Optimal transportation and geometry of Wasserstein spaces

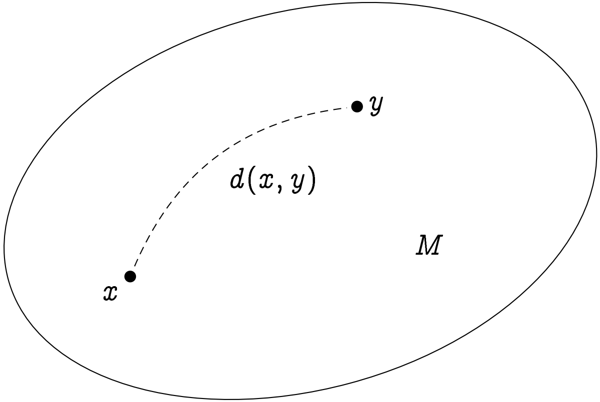
Andrei Sobolevski[1] Aleksei Kroshnin[1,2]

[1] Institute for Information Transmission Problems
of the Russian Academy of Sciences
(Kharkevich Institute)
[2] Moscow Institute of Physics and Technology

Quasilinear Equations, Inverse Problems
and Their Applications
MIPT, September 13, 2016

Gennadi Markovich Henkin

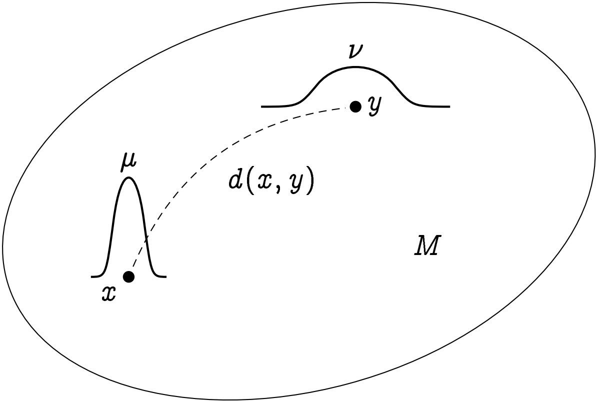
Probability measures as points of geodesic space



M locally compact geodesic space

$\mathcal{P}(M)$ space of probability measures on M

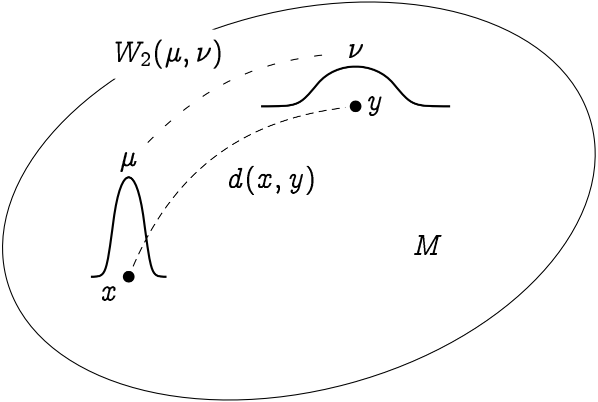
Probability measures as points of geodesic space



M locally compact geodesic space

$\mathcal{P}(M)$ space of probability measures on M

Probability measures as points of geodesic space



M locally compact geodesic space

$\mathcal{P}(M)$ space of probability measures on M

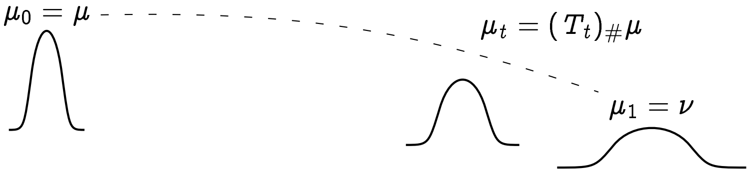
Wasserstein distance $W_2(\mu,\nu)$

Let $M \subset \mathbb{R}^d$, $d(x,y) = |x-y|$

$$(W_2(\mu,\nu))^2 = \inf_{\substack{\gamma \in \mathcal{P}(M \times M) \\ \swarrow \quad \searrow \\ \mu \qquad \nu}} \int |x-y|^2 \, \mathrm{d}\gamma(x,y)$$

Infimum is attained at transport plan $\gamma^* = (\mathrm{id}, T^*)_{\#}\mu$
where $T^*_{\#}\mu = \nu$ and $T = \nabla\Phi$ with Φ convex [Y. Brenier 1991]

Geodesics in $\mathcal{P}(M)$: Displacement interpolation



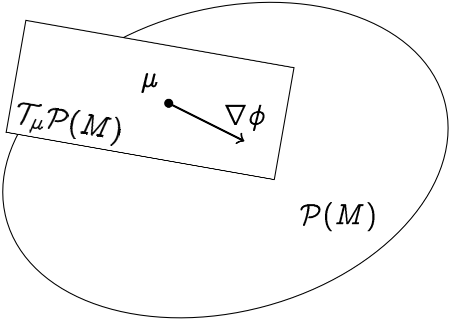
$$T_t: \quad x \mapsto (1-t)x + t\ T^* x$$
$$= x + t(T^* x - x)$$

Curl-free velocity field:

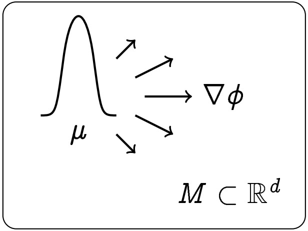
$$T^* x - x = \nabla(\Phi(x) - \frac{|x|^2}{2})n =: \nabla\phi(x)$$

[R. J. McCann 1997]

Formal Riemannian structure of $\mathcal{P}(M)$



$\mathcal{T}_\mu\mathcal{P}(M)$ is formed by all curl-free vector fields $\nabla\phi$ on M



$$\|\nabla\phi\|^2_{\mathcal{T}_\mu\mathcal{P}(M)} = \int_M |\nabla\phi(x)|^2 \, \mathrm{d}\mu$$

[F. Otto 2001]

Barycenter of a measure P

. . . in Euclidean space $\mathbb{R}^d$:

$$\bar{x} = \int x \, \mathrm{dP}(x) = \arg\min_y \frac{1}{2} \int |x - y|^2 \, \mathrm{dP}(x)$$

. . . in Riemannian manifold

$$\bar{x} = \arg\min_y \mathcal{P}(y), \text{ where } \mathcal{P}(y) = \frac{1}{2} \int d(x, y)^2 \, \mathrm{dP}(x)$$

$\bar{x}$ zero of $\nabla \mathcal{P}(y) = -\int \exp_y^{-1}(x) \, \mathrm{dP}(x)$

$\mathcal{P}(\cdot)$ locally convex if sectional curvature is bounded from above

[K. Grove, H. Karcher 1973]

Wasserstein barycenters

A generalization of displacement interpolation

$$\bar{\nu} = \arg\min_{\nu} \frac{1}{2} \sum_i \lambda_i \; W_2^2(\mu_i, \nu) \quad \text{(discrete, } \lambda_i \geq 0, \; \sum_i \lambda_i = 1)$$

$$\bar{\nu} = \arg\min_{\nu} \frac{1}{2} \int W_2^2(\mu, \nu) \, \mathrm{dP}(\mu) \quad \text{(general)}$$

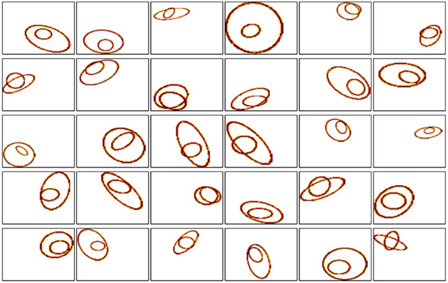
Characterization in terms of $\int \nabla \phi_\mu^* \, \mathrm{dP}(\mu)$

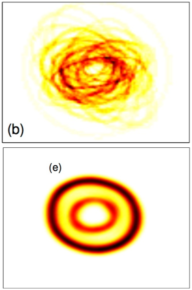
Convexity properties

[M. Agueh, G. Carlier 2011]

Wasserstein barycenters

Conventional vs Wasserstein averaging of measures in 2D:





[M. Cuturi, A. Doucet 2014]

Fréchet barycenters in general metric spaces

Les éléments aléatoires de nature quelconque dans un espace distancié

par

Maurice FRÉCHET.

Éléments aléatoires nouveaux. — Le Calcul des probabilités a été implicitement ou explicitement, jusqu'à une époque récente, l'étude des *nombres* aléatoires et des points aléatoires dans un espace à une, deux ou trois dimensions (probabilités géométriques).

Depuis peu, on a souvent cherché à étendre les résultats obtenus aux séries aléatoires, aux vecteurs aléatoires et aux fonctions numériques aléatoires de variables numériques certaines.

Mais la nature, la science et la technique offrent de nombreux exemples d'éléments aléatoires qui ne sont, ni des nombres, ni des séries, ni des vecteurs, ni des fonctions.

[M. Fréchet 1948]

Fréchet barycenter for measures

$$J(\mu,\nu) = \inf_{\substack{\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \\ \swarrow \qquad \searrow \\ \mu \qquad\qquad \nu}} \int g(x-y)\,\mathrm{d}\gamma(x,y)$$

$g(\cdot) \geq 0$ strictly convex, $g(0) = 0$

$J \geq 0$, $J(\mu,\nu) = 0$ iff $\mu - \nu$ but no triangle inequality

$$\bar{\nu} = \arg\min_{\nu} \int J(\mu,\nu)\,\mathrm{dP}(\mu)$$

[A. Kroshnin, A. Sobolevski 2015]

Fréchet barycenters for measures in $\mathbb{R}^1$

$F_\mu^{-1}\colon (0,1) \to \mathbb{R}^1$ quantile function of measure μ

Theorem

Quantiles of Fréchet barycenter are generalized means of quantiles:

$$F_{\bar{\nu}}^{-1}(\alpha) = \arg\min_y \int g(F_\mu^{-1}(\alpha) - y)\,\mathrm{dP}(\mu), \qquad 0 < \alpha < 1$$

[A. Kroshnin, A. Sobolevski 2015]

Law of large numbers and statistical consistency

Theorem

$\bar{\nu}_n \to \bar{\nu}$ *weakly as* $n \to \infty$
$\bar{\nu}_n$ *barycenter of independent sample of size* n *from* P
$\bar{\nu}$ *barycenter of* P

Theorem

$J(\bar{\nu}_n, \bar{\nu}) \to 0$ *as* $n \to \infty$ *("strong" convergence)*

[A. Kroshnin, A. Sobolevski 2015]

[T. Le Gouic, J.-M. Loubes 2015]:
measures in general geodesic space instead of $\mathbb{R}^1$;
Wasserstein barycenters instead of general Fréchet barycenters

Bibliography I

M. Agueh, G. Carlier.
Barycenters in the Wasserstein Space.
SIAM J. Math. Anal. 43:2 (2011) 904–924.

Y. Brenier.
Polar factorization and monotone rearrangement of vector-valued functions.
Comm. Pure Appl. Math. 44:4 (1991) 375–417.

M. Cuturi, A. Doucet.
Fast computation of Wasserstein barycenters.
Proc. 31 ICML, Beijing, 2014.

M. Fréchet.
Les éléments aléatoires de nature quelconque dans un espace distancié.
Annales de l'IHP 10:4 (1948) 215–310.

Bibliography II

K. Grove, H. Karcher.
How to conjugate C^1-close group actions.
Math. Z. 132 (1973) 11-20.

A. Kroshnin, A. Sobolevski.
Fréchet barycenters and a law of large numbers for measures on the real line.
arXiv:1512.08421v2 [math.PR] (2015) 15 pp.

T. Le Gouic, J.-M. Loubes.
Existence and consistency of Wasserstein barycenters.
arXiv:1506.04153 [math.ST] (2015) 19 pp.

R. J. McCann.
A convexity principle for interacting gases.
Adv. Math. 128:1 (1997) 153–179.

Bibliography III

F. Otto.
The geometry of dissipative evolution equations: the porous medium equation.
Comm. PDE 26:12 (2001) 101–174.