Global Regularity Estimates for the Optimal Transport via Entropic Regularization

Nathael Gozlan (Université Paris Cité,MAP5)
Maxime Sylvestre (Université Paris Dauphine PSL, CEREMADE)

March 2, 2025

Some known regularity results

Framework

Let μ and ν be two probability measures on $\mathbb{R}^n$ with finite second moments. The quadratic transport cost between μ and ν is defined as

$$W_2^2(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int \|y - x\|^2 \, \pi(dxdy). \tag{1}$$

Theorem [Brenier, 1991]

If μ is absolutely continuous with respect to the Lebesgue measure, then there exists a unique optimal transport map T from μ to ν such that $\nu = T_{\#}\mu$ and $\int \| T(x) - x\|^2 \, \mu(dx) = W_2^2(\mu, \nu)$. Moreover, T is the gradient of a convex function ϕ.

Caffarelli's contraction theorem

Theorem [Caffarelli, 1992]

Let $\mu(dx) = e^{-V(x)}dx$, $\nu(dy) = e^{-W(y)}dy$ be two probability measures on $\mathbb{R}^n$ such that $\mathrm{dom}\, V = \mathbb{R}^n$ and $\mathrm{dom}\, W$ is convex with non empty interior. Further assume that V, W are twice continuously differentiable on the interior of their domains and satisfy

$$\nabla^2 V \leq \alpha_V Id, \quad \nabla^2 W \geq \beta_W Id,$$

with $\alpha_V, \beta_W > 0$. Then the optimal transport map for the quadratic transport problem from μ to ν is $\sqrt{\alpha_V/\beta_W}$-Lipschitz.

Entropic framework

The entropic optimal transport problem is the following

$$\mathcal{C}_\epsilon(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int \frac{1}{2}\|x-y\|^2 \, d\pi + \epsilon H(\pi|\mu\otimes\nu)$$

where $H(\pi|\mu\otimes\nu) = \int \log \frac{d\pi}{d(\mu\otimes\nu)} \, d\pi$. It is now well known [Carlier et al., 2017], that $\mathcal{C}_\epsilon(\mu,\nu) \to \frac{1}{2}W_2^2(\mu,\nu)$ as $\epsilon \to 0$. The minimizer of the entropic regularized transport problem is of the form

$$\pi_\epsilon(dxdy) = e^{\frac{\langle x,y\rangle - \phi_\epsilon(x) - \psi_\epsilon(y)}{\epsilon}} \, \mu(dx)\nu(dy),$$

Entropic framework

with $(\phi_\epsilon, \psi_\epsilon)$ a couple of convex functions solution of the following system

$$\phi_\epsilon(x) = \mathcal{L}_{\epsilon,\nu}(\psi_\epsilon)(x) = \epsilon \log\left(\int e^{\frac{\langle x,y\rangle - \psi_\epsilon(y)}{\epsilon}}\, \nu(dy)\right), \qquad \forall x \in \mathbb{R}^n$$
$$\psi_\epsilon(y) = \mathcal{L}_{\epsilon,\mu}(\phi_\epsilon)(y) = \epsilon \log\left(\int e^{\frac{\langle x,y\rangle - \phi_\epsilon(x)}{\epsilon}}\, \mu(dx)\right), \qquad \forall y \in \mathbb{R}^n.$$

Moreover, the entropic Kantorovich potential ϕ_ϵ converges μ-almost everywhere (along to some sequence ϵ_k) to the Kantorovich potential ϕ such that $T = \nabla\phi$ [Nutz and Wiesel, 2021].

A regularity result for the entropic potentials

Theorem [Fathi et al., 2020, Chewi and Pooladian, 2023]

Let $\mu(dx) = e^{-V(x)}dx$, $\nu(dy) = e^{-W(y)}dy$ be two probability measures on $\mathbb{R}^n$ such that $\mathrm{dom}\, V = \mathbb{R}^n$ and $\mathrm{dom}\, W$ is convex with non empty interior. Further assume that V, W are twice continuously differentiable on the interior of their domains and satisfy

$$\nabla^2 V \leq \alpha_V Id, \quad \nabla^2 W \geq \beta_W Id,$$

with $\alpha_V, \beta_W > 0$. Then $\nabla\phi_\epsilon$ is $\sqrt{\alpha_V/\beta_W}$-Lipschitz, for all $\epsilon \geq 0$.

Three steps

The main objective is to relax the $\mathcal{C}^2$ hypothesis on V, W.

(*i*) Define a notion of smoothness which grants regularity of the gradient of a function without refering to its second derivative.

(*ii*) Study the interaction between smoothness and the entropic Legendre transform $\mathcal{L}_{\epsilon,m}$.

(*iii*) Apply the results to the entropic regularized transport problem.

Reminders on convex analysis

Smoothness and Strong convexity

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper, l.s.c. function. We say that f is R-convex, or that R is a convexity modulus for f, if for any $x_0, x_1 \in \mathbb{R}^n$ and $t \in [0, 1]$, we have

$$f((1-t)x_0 + tx_1) + t(1-t)R(x_1 - x_0) \leq (1-t)f(x_0) + tf(x_1).$$

We say that f is S-smooth, or that S is a smoothness modulus for f, if for any $x_0, x_1 \in \mathbb{R}^n$ and $t \in [0, 1]$, we have

$$f((1-t)x_0 + tx_1) + t(1-t)S(x_1 - x_0) \geq (1-t)f(x_0) + tf(x_1).$$

Remarks

Note that when $R(d) = \frac{\alpha}{2}\|d\|^2$ we recover the definition of α strong convexity, and when $S(d) = \frac{\beta}{2}\|d\|^2$ we recover β smoothness of the function.

Remarks

Note that when $R(d) = \frac{\alpha}{2}\|d\|^2$ we recover the definition of α strong convexity, and when $S(d) = \frac{\beta}{2}\|d\|^2$ we recover β smoothness of the function.

Definition

The smallest function S such that f is S-smooth is called the smoothness modulus of f and is denoted S_f. It is convex when f is convex.
The smallest function R sucht that f is R-convex is called the convexity modulus of f and is denoted R_f.

Examples

Quadratic functions Let A be a $n \times n$ matrix, then the function $f(x) = \langle x, Ax \rangle$ admits the following moduli

$$R_f(d) = S_f(d) = \langle d, Ad \rangle, \qquad d \in \mathbb{R}^n,$$

Examples

Quadratic functions Let A be a $n \times n$ matrix, then the function $f(x) = \langle x, Ax \rangle$ admits the following moduli

$$R_f(d) = S_f(d) = \langle d, Ad \rangle, \qquad d \in \mathbb{R}^n,$$

Bounded hessians Let f be a twice continuously differentiable function on $\mathbb{R}^n$, then it admits the following moduli

$$R_f(d) \geq \frac{1}{2} \inf_x \langle d, \nabla^2 f(x) d \rangle, \quad S_f(d) \leq \frac{1}{2} \sup_x \langle d, \nabla^2 f(x) d \rangle, \qquad d \in \mathbb{R}^n.$$

Examples

Radial functions Let $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ be a non-decreasing function such that $\alpha(ct) \geq c\alpha(t)$ for all $t \geq 0$ and $c \geq 1$. Define $A(r) = \int_0^r \alpha(u)\, du$, $r \geq 0$, and $f_\alpha(x) = A(\|x\|)$, $x \in \mathbb{R}^n$. Then, the function f_α is R-convex, with $R(d) = 2A(\|d\|/2)$, $d \in \mathbb{R}^n$.

Examples

Radial functions Let $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ be a non-decreasing function such that $\alpha(ct) \geq c\alpha(t)$ for all $t \geq 0$ and $c \geq 1$. Define $A(r) = \int_0^r \alpha(u)\, du$, $r \geq 0$, and $f_\alpha(x) = A(\|x\|)$, $x \in \mathbb{R}^n$. Then, the function f_α is R-convex, with $R(d) = 2A(\|d\|/2)$, $d \in \mathbb{R}^n$.

Continuous gradients Let f be a continuously differentiable function such that ∇f admits a non-decreasing modulus of continuity ω then

$$S(d) \leq 2\|d\|\omega(\|d\|), \qquad d \in \mathbb{R}^n,$$

is a smoothness modulus for f.

Smoothness and subgradient regularity

Proposition

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a function with a non-empty convex domain and $S : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be even. If f is S-smooth, then for all $x_0, x_1 \in \mathrm{dom} f$ and $y_0 \in \partial f(x_0), y_1 \in \partial f(x_1)$, it holds

$$S^*(y_1 - y_0) \leq S(x_1 - x_0).$$

If f is β-smooth, we have $S(d) \leq \frac{\beta}{2}\|d\|^2$ and $S^*(u) \geq \frac{1}{2\beta}\|u\|^2$, thus we get $\|y_1 - y_0\| \leq \beta\|x_1 - x_0\|$, which is the classical Lipschitz property of the gradient.

Moduli and Legendre transform

Proposition [Azé and Penot, 1995]

Let $f, g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be two functions with non-empty convex domains and $S, R : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ be even functions.

(i) If f is S-smooth, then f^* is S^*-convex.

(ii) If g is R-convex, then g^* is R^*-smooth.

Note that when f is α-strongly convex we recover that f^* is α^{-1}-smooth.

Entropic Legendre transform

Proposition

(a) Let $\psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be an R-convex function with convex domain with positive $\mathcal{H}^n$ measure, then the function $\mathcal{L}_{\epsilon,\mathcal{H}^n}(\psi)$ is R^*-smooth.

(b) Let $\psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be an S-smooth function with full domain, then the function $\mathcal{L}_{\epsilon,\mathcal{H}^n}(\psi)$ is S^*-convex.

Recall that

$$\mathcal{L}_{\epsilon,\mathcal{H}^n}(\psi) = \epsilon \log \left(\int e^{\frac{\langle x,y \rangle - \psi(y)}{\epsilon}} \, dy \right)$$

Prekopa-Leindler inequality

Let $f_0, f_1, h : L \to \mathbb{R}_+$ be measurable functions defined on $\mathbb{R}^n$ such that, for some $t \in]0,1[$, it holds

$$h((1-t)y_0 + ty_1) \geq f_0^{1-t}(y_0) f_1^t(y_1), \qquad \forall y_0, y_1 \in \mathbb{R}^n$$

then

$$\int h \geq \left(\int f_0\right)^{1-t} \left(\int f_1\right)^t.$$

Proof sketch of (a)

Set

$$h(y) = \exp\left(\frac{\langle x_t, y\rangle - \psi(y)}{\epsilon}\right), f_0(y) = \exp\left(\frac{\langle x_0, y\rangle - \psi(y)}{\epsilon}\right), f_1(y) = \exp\left(\frac{\langle x_1, y\rangle - \psi(y)}{\epsilon}\right),$$

Proof sketch of (a)

Set

$$h(y) = \exp\left(\frac{\langle x_t, y\rangle - \psi(y)}{\epsilon}\right), f_0(y) = \exp\left(\frac{\langle x_0, y\rangle - \psi(y)}{\epsilon}\right), f_1(y) = \exp\left(\frac{\langle x_1, y\rangle - \psi(y)}{\epsilon}\right),$$

which satisfy

$$\begin{aligned} h(y_t) &\geq \exp\left(\frac{t(1-t)(R(y_1 - y_0) - \langle x_1 - x_0, y_1 - y_0\rangle)}{\epsilon}\right) f_0^{1-t}(y_0) f_1^t(y_1) \\ &\geq \exp\left(\frac{-t(1-t)R^*(x_1 - x_0)}{\epsilon}\right) f_0^{1-t}(y_0) f_1^t(y_1) \end{aligned}$$

because ψ is R-convex.

Proof sketch of (a)

Applying Prekopa-Leindler inequality to h, f_0, f_1 to grants

$$\int h \geq \exp\left(\frac{-t(1-t)R^*(x_1 - x_0)}{\epsilon}\right)\left(\int f_0\right)^{1-t}\left(\int f_1\right)^t .$$

Finally taking the logarithm and multiplying by ϵ gives

$$\mathcal{L}_{\epsilon,\mathcal{H}^n}(\psi)(x_t) \geq -t(1-t)R^*(x_1 - x_0) + (1-t)\mathcal{L}_{\epsilon,\mathcal{H}^n}(\psi)(x_0) + t\mathcal{L}_{\epsilon,\mathcal{H}^n}(\psi)(x_1),$$

which ensures that $\mathcal{L}_{\epsilon,\mathcal{H}^n}(\psi)$ is R^*-smooth.

Regularity estimates for the optimal transport map

Entropic regularity theorem

Theorem

Let $\mu(dx) = e^{-V(x)}dx$ and $\nu(dy) = e^{-W(y)}dy$ be two measures on $\mathbb{R}^n$ with finite second moment such that V is S_V-smooth, $\mathrm{dom}\, V = \mathbb{R}^n$ and W is R_W-convex. Then for all $\epsilon > 0$ the entropic potentials ϕ_ϵ is S-smooth with S such that

$$S(d) \leq \int_0^1 \sup_{R_W^{**}(p) \leq S_V(td)} \langle p, d \rangle \, dt, \qquad \forall d \in \mathbb{R}^n.$$

Note that the result holds for $\epsilon = 0$ by pointwise convergence of ϕ_ϵ towards ϕ.

Remark on radial moduli

In the case $S_V(.) = \sigma_V(\|.\|)$ and $R(.)_W = \rho_W(\|.\|)$, the estimate on the smoothness modulus S can be rewritten as

$$S(\|d\|) \leq \int_0^{\|d\|} (\rho_W^{**})^{-1}(\sigma_V(s))\, ds.$$

Which gives the following regularity estimate on the gradient of ϕ_ϵ

$$\|\nabla\phi_\epsilon(x) - \nabla\phi_\epsilon(y)\| \leq \frac{2}{\|x-y\|} \int_0^{\|x-y\|} (\rho_W^{**})^{-1}(\sigma_V(s))\, ds.$$

Proof of the entropic regularity theorem

Using the notations above the entropic potentials ϕ_ϵ and ψ_ϵ satisfy

$$\phi_\epsilon = \mathcal{L}_{\epsilon,\nu}(\psi) = \mathcal{L}_{\epsilon,\nu}(\psi + \epsilon W) \qquad \psi_\epsilon = \mathcal{L}_{\epsilon,\mu}(\phi) = \mathcal{L}_{\epsilon,\nu}(\phi + \epsilon V).$$

Then by the entropic Legendre transform property, we have that the modulus of smoothness S_ϵ of ϕ_ϵ and the modulus of convexity R_ϵ of ψ_ϵ satisfy

$$S_\epsilon \leq (R_\epsilon + \epsilon R_W)^*\,, \quad R_\epsilon \geq (S_\epsilon + \epsilon S_V)^*\,.$$

Combining the two inequalities and using the inverse monotonicity of the Legendre transform we get

$$S_\epsilon \leq ((S_\epsilon + \epsilon S_V)^* + \epsilon R_W)^* \leq (S_\epsilon + \epsilon S_V)\square\epsilon R_W.$$

Proof of the entropic regularity theorem

Applying the inequality above at $u + \epsilon v$ and using the convexity of S_ϵ grants

$$\langle \partial S_\epsilon(u), v \rangle \leq \frac{S_\epsilon(u + \epsilon v) - S_\epsilon(u)}{\epsilon} \leq S_V(u) + R_W^*(v).$$

Finally optimizing over v we have $R_W^{**}(\partial S_\epsilon(u)) \leq S_V(u)$. The conclusion follows by integration.

Examples

V	W	$S_V(.)$	$R_W(.)$	$\|\nabla\phi(x)-\phi(y)\| \leq$
$\nabla^2 V \leq \alpha_V Id$	$\nabla^2 W \geq \beta_V Id$	$\frac{\alpha_V}{2}\|.\|^2$	$\frac{\beta_W}{2}\|.\|^2$	$\sqrt{\frac{\alpha_V}{\beta_W}}\|x-y\|$
$V \in \mathcal{C}^1, 1 \leq p \leq 2$	$W \in \mathcal{C}^1, 2 \leq q$	$\alpha_V\|.\|^p$	$\beta_W\|.\|^q$	$\frac{2q}{p+q}\left(\frac{\alpha_V}{\beta_W}\right)^{\frac{1}{q}}\|x-y\|^{\frac{p}{q}}$
V L-Lipschitz	$\nabla^2 W = Id$	$4L\|.\|$	$\frac{1}{2}\|.\|^2$	$\frac{8\sqrt{2}}{3}L\|x-y\|^{\frac{1}{2}}$
$V(.) = n\ln(1+\|x\|^2)+C$	$\nabla^2 W = Id$	$6n\|.\|^2 \wedge 4n\|.\|$	$\frac{1}{2}\|.\|^2$	$2\sqrt{3n}\|x-y\| \wedge \frac{8\sqrt{2}}{3}\|x-y\|^{\frac{1}{2}}$
$\nabla^2 V \leq A^{-1}$	$\nabla^2 W \geq B^{-1}$	A^{-1}	B^{-1}	$B^{1/2}\left(B^{-1/2}A^{-1}B^{-1/2}\right)^{1/2}B^{1/2}$

Corollary

Assume that V is S-smooth with $S(.) = \sigma(\|.\|)$ where $\sigma : \mathbb{R}_+ \to \mathbb{R}_+$ non-decreasing and that W is R-convex with $R(.) = \rho(\|.\|)$ where $\rho : \mathbb{R}_+ \to \mathbb{R} \cup \{+\infty\}$. Then the optimal transport T from μ to ν satisfies

$$\|T(x)\| \leq \|T(0)\| + 2(\rho^{**})^{-1}(\sigma(\|x\|)).$$

Note that it is not required to have $\rho \geq 0$.

Growth estimates

Corollary

Assume that V is S-smooth with $S(.) = \sigma(\|.\|)$ where $\sigma : \mathbb{R}_+ \to \mathbb{R}_+$ non-decreasing and that W is R-convex with $R(.) = \rho(\|.\|)$ where $\rho : \mathbb{R}_+ \to \mathbb{R} \cup \{+\infty\}$. Then the optimal transport T from μ to ν satisfies

$$\|T(x)\| \leq \|T(0)\| + 2(\rho^{**})^{-1}(\sigma(\|x\|)).$$

Note that it is not required to have $\rho \geq 0$.

This estimate is of the same order as the one obtained in [Fathi, 2024]. It is strictly better in the case of a rotationnaly invariant target measure.

Conclusion

(i) This approach also allows us to recover the recent result on a log-subharmonic "contraction" theorem when the target is log-concave [Philippis and Shenfeld, 2024].

(ii) It is possible to deduce results for the entropic Legendre transform with a general cost. However the application to optimal transport is unclear.

(iii) The entropic Legendre transform result is to be compared with the result on weak semiconvexity of Schrödinger potentials [Conforti, 2024].

Thank you for your attention !

References i

Azé, D. and Penot, J.-P. (1995).
Uniformly convex and uniformly smooth convex functions.
Ann. Fac. Sci. Toulouse, Math. (6), 4(4):705–730.

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

Caffarelli, L. A. (1992).
The regularity of mappings with a convex potential.
J. Am. Math. Soc., 5(1):99–104.

Carlier, G., Duval, V., Peyré, G., and Schmitzer, B. (2017).
Convergence of entropic schemes for optimal transport and gradient flows.
SIAM Journal on Mathematical Analysis, 49(2):1385–1418.

Chewi, S. and Pooladian, A.-A. (2023).
An entropic generalization of caffarelli's contraction theorem via covariance inequalities.
Comptes Rendus. Mathématique, 361(G9):1471–1482.

Conforti, G. (2024).
Weak semiconvexity estimates for schrödinger potentials and logarithmic sobolev inequality for schrödinger bridges.
Probability Theory and Related Fields, 189(3–4):1045–1071.

Fathi, M. (2024).
Growth estimates on optimal transport maps via concentration inequalities.

Fathi, M., Gozlan, N., and Prod'homme, M. (2020).
A proof of the Caffarelli contraction theorem via entropic regularization.
Calc. Var. Partial Differ. Equ., 59(3):18.
Id/No 96.

Nutz, M. and Wiesel, J. (2021).
Entropic optimal transport: convergence of potentials.
Probability Theory and Related Fields, 184(1–2):401–424.

Philippis, G. D. and Shenfeld, Y. (2024).
Optimal transport maps, majorization, and log-subharmonic measures.