Disentangling entropy and suboptimality in Entropic optimal transport

Hugo Malamut and Maxime Sylvestre
(Université Paris Dauphine PSL, CEREMADE)

March 2, 2025

The regularized optimal transport

Let $\mu_0, \mu_1 \in \mathcal{P}_{ac}(\mathbb{R}^d)$ such that $H(\mu_i \mid \mathcal{H}^d) < +\infty$.
For $\varepsilon \geq 0$

$$OT_\varepsilon(\mu_0, \mu_1) := \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int c d\gamma + \varepsilon H(\gamma | \mathcal{H}^{2d}) \tag{εEOT}$$

where c is a C^2 cost function and for any measure m :

$$H(\gamma \mid m) = \int \frac{d\gamma}{dm}(x) \ln\left(\frac{d\gamma}{dm}(x)\right) dx \tag{1}$$

Γ-convergence towards $OT_0(\mu, \nu)$ as $\varepsilon \to 0$ for $c(x, y) = \frac{1}{2}\|x - y\|^2$ [CDPS15].

Convergence of the value

Proposition [ADPZ11][EMR15]

Assume $c(x, y) = \frac{1}{2}\|x - y\|^2$, and that μ_i are compactly supported with finite Fisher information then

$$OT_\varepsilon = \frac{1}{2}W_2^2 - \frac{d}{2}\varepsilon ln(2\pi\varepsilon) + \varepsilon\frac{H(\mu_0 \mid \mathcal{H}^d) + H(\mu_1 \mid \mathcal{H}^d)}{2} + o(\varepsilon) \qquad \text{(TE-OT}_\varepsilon\text{)}$$

Proposition [EN22, CPT22]

Assume c is infinitesimally twisted, $H(\mu_i \mid \mathcal{H}^d) < +\infty$ and μ_i are compactly supported then

$$\left(-\frac{d}{2}\varepsilon \ln(\varepsilon) + C'\varepsilon \leq\right) OT_\varepsilon - OT_0 \leq -\frac{d}{2}\varepsilon \ln(\varepsilon) + C\varepsilon \qquad (2)$$

Convergence of the minimizers

When the optimal transport plan γ_0 is unique then $W_2(\gamma_\varepsilon, \gamma_0) \to 0$. Then two natural questions arise:

- Can we find an expansion for $\int c d\gamma_\varepsilon$ and $H(\gamma_\varepsilon \mid \mathcal{H}^{2d})$ as $\varepsilon \to 0$?
- Is there a rate of convergence for $W_2(\gamma_\varepsilon, \gamma_0)$?

Prior Works

Qualitative convergence results.

- Γ-convergence : [Mik04],[MT08],[Lé13],[CDPS15]

Quantitative convergence results.

- Discrete optimal transport : [CM94]
- Semi-discrete optimal transport : [ANWS21],[Del21]
- Finite Fisher information : [ADPZ11],[EMR15],[Con19]
- Finite entropy : [Pal19],[EN22],[CPT22]
- Multimarginal : [NP23]

Fisher information and quadratic cost

Main result

Theorem

Suppose that the cost is quadratic, that is $c(x,y) = \frac{1}{2}\|x-y\|^2$. Further assume that $I(\mu_i) < \infty$ and $Supp(\mu_i)$ compact. Then

$$H(\gamma_\varepsilon \mid \mathcal{H}^{2d}) = -\frac{d}{2} ln(2\pi\varepsilon) + H_m - \frac{d}{2} + o(1) \tag{3}$$

where $H_m = (H(\mu_0) + H(\mu_1))/2$. Moreover

$$(c, \gamma_\varepsilon) = OT_0 + \frac{d}{2}\varepsilon + o(\varepsilon) \tag{4}$$

Recall that

$$I(\mu) = \int \frac{\|\nabla \mu(x)\|^2}{\mu(x)} dx \tag{5}$$

Link with Sinkhorn divergences

It is astoninshing that the first order term in the Taylor expansion of the suboptimality does not depend on the marginals. The estimator $(c, \gamma_\varepsilon) - \frac{d}{2}\varepsilon$ is thus of precision similar to the Sinkhorn divergences [FSV$^+$18, CRL$^+$20]:

$$OT_\varepsilon(\mu_0, \mu_1) - \frac{1}{2}\left(OT_\varepsilon(\mu_0, \mu_0) + OT_\varepsilon(\mu_1, \mu_1)\right) \tag{6}$$

Sketch of proof

The Benamou-Brenier [Lé13] formulation of the problem is the following

$$(\varepsilon EOT) = \varepsilon H_m - \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) + \min_{\rho,v} \int_0^1 \int \frac{1}{2}|v_t|^2 d\rho_t(x)dt + \frac{\varepsilon^2}{8}\int_0^1 \int \frac{\|\nabla \rho_t\|^2}{\rho_t} dxdt \tag{εBB}$$

Sketch of proof

The Benamou-Brenier [Lé13] formulation of the problem is the following

$$(\varepsilon EOT) = \varepsilon H_m - \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) + \min_{\rho,v} \int_0^1 \int \frac{1}{2}|v_t|^2 d\rho_t(x)dt + \frac{\varepsilon^2}{8}\int_0^1 \int \frac{\|\nabla \rho_t\|^2}{\rho_t} dxdt \tag{εBB}$$

Recall that from (TE-OT$_\varepsilon$) we have

$$(\varepsilon EOT) - \frac{1}{2}W_2^2(\mu_0, \mu_1) + \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) - \varepsilon H_m = o(\varepsilon) \tag{7}$$

Sketch of proof

The Benamou-Brenier [Lé13] formulation of the problem is the following

$$(\varepsilon EOT) = \varepsilon H_m - \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) + \min_{\rho,v} \int_0^1 \int \frac{1}{2}|v_t|^2 d\rho_t(x)dt + \frac{\varepsilon^2}{8}\int_0^1 \int \frac{\|\nabla \rho_t\|^2}{\rho_t} dxdt \tag{εBB}$$

Recall that from (TE-OT$_\varepsilon$) we have

$$(\varepsilon EOT) - \frac{1}{2}W_2^2(\mu_0,\mu_1) + \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) - \varepsilon H_m = o(\varepsilon) \tag{7}$$

Thus thanks to (εBB)

$$\min_{\rho,v} \frac{1}{\varepsilon}\left(\int_0^1 \int \frac{1}{2}|v_t|^2 d\rho_t(x)dt - \frac{1}{2}W_2^2(\mu_0,\mu_1)\right) + \frac{\varepsilon}{8}\int_0^1 \int \frac{\|\nabla \rho_t\|^2}{\rho_t} dxdt = o(1) \tag{8}$$

Since both terms are positive they both tend to 0.

From dynamic to static and back

Using the envelope theorem on the static and dynamic formulation we get the following set of identities

$$\begin{cases} (c, \gamma_\varepsilon) &= \displaystyle\int_0^1 \int \frac{1}{2} |v_t^\varepsilon|^2 d\rho_t(x) dt - \frac{\varepsilon^2}{8} \int_0^1 I(\rho^\varepsilon) dt + \frac{d}{2}\varepsilon \\ H(\gamma_\varepsilon) &= \displaystyle\frac{\varepsilon}{4} \int_0^1 I(\rho^\varepsilon) dt - \frac{d}{2} \ln(2\pi\varepsilon) + H_m - \frac{d}{2} \end{cases} \tag{9}$$

Quadratic cost without Fisher information

Main result

Theorem

Suppose that the cost is quadratic, that is $c(x, y) = \frac{1}{2}\|x - y\|^2$. Further assume that μ_i have finite moment of order $2 + \delta$ then

$$(c, \gamma_\varepsilon) = OT_0 + \Theta(\varepsilon), \quad H(\gamma_\varepsilon \mid \mathcal{H}^{2d}) = -\frac{d}{2}\ln(\varepsilon) + O(1), \quad \sqrt{\varepsilon} = O(W_2(\gamma_\varepsilon, \gamma_0)) \tag{10}$$

In the special case where the Monge map ∇f associated to the optimal transport plan γ_0 is Lipschitz then

$$W_2(\gamma_\varepsilon, \gamma_0) = \Theta(\sqrt{\varepsilon}) \tag{11}$$

Disentangling

Once again the goal is to find an upperbound on the sum of two "positive" parts : the "entropy" and the suboptimality. Note that (εEOT) satisfies

$$-\frac{d}{2}\varepsilon \ln(\varepsilon) + C\varepsilon \geq (\varepsilon EOT) - (OT) \geq \int cd\gamma_\varepsilon - \int cd\gamma_0 + \varepsilon H(\gamma_\varepsilon)$$

Introduce the gap function $E = c - \varphi - \psi = f + f^* - \langle ., . \rangle \geq 0$ where φ, ψ are the Kantorovich potentials and f the Brenier potential of the optimal transport problem. Thus $\int cd\gamma_0 = \int \varphi + \psi d\gamma_0$ which in turns grant:

$$-\frac{d}{2}\varepsilon \ln(\varepsilon) + C\varepsilon \geq \int Ed\gamma_\varepsilon + \varepsilon \inf_{\int Ed\gamma \leq \int Ed\gamma_\varepsilon} H(\gamma) \tag{12}$$

Minimizing entropy under energy constraint

Definition

Let $G : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$. We say that G has a *quadratic detachment* if for any $u \in \mathbb{R}^d$ it exists $v_u \in \mathbb{R}^d$ such that

$$\forall (u, v) \in \mathbb{R}^{2d} \quad G(u, v) \geq \frac{1}{2}\|v - v_u\|^2 \tag{13}$$

Minty's trick grants FAIRE UN DESSIN?

$$E(x, y) = f(x) + f^*(y) - \langle x, y\rangle \geq \|x - (Id + \partial f)^{-1}(x + y)\|^2 \tag{14}$$

Which after the change of variable $u = \frac{1}{\sqrt{2}}(x + y), v = \frac{1}{\sqrt{2}}(y - x)$ gives a quadratic detachment for E:

$$E(u, v) \geq \|v - S(u)\|^2 \tag{15}$$

Quadratic detachment and entropy

Proposition

Let G be a function on $\mathbb{R}^d \times \mathbb{R}^d$ and $\gamma \in \mathcal{P}_{ac}(\mathbb{R}^{2d})$ and denote $C_d := -\frac{d}{2}\ln(\frac{4\pi e}{d})$. If G has a quadratic detachment then

$$H(\gamma \mid \mathcal{H}^{2d}) \geq -\frac{d}{2}\ln\left(\int G d\gamma\right) + H(\gamma_1 \mid \mathcal{H}^d) + C_d \tag{16}$$

Where γ_1 is the projection of γ on the first coordinate, ie the first marginal of γ, and where H is the differential entropy.

Sketch of proof

If we disintegrate γ with respect to the projection on the first variable into $\gamma_1 \otimes \gamma_x$ the quadratic detachment of G controls the variance of γ_x.

$$\int 2G(x,.)d\gamma_x \geq \int \|y - y_x\|^2 d\gamma_x \geq \mathrm{Var}(\gamma_x) \tag{17}$$

Thus by minimality of the gaussian under variance constraint for the entropy we have $H(\gamma_x) \geq -\frac{d}{2} \ln \left(\int G d\gamma_x \right) + C_d$. It remains to integrate and use the additivity of entropy.

$$H(\gamma) = H(\gamma_1) + \int H(\gamma_x) d\gamma_1 \tag{18}$$

Expansion of the entropy and suboptimality

Since E has a quadratic detachment after Minty's change of coordinate we can apply last proposition to the rotated transport plan $\hat{\gamma}_\varepsilon(u,v) = \gamma_\varepsilon(\frac{1}{\sqrt{2}}(u-v), \frac{1}{\sqrt{2}}(u+v))$. Since the change of variable is measure preserving we have

$$H(\gamma_\varepsilon) = H(\hat{\gamma}_\varepsilon) \geq -\frac{d}{2}\ln\left(\int Ed\gamma_\varepsilon\right) + H(\pi_u\hat{\gamma}_\varepsilon) + C_d \tag{19}$$

Combining it with the upper bound on $(OT_\varepsilon) - (OT)$ we have

$$-\frac{d}{2}\varepsilon\ln(\varepsilon) + C\varepsilon \geq \int Ed\gamma_\varepsilon - \frac{d}{2}\varepsilon\ln(\int Ed\gamma_\varepsilon) + C'\varepsilon \tag{20}$$

Expansion of the entropy and suboptimality

The last inequality rewrites as

$$C \geq \frac{\int E d\gamma_\varepsilon}{\varepsilon} - \frac{d}{2} \ln \left(\frac{\int E d\gamma_\varepsilon}{\varepsilon} \right) \tag{21}$$

or $x \mapsto x - \frac{d}{2} ln(x)$ diverges near 0 and ∞ thus

$$(c, \gamma_\varepsilon) = OT_0 + \Theta(\varepsilon), \quad H(\gamma_\varepsilon \mid \mathcal{H}^{2d}) = -\frac{d}{2} \ln(\varepsilon) + O(1) \tag{22}$$

Slicing lemma

Lemma

Let $T : \mathbb{R}^d \to \mathbb{R}^d$ an L-Lipschitz map. Let $\mu = (Id \times T)_{\#}\mu_0$ with μ_0 a probability on $\mathbb{R}^d$ with finite moment of order 2. For any $\nu \in \mathcal{P}_{2,ac}(\mathbb{R}^d)$ we have:

$$W_2^2(\mu, \nu) \geq (1 - L)W_2^2(\mu_0, \nu_0) + \frac{1}{L+1}\int Var(\nu_x)d\nu_0(x) \tag{23}$$

In particular

$$(1 + L)^2 W_2^2(\mu, \nu) \geq \int Var(\nu_x)d\nu_0(x) \tag{24}$$

By a similar argument we have

$$H(\nu \mid \mathcal{H}^{2d}) \geq H(\nu_0 \mid \mathcal{H}^d) - \frac{d}{2}\ln\left(W_2^2(\mu, \nu)\right) + C \tag{25}$$

Upper bound on the Wasserstein distance

$S : (x, y) \mapsto (x, \nabla f(x))$ is a transport from γ_ε to γ_0 thus

$$W_2^2(\gamma_\varepsilon, \gamma_0) \leq \int \|y - \nabla f(x)\|^2 d\gamma_\varepsilon(x, y) \tag{26}$$

However since ∇f is Lipschitz an argument by [LN20, Ber20, Gig11] ensures

$$\|y - \nabla f(x)\|^2 \leq 2L\left(f(x) + f^*(y) - \langle x, y\rangle\right) \leq 2LE(x, y) \tag{27}$$

Integration grants $W_2^2(\gamma_\varepsilon, \gamma_0) \leq 2L \int E d\gamma_\varepsilon \leq C\varepsilon$.

Infinitesimally twisted costs and compact supports

Main result

Definition

$c \in \mathcal{C}^2(\Omega^2)$ is said to be infinitesimally twisted if $\nabla^2_{xy} c(x,y) = (\partial^2_{x_i y_j} c(x,y))_{i,j} \in M_d(\mathbb{R})$ is invertible for every $(x,y) \in \Omega^2$.

Theorem

Suppose that the cost is $\mathcal{C}^2$ and infinitesimally twisted . Further assume that μ_i is compactly supported then

$$(c, \gamma_\varepsilon) = OT_0 + \Theta(\varepsilon), \quad H(\gamma_\varepsilon \mid \mathcal{H}^{2d}) = -\frac{d}{2}\ln(\varepsilon) + O(1), \quad \sqrt{\varepsilon} = O(W_2(\gamma_\varepsilon, \gamma_0)) \tag{28}$$

Note that here γ_0 is any optimal transport plan.

Local quadratic detachment

Control of the entropy

Lemma

Let $X \subset \mathbb{R}^d$ compact. $E : X \times X \to \mathbb{R}_+$ continuous with a local quadratic detachment as before then for any $\gamma \in \mathcal{P}(X \times X)$

$$H(\gamma \mid \mathcal{H}^{2d}) \geq -\frac{d}{2} \ln \left(\int E d\gamma \right) + C$$

Gluing of local properties

Lemma[CPT22]

For c infinitesimally twisted, (φ, ψ) a pair of c-conjugate functions. Then $E := c - \varphi - \psi$ has a local quadratic detachment.

Thus the same procedure grants the result on the entropy and the suboptimality.

Lemma [MPW12]

Any optimal transport plan γ_0 is locally supported on the graph of Lipschitz functions.

Using locally the slicing lemma grants the lower bound on the Wasserstein distance between γ_ε and γ_0.

Further questions

- Next order term in the Taylor expansion. H_m for quadratic cost -> General geometric value?
- Upper bound for the Wasserstein distance? Seem to depend on the regularity of the optimal transport plan.
- Other rates of detachment for the Gap function.
- Other problems involving entropy? Entropic mulitmarginal OT, Free energy with temperature, ...

Thank you !

References i

Stefan Adams, Nicolas Dirr, Mark A. Peletier, and Johannes Zimmer.
From a large-deviations principle to the wasserstein gradient flow: A new micro-macro passage.
Communications in Mathematical Physics, 307(3):791–815, September 2011.

Jason M. Altschuler, Jonathan Niles-Weed, and Austin J. Stromme.
Asymptotics for semidiscrete entropic optimal transport.
SIAM Journal on Mathematical Analysis, 54(2):1718–1741, mar 2021.

Robert J. Berman.
Convergence rates for discretized monge–ampère equations and quantitative stability of optimal transport.
Foundations of Computational Mathematics, 21(4):1099–1140, December 2020.

Guillaume Carlier, Vincent Duval, Gabriel Peyré, and Bernhard Schmitzer.
Convergence of entropic schemes for optimal transport and gradient flows, 2015.

R. Cominetti and J. San Martin.
Asymptotic analysis of the exponential penalty trajectory in linear programming.
Mathematical Programming, 67(1-3):169–187, October 1994.

Giovanni Conforti.
A second order equation for schrödinger bridges with applications to the hot gas experiment and entropic transportation cost.
Probability Theory and Related Fields, 174(1-2):1–47, 2019.

- Guillaume Carlier, Paul Pegon, and Luca Tamanini.
  Convergence rate of general entropic optimal transport costs.
  arXiv preprint arXiv:2206.03347, 2022.
- Lenaic Chizat, Pierre Roussillon, Flavien Léger, François-Xavier Vialard, and Gabriel Peyré.
  Faster wasserstein distance estimation with the sinkhorn divergence.
  Advances in Neural Information Processing Systems, 33:2257–2269, 2020.
- Alex Delalande.
  Nearly tight convergence bounds for semi-discrete entropic optimal transport, 2021.

References iv

Matthias Erbar, Jan Maas, and Michiel Renger.
From large deviations to wasserstein gradient flows in multiple dimensions.
Electronic Communications in Probability, 20:1–12, 2015.

Stephan Eckstein and Marcel Nutz.
Convergence rates for regularized optimal transport via quantization, 2022.

Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun ichi Amari, Alain Trouvé, and Gabriel Peyré.
Interpolating between optimal transport and mmd using sinkhorn divergences, 2018.

References v

Nicola Gigli.
On hölder continuity-in-time of the optimal transport map towards measures along a curve.
Proceedings of the Edinburgh Mathematical Society, 54(2):401–409, March 2011.

Wenbo Li and Ricardo H Nochetto.
Quantitative stability and error estimates for optimal transport plans.
IMA Journal of Numerical Analysis, 41(3):1941–1965, July 2020.

Christian Léonard.
A survey of the schrödinger problem and some of its connections with optimal transport, 2013.

Toshio Mikami.
Monge's problem with a quadratic cost by the zero-noise limit of h-path processes.
Probability Theory and Related Fields, 129(2):245–260, March 2004.

Robert J McCann, Brendan Pass, and Micah Warren.
Rectifiability of optimal transportation plans.
Canadian Journal of Mathematics, 64(4):924–934, 2012.

Toshio Mikami and Michèle Thieullen.
Optimal transportation problem by stochastic optimal control.
SIAM Journal on Control and Optimization, 47(3):1127–1139, January 2008.

Luca Nenna and Paul Pegon.
Convergence rate of entropy-regularized multi-marginal optimal transport costs, 2023.

Soumik Pal.
On the difference between entropic cost and the optimal transport cost, 2019.