e^{-u^2/2}\frac{du}{\sqrt{2\pi}}, \quad x \in \mathbb{R}, \qquad \|\Phi\|_{\mathrm{Lip}} \leq 1$$

$$\mathbb{E}\big(W_2^2(\mu_n, \mu)\big) \geq \mathbb{E}\big(W_2^2(\nu_n, \lambda)\big)$$

AKT $\qquad \mathbb{E}\big(W_2^2(\nu_n, \lambda)\big) \approx \dfrac{\log n}{n}$

μ standard Gaussian on $\mathbb{R}^d$

pde-transportation approach

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) \leq D_t + \frac{4}{n} \int_{2t}^{\infty} \int_M \big[p_s(y,y) - 1\big] \, d\mu(y) \, ds$$

$p_t(x,y), \quad t > 0,\ x, y \in \mathbb{R}^d$ Mehler kernel

$$p_t(x,y) = \frac{1}{(1-e^{-2t})^{d/2}} \exp\bigg(-\frac{e^{-2t}}{1-e^{-2t}} \big[|x|^2 + |y|^2 - 2\, e^t \, x \cdot y\big]\bigg)$$

no uniform bounds

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \;\leq\; D_t + \frac{4}{n}\int_{2t}^{\infty}\int_M \big[p_s(y,y) - 1\big]\, d\mu(y)\, ds$$

$$D_t \;=\; \int_M\int_M |x-y|^2\, p_t(x,y)\, d\mu(x) d\mu(y) \;\leq\; 2dt$$

$$\int_M p_s(y,y)\, d\mu(y) \;=\; \frac{1}{(1-e^{-s})^d}$$

optimization in t

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \;=\; O\bigg(\frac{1}{n^{1/d}}\bigg) \qquad \text{if} \quad d \geq 2$$

μ^R normalized restriction of μ to the ball $B = B(0, R)$, $R \sim \sqrt{\log n}$

$Z_1, \ldots, Z_n$ independent with law μ^R

$$X_i^R = \begin{cases} X_i & \text{if} \quad |X_i| \leq R \\ Z_i & \text{if} \quad |X_i| > R \end{cases}$$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu_n^R)\big) = O\Big(\frac{1}{n}\Big), \qquad \mu_n^R = \frac{1}{n} \sum_{i=1}^n \delta_{X_i^R}$$

$$\int_M p_s(y, y)\, d\mu^R(y) = \frac{1}{(1 - e^{-s})^{d/2}} \, \frac{\mu(\theta B)}{\theta^d}$$

$$\theta \sim s^{1/2} \quad \text{as} \quad s \to 0$$

$$\frac{\mu(\theta B)}{\theta^d} \sim \lambda(B) \sim R^d \quad \text{as} \quad \theta R \leq 1$$

μ standard Gaussian on $\mathbb{R}^d$

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) = \begin{cases} O\Big(\frac{\log\log n}{n}\Big) & \text{if} \quad d = 1 \\ O\Big(\frac{(\log n)^2}{n}\Big) & \text{if} \quad d = 2 \\ O\Big(\frac{\log n}{n^{2/d}}\Big) & \text{if} \quad d \geq 3 \end{cases}$$

good enough to cover $d = 1$

(?) unnecessary extra factor $R^2 = \log n$ for $d \geq 2$

μ standard Gaussian on $\mathbb{R}^2$

$$\frac{1}{C}\,\frac{\log n}{n} \;\leq\; \mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \;\leq\; C\,\frac{(\log n)^2}{n}$$

plausible conjecture

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \;\approx\; \frac{\log n}{n}$$

simulation Stra (2016)

$$\lim_{n\to\infty}\frac{n}{\log n}\,\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \;=\; \frac{1}{5} \quad (??)$$

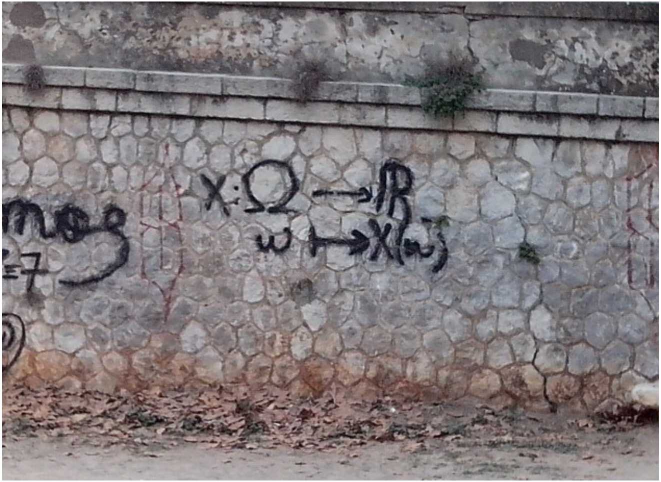
$1 \leq p < 2$

same pde-transportation methodology

μ standard Gaussian on $\mathbb{R}^d$

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n, \mu)\big) \approx \begin{cases} \frac{1}{n^{p/2}} & \text{if } d = 1 \\ \left(\frac{\log n}{n}\right)^{p/2} & \text{if } d = 2 \end{cases}$$

(same as for uniform)



Thank you for your attention

Vershik (2013) wrote a historic essay explaining why it is more fair to fix the name "Kantorovich distance" for all metrics like W_p (calling them Kantorovich power metrics).

Some general topological properties of W_1 were studied in 1970 by Dobrushin, who re-introduced this metric with reference to [Vasershtein 1969] ; apparently, that is why the name "Wasserstein distance" has become rather traditional.

As Vershik writes, "Leonid Vasershtein is a famous mathematician specializing in algebraic K-theory and other areas of algebra and analysis, and ... he is absolutely not guilty of this distortion of terminology, which occurs primarily in Western literature".