# A Fourier-analytic approach to transport inequalities

Sergey Bobkov
University of Minnesota

(Joint work with M. Ledoux)

**Worldwide Asymptotic Geometric Analysis Seminar**
**online 11 April 2020**

# Definitions

Definition. Let $\mu$ and $\nu$ be probability measures on a separable metric space $(M, \rho)$. The Kantorovich transport distance:

$$W(\mu, \nu) \ = \ \inf_{\lambda} \iint \rho(x, y)\, d\lambda(x, y),$$

where the inf is over all $\lambda$ on $M \times M$ with marginals $\mu$ and $\nu$.

Discrete measures with atoms $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$:

$$\mu_n = \frac{1}{n}(\delta_{x_1} + \cdots + \delta_{x_n}), \qquad \nu_n = \frac{1}{n}(\delta_{y_1} + \cdots + \delta_{y_n}),$$

$$W(\mu_n, \nu_n) \ = \ \frac{1}{n} \ \inf_{\sigma \in S_n} \sum_{k=1}^{n} \rho(x_k, y_{\sigma(k)}).$$

The case $M = \mathbb{R}$ with $\rho(x, y) = |x - y|$. Arranging $x_1^* \leq \cdots \leq x_n^*$, $y_1^* \leq \cdots \leq y_n^*$, one has

$$W(\mu_n, \nu_n) \ = \ \frac{1}{n} \ \sum_{k=1}^{n} |x_k^* - y_k^*|.$$

# Random samples and empirical measures

Given i.i.d. $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ (independent samples) distributed according to the law $\mu$ on $M$, let

$$\mu_n = \frac{1}{n}(\delta_{X_1} + \cdots + \delta_{X_n}), \qquad \nu_n = \frac{1}{n}(\delta_{Y_1} + \cdots + \delta_{Y_n}),$$

so that

$$W(\mu_n, \nu_n) = \frac{1}{n} \inf_{\sigma \in S_n} \sum_{k=1}^{n} \rho(X_k, Y_{\sigma(k)}).$$

Theorem (Varadarajan 1958). With probability one, $\mu_n \Rightarrow \mu$ as $n \to \infty$.

Variant: If $\int \rho(x, x_0)\, d\mu(x) < \infty$, then $W(\mu_n, \mu) \to 0$ a.s.

Matching problem: $\mathbb{E}\, W(\mu_n, \mu) = ?$ or $\mathbb{E}\, W(\mu_n, \nu_n) = ?$

The case $M = \mathbb{R}$: In terms of distribution funtions of $\mu_n$ and $\mu$,

$$W(\mu_n, \mu) = \int_{-\infty}^{\infty} |F_n(x) - F(x)|\, dx,$$

$$\mathbb{E}\, W(\mu_n, \mu) \le \frac{1}{\sqrt{n}} J, \qquad J = \int_{-\infty}^{\infty} \sqrt{F(x)(1 - F(x))}\, dx.$$

Theorem (B-L). $\mathbb{E}\, W(\mu_n, \mu) = O(\frac{1}{\sqrt{n}})$ if and only if $J < \infty$.

# Sampling from the uniform distribution

Let $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ be drawn from $M = [0,1]^d$ with Euclidean distance according to the uniform distribution, and let

$$\mu_n = \frac{1}{n}(\delta_{X_1} + \cdots + \delta_{X_n}), \qquad \nu_n = \frac{1}{n}(\delta_{Y_1} + \cdots + \delta_{Y_n}).$$

Case $d = 1$: $\mathbb{E}\, W(\mu_n, \nu_n) \sim \frac{1}{\sqrt{n}}$.

Case $d = 2$: AKT theorem (Ajtai-Komlós-Tusnády 1984)

$$\mathbb{E}\, W(\mu_n, \nu_n) \sim \sqrt{\frac{\log n}{n}}.$$

Other proofs: Shor (mid 1980's), Talagrand-Yukich (1993)
Extention to arbitrary laws $\mu$ on $[0,1]^d$: Talagrand (1992)

Majorizing measure approach: Talagrand (1994)
Fourier-analytic approach: B-Ledoux (2019)
PDE approach, exact asymptotics for $\mathbb{E}\, W_2^2(\mu_n, \nu_n)$: Ambrosio, Stra, Trevisan (2019)

Case $d \geq 3$: $\mathbb{E}\, W(\mu_n, \nu_n) \sim n^{-1/d}$.

# Kantorovich-Rubinstein theorem

Theorem. Given $\mu$ and $\nu$ on $(M, \rho)$,

$$W(\mu, \nu) = \sup_u \Big| \int u \, d\mu - \int u \, d\nu \Big|$$

where the sup is over all Lipschitz $u$ on $M$ with $\|u\|_{\mathrm{Lip}} \le 1$.

For empirical measures

$$\mu_n = \frac{1}{n}(\delta_{X_1} + \cdots + \delta_{X_n}), \qquad \nu_n = \frac{1}{n}(\delta_{Y_1} + \cdots + \delta_{Y_n}),$$

for independent samples drawn from $\mu$ on $M = [0,1]^d$,

$$nW(\mu_n, \nu_n) \; = \; \sup_u \Big[ \sum_{k=1}^n u(X_k) - \sum_{k=1}^n u(Y_k) \Big]$$

$$n \, \mathbb{E} \, W(\mu_n, \nu_n) \; \sim \; \mathbb{E} \sup_u \Big[ \sum_{k=1}^n \varepsilon_k u(X_k) \Big]$$

Subgaussian process (in Talagrand's proof): $\xi(u) = \sum_{k=1}^n \varepsilon_k u(X_k)$.

# Fourier-Stieltjes transform

Any probability measure $\mu$ on $Q^d = (-\pi, \pi]^d$ is uniquely determined by the sequence (characteristic function)

$$f_\mu(m) = \int_{Q^d} e^{im\cdot x}\, d\mu(x), \qquad m = (m_1, \dots, m_d) \in \mathbb{Z}^d.$$

Theorem 1 (Fourier-analytic inequality). If $\mu, \nu$ are supported on $[0, \pi]^d$, for any $t > 0$,

$$W(\mu,\nu) \le \left( \sum_{m \neq 0} e^{-t|m|^2} \frac{|f_\mu(m) - f_\nu(m)|^2}{|m|^2} \right)^{1/2} + 2\sqrt{dt}.$$

Other variants (smoothings): For any $T > 0$,

$$W(\mu,\nu) \le \left( \sum_{1 \le \|m\|_\infty \le T} \frac{|f_\mu(m) - f_\nu(m)|^2}{|m|^2} \right)^{1/2} + \frac{c\sqrt{d}}{T}.$$

# Application to empirical measures

Let

$$\mu_n = \frac{1}{n}(\delta_{X_1} + \cdots + \delta_{X_n}), \qquad \nu_n = \frac{1}{n}(\delta_{Y_1} + \cdots + \delta_{Y_n}).$$

Corollary 1. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be pairwise independent with values in $[0,1]^d \times [0,1]^d$ such that $X_k$ and $Y_k$ have equal distributions $\forall\, k \geq 1$. Then

$$\mathbb{E}\, W(\mu_n, \nu_n) = \begin{cases} O\left(\frac{1}{\sqrt{n}}\right) & \text{if} \quad d = 1, \\ O\left(\sqrt{\frac{\log n}{n}}\right) & \text{if} \quad d = 2, \\ O\left(\frac{1}{n^{1/d}}\right) & \text{if} \quad d \geq 3. \end{cases}$$

Proof. By Fourier-analytic inequality,

$$\mathbb{E}\, W(\mu_n, \nu_n) \leq \left( \sum_{m \neq 0} e^{-t|m|^2}\, \mathbb{E}\, \frac{|f_{\mu_n}(m) - f_{\nu_n}(m)|^2}{|m|^2} \right)^{1/2} + 2\sqrt{d t}.$$

Here

$$f_{\mu_n}(m) - f_{\nu_n}(m) = \frac{1}{n} \sum_{k=1}^{n} \left( e^{im \cdot X_k} - e^{im \cdot Y_k} \right),$$

$$\mathbb{E}\, |f_{\mu_n}(m) - f_{\nu_n}(m)|^2 \leq \frac{4}{n}.$$

# Reduction to torus

Kantorovich-Rubinstein theorem:

$$W(\mu,\nu) \;=\; \sup_{\|u\|_{\mathrm{Lip}}\le 1} \left| \int_{Q^d} u\,d\mu - \int_{Q^d} u\,d\nu \right|.$$

Let $\mathcal{L}_d$ be the set of all $(2\pi)$-periodic functions $u$ on $\mathbb{R}^d$ with $\|u\|_{\mathrm{Lip}} \le 1$. Put

$$\widetilde{W}(\mu,\nu) \;=\; \sup_{u\in\mathcal{L}_d} \left| \int_{Q^d} u\,d\mu - \int_{Q^d} u\,d\nu \right|.$$

Lemma 1. If $\mu$ and $\nu$ are supported on $[0,\pi]^d$, then $W(\mu,\nu) = \widetilde{W}(\mu,\nu)$.

Proof. Let $\|z\| = \mathrm{dist}(z, 2\pi\mathbb{Z}^d)$, $z \in \mathbb{R}^d$. Then $u \in \mathcal{L}_d$ iff $u$ is $1$-Lip with respect to $\rho(x,y) = \|x-y\|$. Since it is a metric on $Q^d$, by the Kantorovich-Rubinstein theorem,

$$\widetilde{W}(\mu,\nu) \;=\; \inf_{\lambda} \iint \|x-y\|\,d\lambda(x,y).$$

If $\mu,\nu$ are supported on $[0,\pi]^d$, then $\lambda$ is supported on $[0,\pi]^d \times [0,\pi]^d$, while

$$\|x-y\| = |x-y|, \qquad x,y \in [0,\pi]^d.$$

# Fourier-analytic inequality on the torus

Lemma 2. For all $\mu$ and $\nu$ on $Q^d$,

$$\widetilde{W}(\mu,\nu) \le \left(\sum_{m\neq 0} \frac{|f_\mu(m) - f_\nu(m)|^2}{|m|^2}\right)^{1/2}.$$

Proof. Given a (smooth) function $u$ participating in Lemma 1 for $\widetilde{W}(\mu,\nu)$, expand it as a multiple Fourier series

$$u(x) = \sum_{m\in\mathbb{Z}^d} a_m\, e^{im\cdot x}.$$

Hence

$$\int_{Q^d} u\, d\mu - \int_{Q^d} u\, d\nu \;=\; \sum_{m\in\mathbb{Z}^d} a_m\,(f_\mu(m) - f_\nu(m)).$$

Also, since $\partial_{x_l} u(x) = i\sum_{m\in\mathbb{Z}^d} m_l\, a_m\, e^{im\cdot x}$. we have

$$\frac{1}{(2\pi)^d}\int_{Q^d} |\partial_{x_l} u(x)|^2\, dx \;=\; \sum_{m\in\mathbb{Z}^d} m_l^2\, |a_m|^2,$$

$$\frac{1}{(2\pi)^d}\int_{Q^d} |\nabla u(x)|^2\, dx \;=\; \sum_{m\in\mathbb{Z}^d} |m|^2\, |a_m|^2 \;\le\; 1.$$

It remains to apply Cauchy's inequality and take the sup over all $u \in \mathcal{L}_d$.

# Convolution

Identifying $(-\pi, \pi]$ with the circle $S^1$ via $x \to e^{ix}$, and $Q^d = (-\pi, \pi]^d$ with $(S^1)^d$, one may introduce the convolution for prob. measures $\mu$ and $\gamma$ on $Q^d$:

$$\int u \, d\,(\mu * \gamma) = \iint u(x+y)\, d\mu(x)\, d\gamma(y)$$

in the class of all continuous $(2\pi)$-periodic functions $u$ on $\mathbb{R}^d$. Applying this equality to $u(x) = e^{im\cdot x}$, it becomes

$$f_{\mu*\gamma}(m) = f_\mu(m)\, f_\gamma(m), \qquad m \in \mathbb{Z}^d.$$

"Gaussian" measures on the torus: For each $t > 0$, there is $\gamma_t$ on $Q^d$ with Fourier-Stieltjes transform

$$f_{\gamma_t}(m) = e^{-t|m|^2/2}, \quad m \in \mathbb{Z}^d.$$

Construction ($d = 1$): $\gamma_t$ is the image of $N(0,t)$ under the map

$$T(x) = x - 2\pi k \quad \text{for } \pi(2k-1) < x \leq \pi(2k+1).$$

In particular, $|T(x)| \leq |x|$, so,

$$\int x^2 \, d\gamma_t(x) \leq t.$$

# Smoothing inequality. Proof of Theorem 1

Put $\mu_t = \mu * \gamma_t$, so that $f_{\mu_t}(m) = f_\mu(m)\, e^{-t|m|^2/2}$, $m \in \mathbb{Z}^d$.

If $u$ is $(2\pi)$-periodic and $\|u\|_{\mathrm{Lip}} \le 1$,

$$\left| \int u\, d\mu_t - \int u\, d\mu \right| = \left| \iint (u(x+y) - u(x))\, d\mu(x)\, d\gamma_t(y) \right| \le \int |y|\, d\gamma_t(y).$$

Conclusion (smoothing inequality): $\widetilde{W}(\mu_t, \mu) \le \sqrt{dt}$.

Proof of Theorem 1 (FA inequality): Using the triangle inequality for $\widetilde{W}$, the smoothing inequality, and applying Lemma 1 to $(\mu, \nu)$, and Lemma 2 to $(\mu_t, \nu_t)$, we have

$$\begin{aligned} W(\mu,\nu) = \widetilde{W}(\mu,\nu) &\le \widetilde{W}(\mu_t,\nu_t) + \widetilde{W}(\mu_t,\mu) + \widetilde{W}(\nu_t,\nu) \\ &\le \left( \sum_{m\neq 0} e^{-t|m|^2} \frac{|f_\mu(m) - f_\nu(m)|^2}{|m|^2} \right)^{1/2} + 2\sqrt{dt}. \end{aligned}$$

# Non-Euclidean metrics

Let the metric

$$\rho(x,y) = \omega(|x-y|), \qquad x,y \in \mathbb{R},$$

be generated by the modulus of continuity $\omega : \mathbb{R}_+ \to \mathbb{R}_+$, i.e. a non-decreasing continuous function such that $\omega(0) = 0$ and $\omega(\delta_1+\delta_2) \le \omega(\delta_1)+\omega(\delta_2)$ for all $\delta_1, \delta_2 \ge 0$.

Kantorovich transport distance:

$$W_\omega(\mu,\nu) \;=\; \inf_\lambda \int\!\!\int \omega(|x-y|)\, d\lambda(x,y),$$

where the inf is over all $\lambda$ on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals $\mu$ and $\nu$.

Theorem 1' (FA inequality). If $\mu,\nu$ are supported on $[0,\pi]^d$, for any $t>0$,

$$W(\mu,\nu) \le \sqrt{d}\left(\sum_{m\neq 0} \omega^2\Big(\frac{\pi}{|m|}\Big)\, e^{-t|m|^2}\, |f_\mu(m) - f_\nu(m)|^2\right)^{1/2} + 6\,\omega\big(\sqrt{dt}\,\big).$$

# Empirical measures

Kantorovich distance:

$$W_\omega(\mu_n, \nu_n) \ = \ \frac{1}{n} \ \inf_{\sigma \in S_n} \sum_{k=1}^{n} \omega(|X_k - Y_{\sigma(k)}|),$$

$$\mu_n = \frac{1}{n}(\delta_{X_1} + \cdots + \delta_{X_n}), \qquad \nu_n = \frac{1}{n}(\delta_{Y_1} + \cdots + \delta_{Y_n}),$$

Corollary 1'. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be pairwise independent with values in $[0,1]^d \times [0,1]^d$ such that $X_k$ and $Y_k$ have equal distributions for each $k \geq 1$. For any $t > 0$,

$$\mathbb{E}\, W_\omega(\mu_n, \nu_n) \ \leq \ \frac{2\sqrt{d}}{\sqrt{n}} \left( \sum_{m \neq 0} \omega^2\Big(\frac{\pi}{|m|}\Big)\, e^{-tm^2} \right)^{1/2} + 6\, \omega(\sqrt{dt}).$$

If all $X_j, Y_k$ are independent, a similar inequality also holds for the $\psi_2$-norm of $W_\omega(\mu_n, \nu_n)$.

# Dimension one (Zolotarev metrics)

For power moduli of continuity $\omega(\delta) = \delta^\alpha$, $0 < \alpha \le 1$,

$$\begin{aligned} W_\omega(\mu,\nu) \;&=\; \zeta_\alpha(\mu,\nu) \;=\; \inf_\lambda \iint |x-y|^\alpha \, d\lambda(x,y) \\ &=\; \sup_{\|u\|_{\mathrm{Lip}(\alpha)\le 1}} \left| \int u\, d\mu - \int u\, d\nu \right|. \end{aligned}$$

Corollary 1''. We have

$$\begin{aligned} \sqrt{2\alpha-1}\;\mathbb{E}\,\zeta_\alpha(\mu_n,\nu_n) \;&\le\; \frac{c}{\sqrt{n}}, && \text{if } \ \frac{1}{2} < \alpha \le 1, \\ \mathbb{E}\,\zeta_\alpha(\mu_n,\nu_n) \;&\le\; \frac{c}{\sqrt{n}}\sqrt{\log(2n)}, && \text{if } \ \alpha = \frac{1}{2}, \\ \sqrt{1-2\alpha}\;\mathbb{E}\,\zeta_\alpha(\mu_n,\nu_n) \;&\le\; \frac{c}{n^\alpha}, && \text{if } \ 0 < \alpha < \frac{1}{2}. \end{aligned}$$

If all $X_j, Y_k$ are independent, similar bounds also hold for the $\psi_2$-norm of the random variable $\zeta_\alpha(\mu_n,\nu_n)$.

The iid case: N. Fournier and A. Guillin (2015).

# Comparison with Bernstein's theorem

S. Bernstein's theorem (1914): If a $(2\pi)$-periodic function

$$u(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos(nx) + b_n \sin(nx))$$

belongs to the Lipschitz class of order $\alpha > \frac{1}{2}$, i.e.,

$$|u(x) - u(y)| \leq L\,\omega(|x-y|) = L\,|x-y|^{\alpha},$$

then

$$\sum_{n=1}^{\infty} (|a_n| + |b_n|) < \infty.$$

# Comparison with order statistics

Given two independent samples $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ drawn from the law $\mu$ on $[0,1]$, consider

$$\xi_n \;=\; \zeta_\alpha(\mu_n, \nu_n) \;=\; \frac{1}{n} \inf_\sigma \sum_{k=1}^n |X_k - Y_{\sigma(k)}|^\alpha,$$

In case $\alpha = 1$, it is simplified to

$$\zeta_1(\mu, \nu) \;=\; W(\mu_n, \nu_n) \;=\; \frac{1}{n} \sum_{k=1}^n |X_k^* - Y_k^*|$$

via order statistics $X_1^* \le \dots \le X_n^*$ and $Y_1^* \le \dots \le Y_n^*$. Define

$$\xi_n^* \;=\; \frac{1}{n} \sum_{k=1}^n |X_k^* - Y_k^*|^\alpha,$$

so that $\xi_n^* \ge \xi_n$, and we have $\xi_n^* = \xi_n$ in case $\alpha = 1$.

Corollary 2. Let $0 < \alpha < 1$. If $\mu$ is log-concave, then

$$\mathbb{E}\, \xi_n^* \;\ge\; \frac{c}{n^{\alpha/2}} \, \mathrm{Var}(X_1).$$

Hence, if $n$ is large, then $\xi_n^* > \xi_n$ with positive probability.

# Minimax grid matching

Minimax matching length: Given two unordered collections $X = \{x_1, \dots, x_n\}$ and $Y = \{y_1, \dots, y_n\}$ of points in $\mathbb{R}^d$, define the distance

$$L(X,Y) = \min_{\sigma \in S_n} \max_{1 \le k \le n} |x_k - y_{\sigma(k)}|.$$

Comparison with Kantorovich: $W \le L$.

Minimax grid matching problem: Find the rate of $\mathbb{E}\, L(X,Y)$ at which it tends to zero as $n \to \infty$, when $X$ and $Y$ are independent samples drawn from $\mu$. Of a special interest – the uniform distribution on $[0,1]^d$.

Case $d \ge 3$ (Shor and Yukich 1991):

$$c_0 \left(\frac{\log(2n)}{n}\right)^{1/d} \le \mathbb{E}\, L(X,Y) \le c_1 \left(\frac{\log(2n)}{n}\right)^{1/d}.$$

Case $d = 2$ (Leighton and Shor 1989):

$$c_0 \frac{\log^{3/4}(2n)}{\sqrt{n}} \le \mathbb{E}\, L(X,Y) \le c_1 \frac{\log^{3/4}(2n)}{\sqrt{n}}.$$

Case $d = 1$:

$$c_0 \frac{1}{\sqrt{n}} \le \mathbb{E}\, L(X,Y) \le c_1 \frac{1}{\sqrt{n}}.$$

# Restricted minimax grid matching

One can improve the rate in dim. one, if counting not all, but most of the points in perfect matchings. For a (non-empty) subset $I$ of $\{1, \dots, n\}$, define

$$L_I(X,Y) \,=\, \min_{\sigma} \max_{k \in I} |x_k - y_{\sigma(k)}|,$$

where the minimum is over all permutations $\sigma$ of $\{1, \dots, n\}$.

Let $Y = (Y_1, \dots, Y_n)$ be an independent copy of the random vector $X = (X_1, \dots, X_n)$ which has independent coordinates with values in $[0,1]$.

Corollary 3. For each $\varepsilon > 0$, with high probability there is a (random) set $I \subset \{1, \dots, n\}$ of $|I| \geq (1-\varepsilon)\, n$ such that

$$L_I(X,Y) \,\leq\, C\, \frac{\log^2(2n)}{\varepsilon^2 n}.$$

"With high probability" means the probability $1 - n^{-p}$, where $p$ can be chosen to be large by a proper choice of $C$. By Corollary 1" with $\alpha = \frac{1}{2}$, for some $\sigma$,

$$|X_k - Y_{\sigma(k)}|^{1/2} \,\leq\, \frac{c}{\sqrt{n}}\, \sqrt{\log(2n)}$$

on average with respect to $k$.