Quantitative stability of push-forwards by optimal maps

Guillaume Carlier [a]

Joint work with Alex Delalande and Quentin Mérigot.

Journées en l'honneur de Christian Léonard, Paris, July, 2025.

[a] CEREMADE, Université Paris Dauphine and MOKAPLAN (Inria-Dauphine).

Christian, un modèle scientifique mais pas que...

Par modestie, il vous l'a caché, mais Christian a déjà rejoint un groupe de doom métal progressif (dont le prochain album s'intitulera évidemment *Schrödinger bridge to Babylon*).

Professor Emeritus

Introduction

The title obviously needs some explanations.

Push-forwards (aka image measures) Let $\rho \in \mathscr{P}(\mathbb{R}^d)$ (Borel probability measure on $\mathbb{R}^d$) and $T : \mathbb{R}^d \to \mathbb{R}^d$ a Borel map. The pushforward of ρ through T is the probability measure $T_\# \rho$ defined by

$$T_\# \rho(B) = \rho(T^{-1}(B)), \forall B \text{ Borel subset of } \mathbb{R}^d.$$

In other words, if X is a random variable with law ρ $(X \sim \rho)$ then $T(X)$ has law $T_\# \rho$.

Very natural object, but tricky to compute in general, examples

- easy case, discrete to discrete: $\rho = \frac{1}{N}\sum_{i=1}^{N}\delta_{x_i}$, then $T_\#\rho$ is just $\frac{1}{N}\sum_{i=1}^{N}\delta_{T(x_i)}$,
- T with finite range: $T = \sum_{i=1}^{N}\mathbf{1}_{A_i}y_i$, then $T_\#\rho = \sum_{i=1}^{N}\rho(A_i)\delta_{y_i}$,
- ρ has a density f and T is a diffeomorphism, change of variables formula: $T_\#\rho$ has density g with

$$|\det(DT)|g \circ T = f.$$

Question: Under which reasonable assumptions can we say that when ρ is close to $\widetilde{\rho}$, then $T_\#\rho$ is close to $T_\#\widetilde{\rho}$?

Imagine ρ is "nice" (easy to draw samples from e.g. uniform on a cube) and $T_\#\rho$ is some target measure, ρ approximated (by quantization or sampling) by $\widetilde{\rho}$, is $T_\#\widetilde{\rho}$ close to the target?

By close, we mean in some weak convergence sense, quantified by Wasserstein distance.

Given $p \geq 1$, Wasserstein p-distance between ρ and $\widetilde{\rho}$:

$$W_p(\rho, \widetilde{\rho})^p := \inf \left\{ \mathbb{E}(|X - \widetilde{X}|^p),\ X \sim \rho,\ \widetilde{X} \sim \widetilde{\rho} \right\}$$
$$= \inf_{\gamma \in \Pi(\rho, \widetilde{\rho})} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - \tilde{x}|^p \mathrm{d}\gamma(x, \tilde{x}) \right\}$$

where $\Pi(\rho, \widetilde{\rho})$ is the set of probability measures having ρ and $\widetilde{\rho}$ as marginals. Projections $\pi_1(x, y) = x$, $\pi_2(x, y) = y$, then

$$\Pi(\rho, \widetilde{\rho}) = \{\gamma \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d) \ : \ {\pi_1}_{\#}\gamma = \rho,\ {\pi_2}_{\#}\gamma = \rho\}$$

Makes sense for ρ and $\widetilde{\rho}$ with finite p-moments. This is an instance of the Monge-Kantorovich optimal transport problem. Existence is easy.

Known facts: W_p is a distance, it metrizes weak convergence (convergence in law for random variables), $W_1 \leq W_p$ (Hölder), and for ρ and $\widetilde{\rho}$ supported on B_R,

$$W_p(\rho, \widetilde{\rho}) \leq (2R)^{\frac{p-1}{p}} W_1(\rho, \widetilde{\rho})^{\frac{1}{p}}$$

most useful cases are $p = 1$ and $p = 2$.

Recall for $p = 1$, the Kantorovich-Rubinstein inequality:

$$\int_{\mathbb{R}^d} u \mathrm{d}(\rho - \widetilde{\rho}) \leq \mathrm{Lip}(u) W_1(\rho, \widetilde{\rho}).$$

If T is Lipschitz, obviously by definition of W_p, we have

$$W_p(T_\# \rho, T_\# \widetilde{\rho}) \leq \mathrm{Lip}(T) W_p(\rho, \widetilde{\rho}).$$

End of the story? No: what if T is discontinuous or even possibly set-valued (in which case $T_\# \rho$ is not even well-defined). And in particular what if T is an "optimal" (in a way I will explain) map in the sense that $T \in \partial \phi$ with ϕ convex?

Outline

① Quadratic optimal transport, optimal maps

② Main result

③ How small is the singular set of a convex function?

The quadratic optimal transport problem

Quadratic OT, Brenier's seminal results (1989, 1991). Let α be a probability supported on B_R. Developing the squared distance, note that

$$\frac{W_2^2(\rho,\alpha)}{2} = \frac{1}{2}\int_{\mathbb{R}^d} |x|^2 \mathrm{d}\rho(x) + \frac{1}{2}\int_{\mathbb{R}^d} |y|^2 \mathrm{d}\alpha(y) - \sup_{\gamma\in\Pi(\rho,\alpha)} \int_{\mathbb{R}^d\times\mathbb{R}^d} x\cdot y \,\mathrm{d}\gamma(x,y)$$

so γ is optimal iff it maximizes the correlation under the marginal constraints.

Duality

$$\sup_{\gamma\in\Pi(\rho,\alpha)} \int_{\mathbb{R}^d\times\mathbb{R}^d} x\cdot y \,\mathrm{d}\gamma(x,y)$$

coincides with the infimum of

$$\int_{B_R} \phi \mathrm{d}\rho + \int_{B_R} \psi \mathrm{d}\alpha$$

among pairs (ϕ,ψ) such that

$$\phi(x)+\psi(y)\geq x\cdot y,\ \forall(x,y)\in B_R\times B_R$$

Given ϕ, the smallest admissible ψ is

$$\psi(y)=\phi^*(y)=\max_{x\in B_R}\{x\cdot y-\phi(x)\},\ y\in B_R.$$

.

We still improve the cost by taking

$$\phi(x) = \psi^*(x) = \max_{y \in B_R}\{x \cdot y - \psi(y)\},\ x \in B_R$$

so that ϕ is convex and R-Lipschitz

Given $\phi = \psi^*$, $\psi = \phi^*$ a pair of conjugate optimal potentials, a plan $\gamma \in \Pi(\rho, \alpha)$ is optimal iff it satisfies the complementary slackness condition

$$\phi(x) + \phi^*(y) = x \cdot y,\ \text{on spt}\gamma.$$

Optimality of γ is therefore equivalent to the fact that $\mathrm{spt}\gamma \subset \partial\phi$ for some convex and R Lipschitz ϕ; we recall that for a convex ϕ,

$$\partial\phi = \{(x, y) \ : \ y \in \partial\phi(x)\}$$

where $\partial\phi(x)$ is the subdifferential of ϕ at x i.e. the set of y's for which

$$\phi(x) + \phi^*(y) = x \cdot y$$

or equivalently

$$\phi(x') \geq \phi(x) + (x' - x) \cdot y, \ \forall x'.$$

γ optimal: disintegrate it with respect to its first marginal

$$\gamma(\mathrm{d}x, \mathrm{d}y) = \rho(\mathrm{d}x) \otimes \gamma^x(\mathrm{d}y)$$

i.e.

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} f(x, y)\gamma(\mathrm{d}x, \mathrm{d}y) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x, y)\gamma^x(\mathrm{d}y)\rho(\mathrm{d}x)$$

then $\mathrm{spt}\gamma^x \subset \partial\phi(x)$ (optimal plans send all the mass at x to $\partial\phi(x)$).

If ρ is absolutely continuous, then $\partial\phi(x) = \{\nabla\phi(x)\}$ for a.e. x and then $\gamma^x = \delta_{\nabla\phi(x)}$.

Brenier's theorem (ρ absolutely continuous): the optimal plan is unique, it is induced by a map, this map is the gradient of a convex function (and if a map in such a form pushes forward ρ to α, it is an optimal map).

Brenier's map $T = \nabla\phi$ is a remarkable change of variable (or transport) between ρ and α. Monotone and has a potential, $DT = D^2\phi$, SDP, related to Monge-Ampère equations etc...

Brenier's map is a (nontrivial) extension to several dimensions of the notion of monotone change of variables.

One can parameterize the set of probability measures, fixing a "nice" reference measure ρ by the bijective map

$$T \in \{\nabla\phi, \text{ with } \phi \text{ convex}\} \mapsto T_{\#}\rho.$$

Note that $T = \nabla\phi$ is the optimal map between ρ and $\mapsto T_{\#}\rho$.

Main result

We wish to investigate the following stability question for optimal plans/maps (i.e. subgradients/gradients of convex functions). Given ϕ convex and Lipschitz, ρ and $\widetilde{\rho}$ in $\mathscr{P}(\mathbb{R}^d)$ supported on B_R, γ and $\widetilde{\gamma}$ in $\mathscr{P}(\mathbb{R}^d, \mathbb{R}^d)$ such that

- $\pi_{1\#}\gamma = \rho$, $\pi_{1\#}\widetilde{\gamma} = \widetilde{\rho}$
- $\mathrm{spt}\gamma \subset \partial\phi$, $\mathrm{spt}\widetilde{\gamma} \subset \partial\phi$ (so that both γ and $\widetilde{\gamma}$ are optimal between their marginals)

Can we bound $W_2(\pi_{2\#}\gamma, \pi_{2\#}\widetilde{\gamma})$ in terms of $W_2(\rho, \widetilde{\rho})$?

The answer is obviously no without additional assumptions: take $d = 1$, $\rho = \widetilde{\rho} = \delta_0$, $\phi = |\cdot|$ and $\gamma = \delta_{(0,1)}$, $\widetilde{\gamma} = \delta_{(0,-1)}$ then $W_2(\rho, \widetilde{\rho}) = 0$ and $W_2(\pi_{2\#}\gamma, \pi_{2\#}\widetilde{\gamma}) = 2$!

We should at least ask that ρ is absolutely continuous (so that it does not see the singular set of ϕ). To make things as simple as possible, we shall always assume:

- ρ and $\widetilde{\rho}$ in $\mathscr{P}(\mathbb{R}^d)$ are supported on B_R, ϕ convex and R-Lipschitz (so $\pi_{2\#}\gamma$, and $\pi_{2\#}\widetilde{\gamma}$ are supported on B_R as well),

- ρ is absolutely continous with a density bounded by M_ρ (so that $\gamma = (\mathrm{id}, \nabla\phi)_\#\rho$ and $\pi_{2\#}\gamma = \nabla\phi_\#\rho$) .

An example. Let $\phi = |\cdot|$ on $\mathbb{R}$ and let $\varepsilon \in (0, \frac{1}{2})$. Set

$$\rho = \lambda_{[-\frac{1}{2},\frac{1}{2}]} \text{ and } \rho^\varepsilon = \lambda_{|[-\frac{1}{2},-\frac{\varepsilon}{2}]\cup[\frac{\varepsilon}{2},\frac{1}{2}]} + \varepsilon\delta_0.$$

Let $\gamma = (\mathrm{id}, \mathrm{sign})_\# \rho$ and

$$\gamma^\varepsilon = \int_{[-\frac{1}{2},-\frac{\varepsilon}{2}]\cup[\frac{\varepsilon}{2},\frac{1}{2}]} \delta_{(x,\mathrm{sign}(x))}\mathrm{d}x + \varepsilon\delta_{(0,1)}.$$

Then

$$\pi_{1\#}\gamma = \rho,\ \pi_{1\#}\gamma^\varepsilon = \rho^\varepsilon,\ \mathrm{spt}(\gamma) \subset \partial\phi,\ \mathrm{spt}(\gamma^\varepsilon) \subset \partial\phi.$$

Easy to compute

$$W_2(\pi_{2\#}\gamma, \pi_{2\#}\gamma^\varepsilon) = (2\varepsilon)^{1/2},\ W_2(\rho, \rho^\varepsilon) = (\varepsilon^3/12)^{1/2},$$

so that $W_2(\pi_{2\#}\gamma, \pi_{2\#}\gamma^\varepsilon) \sim W_2(\rho, \rho^\varepsilon)^{1/3}$.

Theorem 1 (C., Delalande, Mérigot) *Let ρ and $\widetilde{\rho}$ be probability measures supported on the ball B_R of $\mathbb{R}^d$, with ρ absolutely continuous with a density bounded by M_ρ and let ϕ be a convex R-Lipschitz function, let $\widetilde{\gamma}$ in $\mathscr{P}(\mathbb{R}^d, \mathbb{R}^d)$ such that*

- $\pi_{1\#}\widetilde{\gamma} = \widetilde{\rho}$
- $\mathrm{spt}\widetilde{\gamma} \subset \partial\phi$

then we have

$$W_2(\nabla\phi_{\#}\rho, \pi_{2\#}\widetilde{\gamma}) \leq C W_2(\rho, \widetilde{\rho})^{\frac{1}{3}} \tag{1}$$

for the explicit constant

$$C = 2^{8d+\frac{23}{2}} d^2 (1+\omega_d)(1+M_\rho)(1+R)^{4+d},$$

with ω_d denoting the volume of the unit ball of $\mathbb{R}^d$.

The previous example shows that a Hölder quantitative stability estimate with exponent $1/3$ is the best we can hope for. Sharp in terms of exponent (likely not at all for the constant C).

Extension to (sharp in terms of exponents) comparisons in W_p/W_q distances and push forward by optimal maps for W_r (with $r \geq 2$ so that optimal potentials are semiconcave) ....

The first step in the proof is the following. Let S be an optimal map between ρ and $\widetilde{\rho}$ i.e. $S_{\#}\rho = \widetilde{\rho}$ and $W_2(\rho, \widetilde{\rho}) = \|\mathrm{id} - S\|_{L^2(\rho)}$. Disintegrate $\widetilde{\gamma}$ as

$$\widetilde{\gamma} = \widetilde{\rho} \otimes \widetilde{\gamma}^{\tilde{x}}$$

then since $S_{\#}\rho = \widetilde{\rho}$ the plan

$$\int_{B_R} \delta_{\nabla\phi(x)} \otimes \widetilde{\gamma}^{S(x)} \rho(x)dx$$

is a transport plan between $\nabla\phi_{\#}\rho, \pi_{2\#}\widetilde{\gamma}$, we have

$$W_2^2(\nabla\phi_{\#}\rho, \pi_{2\#}\widetilde{\gamma}) \leq \int_{B_R}\int_{B_R} |\nabla\phi(x) - y|^2 \mathrm{d}\widetilde{\gamma}^{S(x)}(y)\rho(x)dx.$$

Let $\eta > 0$, we split the previous integrals into two terms the integral on $\Omega_\eta := \{|\operatorname{id} - S| > \eta\}$, note that by Markov's inequality we have

$$\rho(\Omega_\eta) \leq \frac{W_2^2(\rho, \widetilde{\rho})}{\eta^2}$$

so

$$\int_{\Omega_\eta} \int_{B_R} |\nabla\phi(x) - y|^2 \mathrm{d}\widetilde{\gamma}^{S(x)}(y)\rho(x)dx \leq \frac{4R^2 W_2^2(\rho, \widetilde{\rho})}{\eta^2}$$

We now have to estimate the contribution of $B_R \setminus \Omega_\eta = \{|\,\mathrm{id} - S| \leq \eta\}$. For $x \in B_R \setminus \Omega_\eta$, since $\mathrm{spt}(\widetilde{\gamma}^{S(x)}) \subset \partial\phi(S(x))$ we have

$$\int_{B_R} |\nabla\phi(x) - y|^2 \mathrm{d}\widetilde{\gamma}^{S(x)}(y) \leq \mathrm{diam}(\partial\phi(B(x,\eta)))^2$$

so

$$\begin{aligned}\int_{B_R\setminus\Omega_\eta} \int_{B_R} |\nabla\phi(x) - y|^2 \mathrm{d}\widetilde{\gamma}^{S(x)}(y)\rho(x)dx \\ \leq M_\rho \int_{B_R} \mathrm{diam}(\partial\phi(B(x,\eta))^2 \mathrm{d}x\end{aligned}$$

assume that

$$\int_{B_R} \mathrm{diam}(\partial\phi(B(x,\eta))^2 \mathrm{d}x \leq C\eta \tag{2}$$

then we obtain

$$W_2^2(\nabla\phi_{\#}\rho, \pi_{2\#}\widetilde{\gamma}) \lesssim \frac{W_2^2(\rho,\widetilde{\rho})}{\eta^2} + \eta$$

so choosing $\eta \sim W_2(\rho,\widetilde{\rho})^{\frac{2}{3}}$ we get the desired result.

It remains to prove (2): this is a matter of quantifying the smallness of the singular set of ϕ.

On the singular set of a convex function

$\phi : \mathbb{R}^d \to \mathbb{R}$ convex and Lipschitz, the singular set of ϕ where diam $\partial\phi > 0$ is of measure 0 (and in fact it is $d-1$-rectifiable, a remarkable fine geometric measure theoretical was done by Alberti, Ambrosio and Cannarsa). We want to quantify this smallness in terms of covering number. Recall that if K is a compact subset of $\mathbb{R}^d$ and $\eta > 0$, $\mathcal{N}(K, \eta)$ is the minimal number of balls of radius η needed to cover K.

Theorem 2 (C., Delalande, Mérigot) *Denote*

$$\Sigma_{\eta,\alpha} = \{x \in \mathbb{R}^d \ : \ \operatorname{diam}(\partial\phi(B(x,\eta)) \geq \alpha\},$$

then, we have

$$\mathcal{N}(\Sigma_{\eta,\alpha} \cap B_R, 8\eta) \leq c_{d,R,\eta} \frac{\operatorname{Lip}(\phi)}{\alpha\eta^{d-1}},$$

with $c_{d,R,\eta} = 48d^2(R + 4\eta)^{d-1}$.

The dependence in α and η is sharp.

A direct corollary is the fact that (2) holds:

$$
\begin{aligned}
&\int_{B_R} \operatorname{diam}(\partial\phi(B(x,\eta)))^2 \mathrm{d}x \\
&= \int_0^\infty \left|\{x \in B_R \ : \ \operatorname{diam}(\partial\phi(B(x,\eta)))^2 \geq t\}\right| \mathrm{d}t \\
&\leq \int_0^{(2\mathrm{Lip}(\phi))^2} 48 d^2 (R+4\eta)^{d-1} \frac{\mathrm{Lip}(\phi)}{t^{1/2}\eta^{d-1}} \omega_d (8\eta)^d \mathrm{d}t \\
&= c_{d,R,\eta} \mathrm{Lip}(\phi)^2 \eta.
\end{aligned}
$$

It remains to prove the bound on $\mathcal{N}(\Sigma_{\eta,\alpha} \cap B(0,R), 8\eta)$. Let $\varepsilon = 4\eta$ and Z be a maximal ε packing of $\Sigma := \Sigma_{\eta,\alpha} \cap B_R$ and N be the cardinality of Z, Z being a $2\varepsilon = 8\eta$ covering of Σ, $\mathcal{N}(\Sigma, 8\eta) \leq N$. Now for $x \in Z$, by construction we have

$$\alpha \leq \operatorname{diam}(\partial\phi(B(x,\eta))). \tag{3}$$

With the monotonicity of $\partial\phi$, one can prove that

Lemma 1 *If ϕ is convex*

$$\mathrm{diam}(\partial\phi(B(x,\eta))) \leq \frac{12}{\omega_d \eta^d} \|\nabla\phi\|_{L^1(B(x,4\eta))}.$$

Proof:. Firstly $\mathrm{diam}(\partial\phi(B(x,\eta))) \leq 2\|\nabla\phi\|_{L^\infty(B(x,\eta))}$. But if $y \in B(x,\eta)$ is a differentiability point and $z \in B(y,\eta)$, by convexity we have

$$\mathrm{osc}_{B(x,2\eta)}(\phi) \geq \phi(z) - \phi(y) \geq \nabla\phi(y)\cdot(z-y),$$

so maximing in z yields

$$\|\nabla\phi\|_{L^\infty(B(x,\eta))} \leq \frac{\mathrm{osc}_{B(x,2\eta)}(\phi)}{\eta}.$$

Let y_0 and y_1 be respectively a minimizer and maximizer of ϕ over $B(x, 2\eta)$, let $g_1 \in \partial\phi(y_1)$, $y \in \mathbb{R}^d$ a differentiability point, then

$$\phi(y_1) + g_1(y - y_1) \leq \phi(y) \leq \phi(y_0) + \nabla\phi(y) \cdot (y - y_0).$$

In particular if $y \in B(x, 4\eta) \cap H_+$ where H_+ is the half space where $g_1 \cdot (y - y_1) \geq 0$, we have

$$|\nabla\phi(y)| \geq \frac{\mathrm{osc}_{B(x,2\eta)}}{|y - y_0|}.$$

Now observe that

$$B(y_1 + \eta g_1/|g_1|, \eta) \subset B(y_1, 2\eta) \cap H_+ \subset B(x, 4\eta)$$

integrating the previous inequality yields

$$\begin{aligned}\|\nabla\phi\|_{L^1(B(x,4\eta)} &\geq \operatorname{osc}_{B(x,2\eta)} \int_{B(y_1+\eta g_1/|g_1|,\eta)} \frac{1}{|y-y_1| + |y_1 - y_0|} \mathrm{d}y \\ &\geq \frac{\eta^{d-1}\ \omega_d \operatorname{osc}_{B(x,2\eta)}}{6}\end{aligned}$$

hence, the desired inequality:

$$\operatorname{diam}(\partial\phi(B(x,\eta))) \leq \frac{2 \operatorname{osc}_{B(x,2\eta)}(\phi)}{\eta} \leq \frac{12}{\omega_d \eta^d} \|\nabla\phi\|_{L^1(B(x,4\eta))}.$$

Now if we denote by c the average of $\nabla\phi$ over $B(x, 4\eta)$, the L^1-Poincaré-Wirtinger's inequality together with the fact that $D^2\phi \leq \Delta\phi\mathrm{id}$ (in the sense of SDP matrix-valued measures) yields

$$\|\nabla\phi - c\|_{L^1(B(x,4\eta))} \lesssim \eta\Delta\phi(B(x, 4\eta))$$

so if $x \in Z$ we have

$$\begin{aligned} \alpha \leq \mathrm{diam}(\partial\phi(B(x, \eta)) &= \mathrm{diam}(\partial\phi(B(x, \eta)) - c) \\ &\lesssim \eta^{-d}\|\nabla\phi - c\|_{L^1(B(x,4\eta))} \\ &\lesssim \eta^{1-d}\Delta\phi(B(x, 4\eta)). \end{aligned}$$

Summing over Z, we get, using first the fact that Z is a 4η packing and $\Delta\phi \geq 0$ and then the divergence theorem and the fact that ϕ is Lipschitz

$$\begin{aligned} \alpha N &\lesssim \eta^{1-d} \sum_{x \in Z} \Delta\phi(B(x, 4\eta)) \lesssim \eta^{1-d} \Delta\phi(B_{R+4\eta}) \\ &= \eta^{1-d} \int_{B_{R+4\eta}} \Delta\phi = \eta^{1-d} \int_{\partial(B_{R+4\eta})} \frac{\partial\phi}{\partial n} \\ &\lesssim \eta^{1-d}(R+4\eta)^{d-1}\mathrm{Lip}(\phi). \end{aligned}$$

so

$$\mathcal{N}(\Sigma_{\eta,\alpha} \cap B_R, 8\eta) \leq N \leq \frac{C(R+4\eta)^{d-1}\mathrm{Lip}(\phi)}{\alpha\eta^{d-1}},$$

which is the desired result.

Bonne retraite (!) Christian et merci pour tout!!!