INTERNSHIP REPORT : QUANTITATIVE STABILITY IN OPTIMAL TRANSPORT FOR GENERAL POWER COSTS

OCTAVE MISCHLER

1. Presentation of the internship

This document presents the studies and the work realized during my internship of second year of the Ecole Normale Supérieure. The internship took place from February to June 2024 under the supervision of Dario Trevisan, Associate Professor at the Mathematics Department of the University of Pisa (IT). During this four months internship I attended the lectures of Luigi Ambrosio on optimal transport at the Scuola Normale Superiore di Pisa and a lecture group on Stochastic Differential Equations with Mario Maurelli at the University of Pisa. Simultaneously I worked on different research subjects with Dario Trevisan. Upon my arrival, I studied the convergence of the entropic optimal transport problem to classical optimal transport as the regularization goes to zero. After a few weeks, I turned to the improvement of quantitative stability estimates. We first tried other convex regularization than the entropic one, before turning to the ideas and results presented in the last part of this report. During my stay, with my supervisor I met and discussed about research topics with Luigi Ambrosio and Aldo Pratelli in Pisa, Simone Di Marino and Augusto Gerolin in Genova.

2. General introduction to the goal

The optimal transport problem was first introduced by Monge in 1781 in the study of the optimal way of filling holes starting from piles of sand. The problem was left unsolved until 1942, when Kantorovitch reformulated Monge's problem. Monge's original problem was first solved in 1976 by Sudakov, whose proof was later corrected in 2003 by Ambrosio. Since the end of the 90s, optimal transport has been intensively studied for its links with PDEs and lately for its applications in machine learning.

The Monge problem for the cost $|.|^p$ between two probability measures $\rho \in \mathcal{P}(\mathcal{X})$, $\mu \in \mathcal{P}(\mathcal{Y})$ with $\mathcal{X}$, $\mathcal{Y} \subset \mathbb{R}^d$ compact, is defined as the following minimization problem

$$\mathcal{M}_p(\rho,\mu) := \inf\{\int_{\mathcal{X}} |T(x)-x|^p d\rho(x) | T : \mathcal{X} \to \mathcal{Y} \text{ with } T_\sharp\rho = \mu\} \tag{2.1}$$

where $T_\sharp\rho$ denotes the push-forward of ρ through T and $|\cdot|$ denotes the Euclidean norm. It was originally introduced by Monge for $p = 1$. Let us also recall that the push-forward $T_\sharp\rho \in \mathcal{P}(\mathcal{Y})$ of a measure $\rho \in \mathcal{P}(\mathcal{X})$ through $T : \mathcal{X} \to \mathcal{Y}$ is defined as $T_\sharp\rho(B) = \rho(T^{-1}(B))$ for any ρ-measurable $B \subset \mathcal{Y}$. Kantorovitch introduced the following relaxation :

{eq:Wp}

$$\mathcal{W}_p(\rho,\mu)^p := \inf_{\pi\in\Pi(\rho,\mu)} \int_{\mathcal{X}\times\mathcal{Y}} |y-x|^p d\pi(x,y). \tag{2.2}$$

where $\Pi(\rho,\mu)$ denotes the set of couplings between ρ and μ, i.e., (joint) probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals distributions ρ, μ. It is indeed a relaxation of the previous, since $(Id \times T)_\sharp\rho \in \Pi(\rho,\mu)$ when $T_\sharp\rho = \mu$. A major improvement of this relaxation is that an optimizer always exists for (2.2). $\mathcal{W}_p(\rho,\mu)$ is called Wasserstein distance between ρ and μ.

Date: August 15, 2024.

The first major result of the optimal transport theory, is to exploit convex duality and obtain the equivalent formulation

$$\mathcal{W}_p(\rho,\mu)^p = \sup_{(\phi,\psi)} \left\{ \int_{\mathcal{X}} \phi d\rho + \int_{\mathcal{Y}} \psi d\mu \right\}, \tag{2.3}$$

where the supremum runs among upper semi-continuous functions (ϕ,ψ) such that $\phi(x) + \psi(y) \le |x-y|^p$ for every $x \in \mathcal{X}$, $y \in \mathcal{Y}$. An optimizing pair (ϕ,ψ) is called an optimal (or Kantorovich) potential for $\mathcal{W}_p(\rho,\mu)$. Under minimal assumptions, one can always argue that an optimizer exists.

A second major result of the optimal transport theory is that under mild assumptions, in particular if ρ is absolutely continuous with respect to the Lebesgue measure and $p > 1$, one can argue that the optimizer in (2.2) is unique and given by a optimal transport map T, i.e., one has $y = T(x)$ for π-a.e. (x,y) (with necessarily $T_\sharp \rho = \mu$) and therefore

$$\mathcal{W}_p(\rho,\mu)^p = \int_{\mathcal{X}} |T(x) - x|^p d\rho(x). \tag{2.4}$$

For the quadratic cost $p = 2$, the optimal transport map is often referred to as Brenier's map.

A well-posed mathematical problem in the Hadamard sense is one for which existence and uniqueness of solutions hold, but also stability with respect to perturbations in the data. Although qualitative stability in most situations emerges as a by-product of existence and uniqueness arguments, quantitative stability becomes particularly crucial in applications, for it directly relates to the convergence of numerical methods. This applies in particular to optimal transport problems.

Aim of this work is to give a short introduction to optimal transport, highlighting the key ideas and avoiding technicalities, before presenting new quantitative stability results for optimal transport potentials with respect to perturbations of the "target" measure μ, keeping one "source" measure ρ fixed. In a slightly simplified form, to ease the notation, this goal can be summarized in the following result.

Theorem 2.1. *Let $d \ge 1$, ρ be a* log*-concave probability measure on $\mathbb{R}^d$ with bounded support, let $\mathcal{Y} \subseteq \mathbb{R}^d$ be compact and $p > 1$. Then, there exists $C = C(\rho,\mathcal{Y},p) < \infty$ such that, for any μ, ν, probability measures supported on $\mathcal{Y}$, it holds*

$$\|\phi_\mu - \phi_\nu\|_{L^2(\rho)} \le C \mathcal{W}_1(\mu,\nu)^\theta, \tag{2.5}$$

with

$$\theta = \begin{cases} 1 - \frac{1}{p} & \text{if } 1 < p < 2, \\ \frac{1}{2} & \text{if } p \ge 2, \end{cases} \tag{2.6}$$

and where (ϕ_μ,ψ_μ) denotes the Kantorovich potentials for $\mathcal{W}_p(\rho,\mu)$, i.e., the optimizer in (2.3), with zero ρ-mean, i.e., $\int \phi_\mu d\rho = 0$ (and similarly (ϕ_ν,ψ_ν), for $\mathcal{W}_p(\rho,\nu)$).

In Section 3 we start by giving the most classical and important results of the optimal transport theory. For this part our sources are the usual references on optimal transport such as [AGS08] and [Vil+09]. Starting from Section 4, we dive into more specific results necessary for Section 5. This last two parts are mostly extracted from my article with Dario Trevisan [MT24].

3. Introduction to optimal transport, various results

3.1. Optimal transport, definitions and existence.

Definition 3.1. Let $\mathcal{X}$, $\mathcal{Y}$ be Polish spaces and let $c : \mathcal{X} \times \mathcal{Y} \to [0,\infty]$ be continuous. Given $\rho \in \mathcal{P}(\mathcal{X})$ and $\mu \in \mathcal{P}(\mathcal{Y})$, define

$$\Pi(\rho,\mu) := \{\pi \in \mathcal{P}(\mathcal{X}\times\mathcal{Y}) | \pi(A \times \mathcal{Y}) = \rho(A), \pi(\mathcal{X} \times B) = \mu(B) \quad \forall \text{ Borel } A, B\} \tag{3.1}$$

Kantorovitch's formulation of the optimal transport problem asks to find

$$\mathcal{T}^c(\rho,\mu) := \inf_{\pi \in \Pi(\rho,\mu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y) d\pi(x,y) \tag{OT$_c$}$$

Remark 3.2. (1) $\Pi(\rho,\mu)$ is said to be the set of couplings between ρ and μ. If we note $p_{\mathcal{X}} : \mathcal{X}\times\mathcal{Y}\to\mathcal{X}$ (resp. $p_{\mathcal{Y}} : \mathcal{X}\times\mathcal{Y}\to\mathcal{Y}$) the projection on $\mathcal{X}$ (resp. $\mathcal{Y}$), we can also write

$$\Pi(\rho,\mu) := \{\pi\in\mathcal{P}(\mathcal{X}\times\mathcal{Y}) | (p_{\mathcal{X}})_{\#}\,\pi = \rho, (p_{\mathcal{Y}})_{\#}\,\pi = \mu\} \tag{3.2}$$

where $(p_{\mathcal{X}})_{\#}\,\pi$ is the push-forward measure through $p_{\mathcal{X}}$, i.e. $(p_{\mathcal{X}})_{\#}\,\pi(A) = \pi(A\times\mathcal{Y})$ $\forall$ Borel A.

(2) When $\mathcal{X}$, $\mathcal{Y}$ are compact subsets of $\mathbb{R}^d$ and c is continuous, it holds $\mathcal{T}^c(\rho,\mu)<\infty$ for every $\rho\in\mathcal{P}(\mathcal{X})$, $\mu\in\mathcal{P}(\mathcal{Y})$.

(3) When $p\geq 1$ and c is the so-called p-cost, $c(x,y)=\frac{1}{p}|x-y|^p$ for $x\in\mathcal{X}$, $y\in\mathcal{Y}$, we write $\mathcal{T}^p$ instead of $\mathcal{T}^c$ and notice that

{pb:otp}
$$\mathcal{T}^p(\rho,\mu) = \frac{1}{p}\mathcal{W}_p(\rho,\mu)^p. \tag{3.3}$$

When $p=2$, writing $|x-y|^2 = |x|^2+|y|^2-2\langle x,y\rangle$, we collect the identity

{eq:quadratic}
$$\mathcal{T}^2(\rho,\mu) = \frac{1}{2}\int_{\mathcal{X}}|x|^2 d\rho(x) + \frac{1}{2}\int_{\mathcal{Y}}|y|^2 d\mu(y) + \mathcal{T}'(\rho,\mu), \tag{3.4}$$

where $'$ denotes the linear cost $(x,y)\mapsto -\langle x,y\rangle$.

thm:existK
Theorem 3.3 (Existence of optimal plans.)**.** *Let $\mathcal{X}$, $\mathcal{Y}$ be Polish spaces and let $c:\mathcal{X}\times\mathcal{Y}\to[0,\infty]$ be continuous. Given $\rho\in\mathcal{P}(\mathcal{X})$ and $\mu\in\mathcal{P}(\mathcal{Y})$, the infimum in* ($\mathrm{OT}_c$) *is a minimum.*

Proof of Theorem 3.3. The proof follows an elementary scheme from calculus of variations. We consider the weak topology on $\mathcal{M}(\mathcal{X}\times\mathcal{Y})$ and we divide the proof in the two following steps :

(1) the application $\mathcal{C} : \pi\in\Pi(\mathcal{X}\times\mathcal{Y}) \to \int\limits_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\pi(x,y)$, extended to $\mathcal{M}(\mathcal{X}\times\mathcal{Y})$ by $+\infty$ outside $\Pi(\mathcal{X}\times\mathcal{Y})$, is lower semi-continuous for the weak topology on $\mathcal{M}(\mathcal{X}\times\mathcal{Y})$,

(2) the set $\Pi(\mathcal{X}\times\mathcal{Y})$ is compact for the weak topology on $\mathcal{M}(\mathcal{X}\times\mathcal{Y})$.

Before proving these two facts let us conclude the proof. Considering a minimizing sequence of probabilities $(\pi_n)_{n\in\mathbb{N}}$ in (OT_c), from the compacity of $\Pi(\mathcal{X}\times\mathcal{Y})$ we extract a weakly converging subsequence $\pi_{n_k}\rightharpoonup\pi^*\in\Pi(\rho,\mu)$. Now the lower semi-continuity yields $\mathcal{C}(\pi^*)\leq\liminf_{k\to\infty}\mathcal{C}(\pi_{n_k}) = \mathcal{T}^c(\rho,\mu)$ since $(\pi_n)_n$ is a minimizing sequence. π^* is thus a minimizing in (OT_c).

Proof of the lower semi-continuity of $\mathcal{C}$. For $k\in\mathbb{N}$, let us define the regularized cost

$$c_k(x,y) = \inf_{x'\in\mathcal{X},y'\in\mathcal{Y}} c(x',y') + kd_{\mathcal{X}}(x,x') + kd_{\mathcal{Y}}(y,y') \tag{3.5}$$

which satisfy $0\leq c_k\leq c_{k+1}\leq c\wedge k$. Moreover, since c_k is the pointwise infimum of a family of equi-Lipschitz functions, one has $c_k\in\mathrm{Lip}_b(\mathcal{X}\times\mathcal{Y})$.
The lower semi-continuity of c ensures $c_k\to c$ as $k\to\infty$. Indeed, $\sup_k c_k\leq c$ is immediate and if $(x,y)\in\mathcal{X}\times\mathcal{Y}$ with $\sup_k c_k<\infty$, for all k it exists $x_k,y_k\in\mathcal{X}\times\mathcal{Y}$ such that

$$c(x_k,y_k)\wedge k + kd_{\mathcal{X}}(x,x_k) + kd_{\mathcal{Y}}(y,y_k) \leq c_k(x,y) + \frac{1}{k}. \tag{3.6}$$

From this inequality we immediately get $x_k\to x$ and $y_k\to y$ as $k\to\infty$. Moreover, from $c(x_k,y_k)\wedge k\leq c_k(x,y)+\frac{1}{k}$ and the lower semi-continuity of c it follows that $c(x,y)\leq\sup_k c_k(x,y)$.
Now, if $(\pi_n)_n$ weakly converge to $\pi\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$, for any $k\in\mathbb{N}$ one has

$$\liminf_{n\to\infty}\mathcal{C}(\pi_n) \geq \liminf_{n\to\infty}\int_{\mathcal{X}\times\mathcal{Y}} c_k d\pi_n = \int_{\mathcal{X}\times\mathcal{Y}} c_k d\pi \tag{3.7}$$

Finally, taking the limit $k\to\infty$, the monotone convergence theorem yields

$$\liminf_{n\to\infty}\mathcal{C}(\pi_n) \geq \sup_k\int_{\mathcal{X}\times\mathcal{Y}} c_k d\pi = \mathcal{C}(\pi) \tag{3.8}$$

Proof of the compacity of $\Pi(\rho,\mu)$. We start by recalling here without proof Prokhorov theorem and Ulam lemma

Theorem 3.4 (Prokhorov theorem)**.** *Let* $(\mathcal{Z},d)$ *be a Polish space and let* $\mathcal{F}\subset\mathcal{M}_+(\mathcal{Z})$[1] *with* $\sup_{\mu\in\mathcal{F}}\mu(\mathcal{Z})<\infty$*. Then* $\mathcal{F}$ *is relatively compact with respect to the weak topology if and only if* $\mathcal{F}$ *is equi-tight, i.e. for every* $\varepsilon>0$ *there exists* $K\subset\mathcal{Z}$ *compact such that* $\mu(\mathcal{Z}\backslash K)<\varepsilon$ *for all* $\mu\in\mathcal{F}$*.*

Lemma 3.5 (Ulam lemma)**.** *Let* $(\mathcal{Z},d)$ *be a Polish space and* $\mu\in\mathcal{M}_+(\mathcal{Z})$ *then* $\{\mu\}$ *is equi-tight.*

The closedness of $\Pi(\rho,\mu)$ is an easy consequence of rewriting the marginal condition as

$$\int_{\mathcal{X}}\phi d\rho=\int_{\mathcal{X}\times\mathcal{Y}}\phi d\pi\quad\forall\phi\in C_b(\mathcal{X})\tag{3.9}$$

and similarly for the second marginal condition.
Thanks to Prokhorov theorem it is now sufficient to prove the equitightness of $\Pi(\rho,\mu)$. Fixing $\varepsilon>0$, thanks to Ulam's lemma there exist $K\subset\mathcal{X}$ and $\tilde K\subset\mathcal{Y}$ compact sets such that $\rho(\mathcal{X}\backslash K)<\frac{\varepsilon}{2}$ and $\mu(\mathcal{Y}\backslash\tilde K)<\frac{\varepsilon}{2}$.
Thus

$$\pi(\mathcal{X}\times\mathcal{Y}\backslash K\times\tilde K)\le\pi((\mathcal{X}\backslash K)\times\mathcal{Y})+\pi(\mathcal{X}\times(\mathcal{Y}\backslash\tilde K)\le\varepsilon\tag{3.10}$$

this proves equi-tightness and concludes the proof. □

sec:dual

3.2. **Duality formula.** One of the most important results in the Optimal Transport theory is the following Kantorovitch duality formula.

thm:KantoDual

Theorem 3.6 (Kantorovitch duality)**.** *Let* $\mathcal{X}$ *and* $\mathcal{Y}$ *be Polish spaces, let* $\rho\in\mathcal{P}(\mathcal{X})$ *and* $\mu\in\mathcal{P}(\mathcal{Y})$*, and let* $c:\mathcal{X}\times\mathcal{Y}\to[0,+\infty]$ *be a continuous cost function. Then,*

{eq:KantoDual}

$$\begin{aligned}\mathcal{T}^c(\rho,\mu)&=\sup_{(\phi,\psi)\in I_c}\int_{\mathcal{X}}\phi(x)d\rho(x)+\int_{\mathcal{Y}}\psi(y)d\mu(y)\\&=\sup_{(\phi,\psi)\in J_c}\int_{\mathcal{X}}\phi(x)d\rho(x)+\int_{\mathcal{Y}}\psi(y)d\mu(y)\end{aligned}\tag{3.11}$$

where

$$I_c:=\{(\phi,\psi)\in Lip_b(\mathcal{X})\times Lip_b(\mathcal{Y})|\phi(x)+\psi(y)\le c(x,y)\}\tag{3.12}$$

and

$$J_c:=\{(\phi,\psi)\in L^1(\rho)\times L^1(\mu)|\phi(x)+\psi(y)\le c(x,y)\}\tag{3.13}$$

Remark 3.7. (1) *Easy part of Theorem 3.6:* in (3.11), one inequality is immediate and one requires more work. Indeed for every $(\phi,\psi)\in I_c$ (resp. J_c) and $\pi\in\Pi(\rho,\mu)$ it holds

$$\int_{\mathcal{X}\times\mathcal{Y}}c(x,y)d\pi(x,y)\ge\int_{\mathcal{X}\times\mathcal{Y}}\phi(x)+\psi(y)d\pi(x,y)\ge\int_{\mathcal{X}}\phi(x)d\rho(x)+\int_{\mathcal{Y}}\psi(y)d\mu(y)\tag{3.14}$$

so that

$$\mathcal{T}^c(\rho,\mu)\ge\sup_{(\phi,\psi)\in I_c}\int_{\mathcal{X}}\phi(x)d\rho(x)+\int_{\mathcal{Y}}\psi(y)d\mu(y)\tag{3.15}$$

and analogously with J_c in place and role of I_c. As $I_c\subset J_c$, the reverse inequality for I_c is enough to prove the theorem.

(2) Before giving a rigorous proof of Theorem 3.6, we derive a formal one. The idea, which is standard in optimization problems with contraints, is to rewrite the constrained infimum problem as an inf sup problem, and exchange the two operations by formally applying a minimax principle, i.e. replacing an "inf sup" by a "sup inf".
Let us introduce the indicator function χ_A of a set defined as

$$\chi_A(\pi)\quad=\quad\begin{cases}0&\text{if }\pi\in A\\+\infty&\text{otherwise.}\end{cases}\tag{3.16}$$

[1]define mm+

so that we can write the Optimal Transport problem as

{eq:otpbForDual} (3.17)
$$\mathcal{T}^c(\rho,\mu) = \inf_{\pi\in\mathcal{M}_+(\mathcal{X},\mathcal{Y})} \Big(\mathcal{C}(\pi) + \chi_{\Pi(\rho,\mu)}(\pi)\Big).$$

Moreover, it is easy to check that

(3.18)
$$\chi_{\Pi(\rho,\mu)}(\pi) = \sup_{(\phi,\psi)\in \mathrm{Lip}_b(\mathcal{X})\times\mathrm{Lip}_b(\mathcal{Y})} \Big(\int_{\mathcal{X}} \phi d\rho + \int_{\mathcal{Y}} \psi d\mu - \int_{\mathcal{X}\times\mathcal{Y}} [\phi(x)+\psi(y)] d\pi(x,y)\Big).$$

As a consequence, the right hand side of (3.17) is given by

(3.19)
$$\begin{aligned}\inf_{\pi\in\mathcal{M}_+(\mathcal{X}\times\mathcal{Y})} \sup_{(\phi,\psi)} \Big(&\int_{\mathcal{X}\times\mathcal{Y}} c(x,y) d\pi(x,y) \\ &+ \int_{\mathcal{X}} \phi d\rho + \int_{\mathcal{Y}} \psi d\mu - \int_{\mathcal{X}\times\mathcal{Y}} [\phi(x)+\psi(y)] d\pi(x,y)\Big).\end{aligned}$$

Taking for granted that a minimax principle can be invoked, we rewrite this as

(3.20)
$$\begin{aligned}&\sup_{(\phi,\psi)} \inf_{\pi\in\mathcal{M}_+(\mathcal{X}\times\mathcal{Y})} \Big(\int_{\mathcal{X}\times\mathcal{Y}} c(x,y) d\pi(x,y) \\ &\qquad + \int_{\mathcal{X}} \phi d\rho + \int_{\mathcal{Y}} \psi d\mu - \int_{\mathcal{X}\times\mathcal{Y}} [\phi(x)+\psi(y)] d\pi(x,y)\Big) \\ &= \sup_{(\phi,\psi)} \Big(\int_{\mathcal{X}} \phi d\rho + \int_{\mathcal{Y}} \psi d\mu \\ &\qquad - \sup_{\pi\in\mathcal{M}_+(\mathcal{X}\times\mathcal{Y})} \int_{\mathcal{X}\times\mathcal{Y}} [\phi(x)+\psi(y)-c(x,y)] d\pi(x,y)\Big)\end{aligned}$$

We now first compute the supremum over π. If $(\phi,\psi)\in I_c$ the supremum is clearly 0 and attained for $\pi=0$, otherwise, there exists $(x,y)\in\mathcal{X}\times\mathcal{Y}$ such that $\phi(x)+\psi(y)-c(x,y)>0$, then taking $\pi = \lambda\delta_{(x,y)}$ and letting $\lambda\to+\infty$, we see that the supremum is infinite. Therefore,

(3.21)
$$\sup_{\pi\in\mathcal{M}_+(\mathcal{X}\times\mathcal{Y})} \int_{\mathcal{X}\times\mathcal{Y}} [\phi(x)+\psi(y)-c(x,y)] d\pi(x,y) = \chi_{I_c}((\phi,\psi)),$$

so that previous lines, yields

(3.22)
$$\mathcal{T}^c(\rho,\mu) \geq \sup_{(\phi,\psi)\in I_c} \int_{\mathcal{X}} \phi(x) d\rho(x) + \int_{\mathcal{Y}} \psi(y) d\mu(y)$$

Proof of Theorem 3.6. We now prove rigorously Theorem 3.6, in the case where $\mathcal{X}$ and $\mathcal{Y}$ are compact, c is continuous. This is the first step of most proofs for Theorem 3.6. The additional assumptions on $\mathcal{X}$, $\mathcal{Y}$ and c are then relaxed through approximations procedures. The minimax principle is made rigorous through a convex analysis theorem known as Fenchel-Rockafellar duality theorem. We first need to introduce a few notations. Let E be a normed vector space, and Θ a convex function on E with values in $\mathbb{R}\cup\{+\infty\}$, the Legendre-Fenchel transform of Θ is the function Θ^*, defined on the topological dual E^* of E by the formula

(3.23)
$$\Theta^*(z^*) = \sup_{z\in E} [\langle z^* | z\rangle - \Theta(z)].$$

thm:fenrock **Theorem 3.8** (Fenchel-Rockafellar duality)**.** *Let E be a normed vector space, E^* its topological dual space, and Θ, Ξ two convex functions on E^* with values in $\mathbb{R}\cup\{+\infty\}$. Let Θ^*, Ξ^* be the Legendre-Fenchel transforms of Θ, Ξ respectively. Assume that there exists $z_0\in E$ such that*

(3.24)
$$\begin{gathered}\Theta(z_0)<+\infty, \quad \Xi(z_0)<+\infty, \\ \Theta \text{ is continuous at } z_0\end{gathered}$$

Then,

{eq:fenrock} (3.25)
$$\inf_E [\Theta+\Xi] = \max_{z^*\in E^*} [-\Theta^*(-z^*) - \Xi^*(z^*)]$$

Let $E = C(\mathcal{X}\times\mathcal{Y})$ be the set of all continuous functions on $\mathcal{X}\times\mathcal{Y}$, equipped with its usual supremum norm $\|.\|_\infty$. By Riesz theorem we have $E^* = \mathcal{M}(\mathcal{X}\times\mathcal{Y})$ the space of Radon measures normed by total variation.
Then we introduce

eq:fenrock

$$\Theta : u \in C(\mathcal{X}\times\mathcal{Y}) \quad \to \quad \begin{cases} 0 & \text{if } u(x,y) \ge -c(x,y) \\ +\infty & \text{otherwise,} \end{cases} \tag{3.26}$$

$$\Xi : u \in C(\mathcal{X}\times\mathcal{Y}) \quad \to \quad \begin{cases} \int_{\mathcal{X}} \phi d\rho + \int_{\mathcal{Y}} \psi d\mu & \text{if } u(x,y) = \phi(x)+\psi(y) \\ +\infty & \text{otherwise,} \end{cases} \tag{3.27}$$

and we apply Theorem 3.8. The assumptions are obviously satisfied with $z_0 \equiv 1$ so (3.25) holds true.

The left hand side reads,

$$\inf\Big\{\int_{\mathcal{X}} \phi d\rho + \int_{\mathcal{Y}} \psi d\mu; \phi(x)+\psi(y) \ge -c(x,y)\Big\} = -\sup\{\int_{\mathcal{X}} \phi d\rho + \int_{\mathcal{Y}} \psi d\mu; (\phi,\psi)\in I_c\} \tag{3.28}$$

Next, we compute the Legendre-Fenchel transforms of Θ, Ξ. First, for any $\pi \in \mathcal{M}(\mathcal{X}\times\mathcal{Y})$,

$$\begin{aligned} \Theta^*(-\pi) &= \sup_{u\in C(\mathcal{X}\times\mathcal{Y})} \Big\{ -\int u(x,y) d\pi(x,y); u(x,y) \ge -c(x,y)\Big\} \\ &= \sup_{u\in C(\mathcal{X}\times\mathcal{Y})} \{\int u(x,y) d\pi(x,y); u(x,y) \le c(x,y)\Big\} \end{aligned} \tag{3.29}$$

- If π is not a nonnegative measure, then there exists a nonpositive function $v \in C(\mathcal{X}\times\mathcal{Y})$, such that $\int v d\pi > 0$. Then, the choice $u = \lambda v$, with $\lambda \to +\infty$, shows that the supremum is $+\infty$.
- On the other hand, if π is nonnegative, then the supremum is clearly $\int c d\pi$.

Thus,

$$\Theta^*(-\pi) \quad = \quad \begin{cases} \int c(x,y) d\pi(x,y) & \text{if } \pi \in \mathcal{M}_+(\mathcal{X}\times\mathcal{Y}) \\ +\infty & \text{otherwise,} \end{cases} \tag{3.30}$$

and similarly,

$$\Xi^*(\pi) \quad = \quad \begin{cases} 0 & \text{if } \pi \in \Pi(\rho,\mu) \\ +\infty & \text{otherwise.} \end{cases} \tag{3.31}$$

Putting everything together, we recover

$$\inf_{\pi\in\Pi(\rho,\mu)} \mathcal{C}(\pi) = \sup_{(\phi,\psi)\in I_c} \int \phi d\rho + \int \psi d\mu \tag{3.32}$$

which concludes the proof. □

sec:uni

3.3. c-cyclically monotonicity, c-concavity and optimality conditions.

Let us start by defining c-conjugate functions, c-cyclically monotonicity and c-concavity, which provide an extension of the notion of convex functions useful for studying unicity and optimality conditions in both primal and dual optimal transport problems.

Definition 3.9 (c-conjugate function)**.** Given $\phi : \mathcal{X} \to [-\infty, +\infty)$ we define the c-conjugate function ϕ^c by

$$\phi^c(y) := \inf_{x\in\mathcal{X}} [c(x,y) - \phi(x)] \tag{3.33}$$

Anagolously if $\psi : \mathcal{Y} \to [-\infty, +\infty)$, we define

$$\psi^c(x) := \inf_{y\in\mathcal{Y}} [c(x,y) - \psi(y)] \tag{3.34}$$

Definition 3.10 (c-cyclically monotonicity)**.** We say that $\Gamma \subset \mathcal{X} \times \mathcal{Y}$ is c-cyclically monotone if

$$\sum_{i=1}^{N} c(x_i, y_{\sigma(i)}) \geq \sum_{i=1}^{N} c(x_i, y_i) \tag{3.35}$$

for every $N \geq 1$, σ permutation of $\{1, ..., N\}$ and $(x_i, y_i) \in \Gamma$ for $i = 1, ..., N$.

Definition 3.11 (c-concavity)**.** A function $\phi : \mathcal{X} \to [-\infty, +\infty)$ is said to be c-concave if it is the infimum of a family of c-affine function $\phi(x) = \inf_{(y,\alpha)\in Y\times A} c(x, y)+\alpha$, $Y \subset \mathcal{Y}$, $A \subset \mathbb{R}$. Analogously, $\psi : \mathcal{Y} \to [-\infty, +\infty)$ is said to be c-concave if $\psi(y) = \inf_{(x,\alpha)\in X\times A} c(x, y) + \alpha$, $X \subset \mathcal{X}$, $A \subset \mathbb{R}$.

em:c-concavity *Remark* 3.12. (1) When $c(x, y) = -\langle x, y\rangle$, up to a change of sign, one recovers the usual notions of convexity, Legendre-Fenchel transform and monotone sets.

(2) An important fact for the following is that the modulus of continuity of a c-concave function is bounded by the modulus of continuity of c. In particular if c is L-Lipschitz and ϕ c-concave, then ϕ is L-Lipschitz as well.

(3) Finally, elementary manipulations show that if ϕ is c-concave, $\phi^{cc} = \phi$.

The following theorem yields a characterization of optimal plans

thm:charcopti **Theorem 3.13.** *Under the assumptions of Theorem 3.6 assume moreover that there exists $a \in L^1(\rho)$, $b \in L^1(\mu)$ such that $c(x, y) \leq a(x) + b(y)$ then*

(1) *$\pi \in \Pi(\rho, \mu)$ is optimal in (OT_c) if and only if it is concentrated on a c-cyclically monotone σ-compact set.*

it:potL1 (2) *there exists a c-concave function $\phi : \mathcal{X} \to [-\infty, +\infty)$ such that $\phi \in L^1(\rho)$, $\phi^c \in L^1(\mu)$ and*

{eq:dualat}
$$\mathcal{T}^c(\rho, \mu) = \int \phi d\rho + \int \phi^c d\mu \tag{3.36}$$

Sketch of proof of Theorem 3.13. Assume by contradiction that the conclusion is false. Then there exist $N \geq 1$, a permutation σ and points $(x_1, y_1), ..., (x_N, y_N) \in \mathrm{supp}\pi$ such that

{eq:notcmon}
$$\sum_{i=1}^{N} c(x_i, y_{\sigma(i)}) < \sum_{i=1}^{N} c(x_i, y_i). \tag{3.37}$$

By the continuity of the cost function c there exist neighbourhoods $U_i \times V_i \ni (xi, yi)$ where inequality (3.37) continues to hold, replacing (x_i, y_i) with any choice of $(x_i', y_i') \in U_i \times V_i$. Now we want to construct a measure $\theta \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ such that:

(1) θ has null marginals on $\mathcal{X}$ and $\mathcal{Y}$;

(2) $\theta^- \leq \pi$ (where that θ^- denotes the negative part of the measure θ);

(3) $\int c d\theta < 0$

If we are able to construct such a measure θ, then we get a contradiction since $\pi+\theta \in \Pi(\rho, \mu)$ and

$$\int c d(\pi + \theta) < \mathcal{T}^c(\rho, \mu). \tag{3.38}$$

We will not get into the details of this construction, but formally the measure θ should simply deduce the π-mass of $U_i \times V_i$ and put it on $U_i \times V_{\sigma(i)}$, (3.37) would then ensure $\int c d\theta < 0$ holds.
We turn to the proof of (2). Fix $(x_0, y_0) \in \mathrm{supp}\pi$ we define;
(3.39)
$$\phi(x) := \inf\{c(x, y_N), -c(x_N, y_N)+c(x_N, y_{N-1})-c(x_{N-1}, y_{N-1})+...+c(x_1, y_0)-c(x_0, y_0)\}$$

where the infimum is made over $N \leq 1$ and $(x_1, y_1), ..., (x_N, y_N) \in \mathrm{supp}\pi$. One can check that $\phi(x_0) = 0$, moreover the construction of ϕ is made so that $\phi + \phi^c = c$ π a.e.. Indeed if we first minimize keeping (x_N, y_N) fixed, then we minimize with respect to (x_N, y_N) we see that

$$\phi(x) = \inf_{(x_N,y_N)} c(x, y_N) - c(x_N, y_N) + \phi(x_N) \tag{3.40}$$

which implies, for any $(x', y') \in \text{supp}\pi$, for any $x \in \mathcal{X}$

$$c(x', y') \le c(x, y') - \phi(x) + \phi(x') \tag{3.41}$$

and thus for any $(x', y') \in \text{supp}\pi$

$$c(x', y') \le \phi^c(y') + \phi(x') \tag{3.42}$$

the reverse inequality being always true, by definition of the c-transform, we conclude $\phi + \phi^c = c\ \pi$ a.e.. Finally, assuming $\phi \in L^1(\rho)$ and $\mu \in L^1(\mu)$, we get

{eq:optidual}

$$0 \le \mathcal{C}(\pi) = \int_{\mathcal{X}\times\mathcal{Y}} c d\pi = \int_{\mathcal{X}\times\mathcal{Y}} (\phi + \phi^c) d\pi = \int_{\mathcal{X}} \phi d\rho + \int_{\mathcal{Y}} \phi^c d\mu. \tag{3.43}$$

To conclude, we need to show the integrability. First notice that $\phi + \phi^c = c\ \pi$ a.e. implies $\phi > -\infty\ \rho$ a.e. and $\phi^c > -\infty\ \nu$ a.e.. Choosing $x \in \mathcal{X}$ such that both $a(x) < \infty$ and $\phi(x) > -\infty$, it follows that$\phi^c(y) \le c(x, y) - \phi(x) \le b(y) + a(x) - \phi(x)$ and

$$\int_{\mathcal{Y}} (\phi^c)^+ d\mu \le \int b d\mu + a(x) - \phi(x) < \infty. \tag{3.44}$$

Similarly, using $\phi^{cc} = \phi$, we get $\int_{\mathcal{X}} (\phi)^+ d\rho < \infty$. This is enough for 3.43 to hold and we get that $\int_{\mathcal{Y}} (\phi^c)^- d\mu < \infty$, $\int_{\mathcal{X}} (\phi)^- d\rho < \infty$ and the desired conclusion $\phi \in L^1(\rho)$, $\phi^c \in L^1(\mu)$ □

Remark 3.14. Even when $(\phi, \phi^c) \notin \text{Lip}_b(\mathcal{X}) \times \text{Lip}_b(\mathcal{Y})$, we call a couple (ϕ, ϕ^c) that satisfies (3.36) a pair of Kantorovitch potentials. Moreover, if ϕ is the Kantorovitch potential with 0 ρ-mean, $\phi - C$ as well for any $C \in \mathbb{R}$, as a consequence we will usually choose a Kantorovitch potential with $\int_{\mathcal{X}} \phi d\rho = 0$.

thm:BrenKS

Theorem 3.15 (Brenier, Knott-Smith)**.** *Assume $\mathcal{X}$, $\mathcal{Y} \subset \mathbb{R}^d$ compact, $c(x, y) = \frac{1}{p}|x - y|^p$ for some $p > 1$ and let $\rho \in \mathcal{P}(\mathcal{X}), \mu \in \mathcal{P}_(\mathcal{Y})$ with ρ absolutely continuous with respect to Lebesgue's measure. Then the problem (OT) has a unique solution π. In addition, π is induced by a transport map T :*

$$\pi = (Id \times T)_{\#}\ \rho$$
$$i.e.\ \ d\pi(x, y) = \rho(x)\delta_{T(x)=y}$$

and $T(x) = x - |\nabla\phi(x)|^{\frac{2-p}{p-1}} \nabla\phi(x)$, where $\phi : \mathbb{R}^d \to [-\infty, \infty)$ is a lower semicontinuous c-concave function differentiable ρ-almost everywhere. Finally, as its gradient is uniquely determined, the Kantorovitch potential is unique up to a constant.

Sketch of proof of Theorem 3.15. The existence of π optimal and of a pair (ϕ, ϕ^c) of Kantorovitch potentials follow from the previous results. Both ϕ and ϕ^c are Lipschitz, thus differentiable Lebesgue a.e. and therefore differentiable ρ a.e.. Now take $x \in \mathcal{X}$ such that $\nabla\phi(x)$ exists. We want to prove there exists a unique $y \in \mathcal{Y}$ such that $(x, y) \in$ suppπ. We already know, since $\phi + \phi^c = c$ on suppπ, the existence part and that for such a y, $c(x', y) - \phi(x)$ is minimal at $x = x'$. The first order condition yields $y = x - |\nabla\phi(x)|^{\frac{2-p}{p-1}} \nabla\phi(x) := T(x)$. As a consequence π is supported on the graph of T. Finally, if π, π' optimal plans, T, T' associated maps as we just constructed, $\pi'' := \frac{1}{2}(\pi + \pi') = \frac{1}{2}\rho(x)(\delta_{T(x)}(y) + \delta_{T'(x)}(y))$ is also optimal. It follows that π'' should be concentrated on the graph of a map. However $\pi''(x, y) = \frac{1}{2}\rho(x)(\delta_{T(x)}(y) + \delta_{T'(x)}(y))$ holds, which can happen if and only if $T = T'\ \rho$ a.e. in $\mathcal{X}$. □

Remark 3.16. The conclusion of Theorem 3.15 also holds without the compactness assumption, however the proof is more technical. As the compactness is needed for Theorem 2.1 and to keep things simple we chose this setting for Theorem 3.15

3.4. **Qualitative stability.** After having studied the question of existence of optimizers for the primal and dual problems in Section 3.1, Section 3.2 and of uniqueness in Section 3.3, we turn to stability.

thm:qualistab **Theorem 3.17** (Qualitative stability). *Let $\mathcal{X}$ be a compact subset of $\mathbb{R}^d$ and $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ continuous. Let $(\rho_n)_{n\in\mathbb{N}}$, $(\mu_n)_{n\in\mathbb{N}}$ be sequences in $\mathcal{P}(\mathcal{X})$ and assume that $\rho_n \rightharpoonup \rho \in \mathcal{P}(\mathcal{X})$ and $\mu_n \rightharpoonup \mu \in \mathcal{P}(\mathcal{X})$. Then for each $n \geq 0$, denoting $\pi_n \in \Pi(\rho_n, \mu_n)$ an optimal transport plan and $\phi_n \in L^1(\rho_n)$ a Kantorovitch potential for the transport problem between ρ_n and μ_n w.r.t. c, one has, up to a subsequence:*

it:cvplan (1) *$\pi_n \rightharpoonup \pi \in \Pi(\rho, \mu)$, where π is an optimal transport plan between ρ and μ w.r.t. c.*

it:cvval (2) *$\min_{\pi\in\Pi(\rho_n,\mu_n)} \int c d\pi \to \min_{\pi\in\Pi(\rho,\mu)} \int c d\pi$.*

it:cvpot (3) *$\|\phi_n - \phi\|_{L^\infty(\mathcal{X})} \to 0$ and $\|\phi_n^c - \phi^c\|_{L^\infty(\mathcal{X})} \to 0$, where $\phi \in L^1(\rho)$ is a Kantorovitch potential between ρ and μ w.r.t. c.*

Sketch of proof of Theorem 3.17. (1) is proved using Porkhorov's theorem to extract a converging subsequence $\pi_{n_k} \rightharpoonup \pi \in \Pi(\rho, \mu)$ and checking the optimality condition holds in the limit. To this aim use that for any $(x, y) \in \operatorname{supp}\pi$, there exists a sequence $(x_k, y_k) \in \operatorname{supp}\pi_{n_k}$ such that $(x_k, y_k) \to (x, y)$. (2) is a consequence of (1). For (3), using that c-concave functions have the same modulus of continuity as c, Arzela-Ascoli theorem allows to extract convergent subsequences of ϕ_n, ϕ_n^c and the optimality of the limit is shown using (2) . □

4. Regularization, approximations and useful inequalities

sec:regapprox

In this section, we present and state some additional results usually without any proof. It is mainly a toolbox for Section 5. Theorem 2.1 is an extension to any $p > 1$ of a similar result for $p = 2$ in [Del22]. We use the same regularization strategy than in [Del22] but our proof then rely on a new generalization of the Prékopa-Leindler inequality that one can find in the proof of Proposition 5.7.

delalande2022quantitative delalande2022quantitative

sec:entrop

4.1. **Entropic regularization.** For any $\varepsilon > 0$, we introduce the ε-entropic regularized optimal transport problem, with cost c, as the following minimization problem:

{pb:eot}
$$(\mathrm{EOT}_{c,\varepsilon}) \qquad \mathcal{T}^{c,\varepsilon}(\rho, \mu) := \inf_{\pi\in\Pi(\rho,\mu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y)d\pi(x, y) + \varepsilon \mathsf{Ent}(\pi||\rho \otimes \mu)$$

where Ent denotes the relative entropy functional

pb:eot
$$\mathsf{Ent}(.||\rho \otimes \mu) : \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \longrightarrow \mathbb{R} \cup \{+\infty\} \tag{4.1}$$
$$\mathsf{Ent}(\pi||\rho \otimes \mu) = \begin{cases} \displaystyle\int_{\mathcal{X}\times\mathcal{Y}} \log\left(\frac{d\pi}{d(\rho\otimes\mu)}\right)d\pi & \text{if } \pi \ll \rho\otimes\mu \\ +\infty & \text{otherwise.} \end{cases} \tag{4.2}$$

Remark 4.1. (1) Note also that for $\sigma \in \mathcal{P}(\mathcal{Y})$ such that $\mu \ll \sigma$ and $\sigma \ll \mu$, it admits the following rewriting

$$\mathcal{T}^{c,\varepsilon}(\rho, \mu) := \inf_{\pi\in\Pi(\rho,\mu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y)d\pi(x, y) + \varepsilon \mathsf{Ent}(\pi||\rho \otimes \sigma) - \varepsilon \mathsf{Ent}(\mu||\sigma), \tag{4.3}$$

which will be quite useful in our derivations.

Since the "source" measure $\rho \in \mathcal{P}(\mathcal{X})$ plays a distinguished role in our results, we write the dual formulation of the optimal transport problem in the following asymmetric way (also called semi-dual in [Del22]):

delalande2022quantitative

$$\mathcal{T}^c(\rho, \mu) = \sup_{\psi\in\mathcal{C}(\mathcal{Y})} \left\{ \int_{\mathcal{X}} \psi^c d\rho + \int_{\mathcal{Y}} \psi d\mu \right\}, \tag{4.4}$$

For the ε-entropic regularized version, duality reads [MG20; NW22]

$$\mathcal{T}^{c,\varepsilon}(\rho,\mu) = \sup_{\psi\in\mathcal{C}(\mathcal{Y})} \left\{ \int_{\mathcal{X}} \psi^{c,\varepsilon} d\rho + \int_{\mathcal{Y}} \psi d\mu \right\} - \varepsilon \mathsf{Ent}(\mu||\sigma), \tag{4.5}$$

where the (c,ε)-transform of ψ is given by

$$\psi^{c,\varepsilon}(x) := -\varepsilon \log \left(\int_{\mathcal{Y}} \exp\left(-\frac{c(x,y)-\psi(y)}{\varepsilon} \right) d\sigma \right). \tag{4.6}$$

We notice that this does not depends anymore upon μ. Being able to use the same definition of (c,ε)-transform for different target measure μ is a key element of the proof. Symmetrically, c-transform and (c,ε)-transform can be defined for continuous functions on $\mathcal{X}$. When a function in $\mathcal{C}(\mathcal{Y})$ can be written as a c-transform (resp. (c,ε)-transform) we say it is c-concave (resp. (c,ε)-concave function). In this formulation, a c-concave or (c,ε)-concave optimizer $\psi \in \mathcal{C}(\mathcal{Y})$ is called a Kantorovich potential (for $\mathcal{T}^c(\rho,\mu)$ or $\mathcal{T}^{c,\varepsilon}(\rho,\mu)$), and $(\phi,\psi) = (\psi^c,\psi) \in \mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{Y})$ is called a pair of Kantorovitch potentials (and similarly for $\varepsilon > 0$). We notice that such transforms define monotone decreasing maps, i.e., for $\psi, \tilde{\psi} \in \mathcal{C}(\mathcal{Y})$

$$\psi(y) \le \tilde{\psi}(y) \quad \text{for every } y \in \mathcal{Y} \quad \Rightarrow \quad \psi^{c,\varepsilon}(x) \ge \tilde{\psi}^{c,\varepsilon}(x) \quad \text{for every } x \in \mathcal{X}, \tag{4.7}$$

and are concave,

$$((1-t)\tilde{\psi} + t\psi)^{c,\varepsilon}(x) \ge (1-t)\tilde{\psi}^{c,\varepsilon}(x) + t\psi^{c,\varepsilon}(x) \quad \text{for every } t \in [0,1],\ x \in \mathcal{X}, \tag{4.8}$$

as a consequence of Hölder's inequality. Moreover, $(\psi+\lambda)^{c,\varepsilon} = \psi^{c,\varepsilon} - \lambda$ for every constant $\lambda \in \mathbb{R}$.

We introduce, for ϕ_0, $\phi_1 \in \mathcal{C}(\mathcal{X})$, the quantity

$$M^c_{\phi_0,\phi_1} = \sup_{t\in[0,1]} \{\mathrm{osc}(\phi_t)\}, \tag{4.9}$$

where $\phi_t := ((1-t)\phi_0^c + t\phi_1^c)^c$.

Example 4.2. If the cost c is Lipschitz continuous, then any function $\phi = \psi^c$ that is a c-transform is Lipschitz continuous. In the case of the p-cost, we obtain that $\mathrm{Lip}(\phi) \le (R_{\mathcal{X}} + R_{\mathcal{Y}})^{p-1}$, hence $\mathrm{osc}(\phi) \le 2R_{\mathcal{X}}(R_{\mathcal{X}} + R_{\mathcal{Y}})^{p-1}$. In particular, we can always bound from above

$$M^c_{\phi_0,\phi_1} \le 2R_{\mathcal{X}}(R_{\mathcal{X}} + R_{\mathcal{Y}})^{p-1}, \tag{4.10}$$

where $R_{\mathcal{X}}$ (resp. $R_{\mathcal{Y}}$) is the smallest R such that $\mathcal{X} \subset B(0,R)$ (resp. $\mathcal{Y} \subset B(0,R)$).

As in [Del22] we introduce the c-Kantorovich functional

$$\psi \in \mathcal{C}(\mathcal{Y}) \quad \mapsto \quad \mathcal{K}^c(\psi) = -\int_{\mathcal{X}} \psi^c(x) d\rho(x) \tag{4.11}$$

and similarly the (c,ε)-Kantorovich functional

$$\psi \in \mathcal{C}(\mathcal{Y}) \quad \mapsto \quad \mathcal{K}^{c,\varepsilon}(\psi) = -\int_{\mathcal{X}} \psi^{c,\varepsilon}(x) d\rho(x). \tag{4.12}$$

We remark that the notation does not highlight the dependence upon ρ, which will be clear from the context. Both functionals $\mathcal{K}^c$, $\mathcal{K}^{c,\varepsilon}$ are convex, and the duality formulas read

$$\begin{aligned} \mathcal{T}^c(\rho,\mu) &= \sup_{\psi\in\mathcal{C}(\mathcal{Y})} \{\langle \mu|\psi\rangle - \mathcal{K}^c(\psi)\} \\ \mathcal{T}^{c,\varepsilon}(\rho,\mu) &= \sup_{\psi\in\mathcal{C}(\mathcal{Y})} \{\langle \mu|\psi\rangle - \mathcal{K}^{c,\varepsilon}(\psi)\} - \varepsilon \mathsf{Ent}(\mu||\sigma). \end{aligned} \tag{4.13}$$

Remark 4.3. With the above definitions, we have convergence of the regularized problems to the usual transport problem as $\varepsilon \downarrow 0$, as discussed in [NW22]. More precisely, since both $\mathcal{X}$ and $\mathcal{Y}$ are bounded and the cost function is (uniformly) continuous, any family of functions that are (c,ε)-transforms has a uniform modulus of continuity, also uniformly with respect to $\varepsilon > 0$. If we require that optimal potential $\phi = \psi^{c,\varepsilon}$ have e.g. zero mean with respect to ρ, we get by Arzelà-Ascoli that the families are relatively compact, hence up to a subsequence they converge uniformly. Reversing the roles of ϕ and ψ leads to compactness also for the optimizers ψ's. If moreover uniqueness of the (zero mean) optimal potentials holds, as is the case of p-costs for $p > 1$ in our setting, one obtains uniform convergence of the optimizers (and also of the costs), otherwise one can still extract a uniformly converging subsequence.

4.2. **Semi-discrete problem.** It will be convenient to further approximate by assuming that $\mathcal{Y}$ is a finite set. The following result extends [Del22, Lemma 2.7] and allows us to pass from such semi-discrete case to the general case of $\mathcal{Y}$ compact.

lem:approx **Lemma 4.4.** *Let $\mathcal{X} \subset \mathbb{R}^d$ be compact and convex and $\mathcal{Y} \subset \mathbb{R}^d$ be compact, $\rho \in \mathcal{P}(\mathcal{X})$ be absolutely continuous, $\mu_0, \mu_1 \in \mathcal{P}(\mathcal{Y})$ and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be continuous. For $k \in \{0, 1\}$, let $\phi_k \in \mathcal{C}(\mathcal{X})$ assume that there exists a unique Kantorovich potential with zero ρ-mean for $\mathcal{T}^c(\rho, \mu_k)$ and ψ_k its c-transform. Then, there exists sequences of finite sets $(\mathcal{Y}^n)_{n=1}^\infty$, discrete measures $(\mu_k^n)_{n=1}^\infty$, Kantorovich potentials $(\phi_k^n, \psi_k^n)_{n=1}^\infty$ such that, for $k \in \{0, 1\}$,*

(1) $\operatorname{supp}(\mu_k^n) = \mathcal{Y}^n$ *for every* $n \geq 1$,

(2) $\|\phi_k^n - \phi_k\|_\infty \to 0$ *and* $\|\psi_k^n - \psi_k\|_\infty \to 0$ *as* $n \to \infty$

(3) *it holds* $\langle \mu_k^n | \psi_1^n - \psi_0^n \rangle \to \langle \mu_k | \psi_1 - \psi_0 \rangle$ *as* $n \to \infty$.

Without the uniqueness assumption, one may still extract a converging subsequence to a pair of Kantorovich potentials.

Proof. For any $n \geq 1$, consider a finite partition $\mathcal{Y} = \bigsqcup_{1 \leq i \leq n} \mathcal{Y}_i^n$ with $\varepsilon_n := \max_{1 \leq i \leq n} \operatorname{diam}(\mathcal{Y}_{i,N})$ infinitesimal as $n \to \infty$. Define $\mathcal{Y}^n = \{y_i^n\}_{i=1}^n$, with $y_i^n \in \mathcal{Y}_i^n$ and set

$$\mu_k^n = \sum_{i=1}^n \left[\left(1 - \frac{1}{n}\right) \mu_k(\mathcal{Y}_i^n) + \frac{1}{n^2} \right] \delta_{y_i^n}, \tag{4.14}$$

so that (1) is verified.

Letting ϕ_k^n be the Kantorovich potential for $\mathcal{T}^c(\rho, \mu_k^n)$ with zero ρ-mean, then (2) holds by Arzelà-Ascoli and the fact that the limiting potential is unique.

To show (3), notice that

$$\mathcal{W}_1(\mu_k, \mu_k^n) \leq \varepsilon_n + \frac{\operatorname{diam}(\mathcal{Y})}{n}, \tag{4.15}$$

so that $\mu_k^n \to \mu_k$ weakly and therefore the limit holds, since the functions ψ_k^n converge uniformly. ☐

In the case where $\mathcal{Y}$ is finite, and fixing σ as the uniform measure on $\mathcal{Y}$, we recall some formulas for the derivatives of $\mathcal{K}^{c,\varepsilon}$ which is easily seen to be a smooth map from $\mathcal{C}(\mathcal{Y})$ (identified with $\mathbb{R}^n$) to $\mathbb{R}$. These are proved in [Del22, lemma 4.2, proposition 3.6] for the case of the "linear" cost $'$ defined above (3.4), but the proof can straightforwardly be generalized to any cost.

$$\partial_i(\psi^{c,\varepsilon})(x) = -\frac{\exp\left(\frac{\psi_i - c(x, y_i)}{\varepsilon}\right)\sigma_i}{\sum_{j=1}^n \exp\left(\frac{\psi_j - c(x, y_j)}{\varepsilon}\right)\sigma_j}, \quad \text{for } i = 1 \dots, n, \tag{4.16}$$

which we notice defines for every $x \in \mathcal{X}$ a probability vector, i.e., $-\partial_i(\psi^{c,\varepsilon})(x) \in [0, 1]$ and $-\sum_i \partial_i(\psi^{c,\varepsilon})(x) = 1$, which we naturally identify with a probability on $\mathcal{Y}$, and plays the role of a generalized optimal transport map in this context. We write therefore

$$\pi_{\varepsilon,\psi}^c(y_i|x) := -\partial_i(\psi^{c,\varepsilon})(x). \tag{4.17}$$

which in turn play a role in the expression for the gradient and Hessian of $\mathcal{K}^{c,\varepsilon}$:

{eq:hessian-K}

$$\begin{aligned} \langle \nabla \mathcal{K}^{c,\varepsilon}(\psi), v \rangle &= \int_{\mathcal{X}} \mathsf{m}_{\pi_{\varepsilon,\psi}^c(\cdot|x)}(v) d\rho(x), \\ \langle v, \nabla^2 \mathcal{K}^{c,\varepsilon}(\psi) v \rangle &= \frac{1}{\varepsilon} \int_{\mathcal{X}} \mathsf{var}_{\pi_{\varepsilon,\psi}^c(\cdot|x)}(v) d\rho(x). \end{aligned} \tag{4.18}$$

We also introduce the following "partition function"

{eq:Z}
$$\psi \in \mathcal{C}(\mathcal{Y}) \mapsto \mathcal{Z}_\beta(\psi) := \int_{\mathcal{X}} e^{\beta\psi^{c,\varepsilon}} d\rho. \tag{4.19}$$

As with the function $\mathcal{K}$, the notation does not highlight the dependence upon ρ, which will be clear from the context. The function $\mathcal{Z}$ is smooth and in particular the following formula

{eq:hessI}
$$\langle v, \nabla^2 \log \mathcal{Z}_\beta(\psi) v\rangle = -\frac{\beta}{\varepsilon}\int_{\mathcal{X}} \mathrm{Var}_{\pi^c_{\varepsilon,\psi}(\cdot|x)}(v) d\rho^\varepsilon_\psi(x) + \beta^2 \mathrm{Var}_{\rho^\varepsilon_\psi}\left(\mathsf{m}_{\pi^c_{\varepsilon,\psi}(\cdot|x)}(v)\right), \tag{4.20}$$

holds with

{eq:defrhoI}
$$\rho^\varepsilon_{\beta,\psi} := \frac{1}{\mathcal{Z}_\beta(\psi)} e^{\beta\psi^{c,\varepsilon}}\rho. \tag{4.21}$$

Remark 4.5. In most proofs, to keep notation simple we will fix $\beta = 1$ and write $\mathcal{Z} := \mathcal{Z}_1$.

Finally, recalling that we denote with $'$ the cost $(x, y) \mapsto -\langle x, y\rangle$, we obtain

$$\psi'(x) = -\sup_{y\in\mathcal{Y}} \{\langle x, y\rangle + \psi(y)\}, \quad \psi'^{,\varepsilon}(x) = -\varepsilon \log\left(\int_{\mathcal{Y}} \exp\left(\frac{\langle x, y\rangle + \psi(y)}{\varepsilon}\right) d\sigma(y)\right). \tag{4.22}$$

4.3. **Displacement interpolation and L^2 bound on $\mathcal{W}_2$.** We will make crucial use of the following result on the so-called displacement interpolation between two absolutely continuous probabilities, which is well-known for the quadratic case (see e.g. McCann's proof of Prékopa-Leindler [McC94, appendix D]), while for general power cost a proof can be found as a byproduct of the proof of [AGS08, proposition 9.3.9] applied to the (opposite of) the differential entropy functional $\mathcal{F}(\rho) = \int \rho \log(\rho)$.

lem:boundgeo **Lemma 4.6.** *Let $p > 1$, ρ_0, $\rho_1 \in \mathcal{P}(\mathcal{X})$ be absolutely continuous and let T be the optimal transport map for $\mathcal{W}_p(\rho_0, \rho_1)$. For $t \in [0, 1]$, let $T_t(x) = (1-t)x + tT(x)$ and set $\rho_t = (T_t)_\sharp\rho_0$ the displacement interpolation between ρ_0 and ρ_1. Then ρ_t is absolutely continuous and*

$$\rho_t(T_t(x)) \le \rho_0(x)^{1-t}\rho_1(T(x))^t \quad \text{for } \rho_0\text{-a.e. } x \in \mathcal{X}. \tag{4.23}$$

We also recall the following upper bound for the quadratic Wasserstein distance between two densities, in terms of the L^2 norm of their difference, with respect to a log-concave reference measure ρ.

lem:peyre **Lemma 4.7.** *Let $\rho \in \mathcal{P}(\mathbb{R}^d)$ be* log-concave*. Then, there exists $C = C(\rho) < \infty$ such that the following holds. For every $\mu_0 = f_0\rho$, $\mu_1 = f_1\rho \in \mathcal{P}(\mathbb{R}^d)$, both absolutely continuous with respect to ρ,*

$$\mathcal{W}_2(\mu_0, \mu_1)^2 \le \frac{C}{\text{ess-inf } f_1} \|f_1 - f_0\|^2_{L^2(\rho)}. \tag{4.24}$$

Inequalities of this kind have been used by many authors, see e.g. [Pey18; AST19; GHO21; Led17] and can be proved combining the Benamou-Brenier formula for $\mathcal{W}_2$ and the Poincaré-Wirtinger inequality with respect to the measure ρ.

sec:pot

5. Quantitative stability of potentials

Aim of this section is to establish our main result, Theorem 2.1. As we actually are able to obtain more precise bounds, we split the argument into three subsections. We deal first with the quadratic case, in particular recovering and slightly extending the results in [Del22], but mainly in order to point out where our arguments differ and allow for extensions respectively to the case $1 < p < 2$ (Section 5.2). The case $p > 2$ does not differ much from $p = 2$ and we refer to [MT24] for a detailed argument.

5.1. **The quadratic case.** In this section, we prove Theorem 2.1 in the case $p = 2$. In view of the identity (3.4), it is sufficient to argue in the case of the linear cost $c(x, y) = -\langle x, y\rangle$ (denoted with $'$) above. Therefore, we establish the following result.

Theorem 5.1. *Let ρ be a* log*-concave probability measure with bounded (convex) support $\mathcal{X} \subseteq \mathbb{R}^d$ and let $\mathcal{Y} \subseteq \mathbb{R}^d$ be compact. Given μ_0, $\mu_1 \in \mathcal{P}(\mathcal{Y})$, let $(\phi_{\mu_0}, \psi_{\mu_0})$, $(\phi_{\mu_1}, \psi_{\mu_1}) \in \mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{Y})$ be pairs of Kantorovich potentials respectively for $\mathcal{T}'(\rho, \mu_0)$ and $\mathcal{T}'(\rho, \mu_1)$, then*

$$\mathrm{Var}_\rho\left(\phi_{\mu_0} - \phi_{\mu_1}\right) \leq 2M_{\phi_{\mu_0},\phi_{\mu_1}} \langle \mu_0 - \mu_1 | \psi_{\mu_0} - \psi_{\mu_1}\rangle. \tag{5.1}$$

Recall that, by (4.10) for $p = 2$, we have $M_{\phi_{\mu_0},\phi_{\mu_1}} \leq 2R_{\mathcal{X}}(R_{\mathcal{X}} + R_{\mathcal{Y}})$.

Proof of Theorem 2.1 for $p = 2$. Let us now prove Theorem 2.1 from Theorem 5.1. As it is a Kantorovitch potential for the $'$-cost, ψ_{μ_0} is the $'$-transform of $\phi_{\mu_0} \in \mathcal{C}(\mathcal{X})$. As a consequence, ψ_{μ_0} is $R_{\mathcal{X}}$-Lipschitz. Similarly, ψ_{μ_1} is $R_{\mathcal{X}}$-Lipschitz. Then, using Kantorovitch-Rubinstein duality formula, one gets

$$\langle \mu_0 - \mu_1 | \psi_{\mu_0} - \psi_{\mu_1}\rangle \leq R_{\mathcal{X}} W_1(\mu_0, \mu_1) \tag{5.2}$$

which, put along with Theorem 5.1 yields Theorem 2.1.

□

We collect also as a corollary the following slightly different bound, where we change the point of view by taking a single target measure $\mu \in \mathcal{P}(\mathcal{Y})$ and consider any $\psi \in \mathcal{C}(\mathcal{Y})$ (not necessarily a potential).

Corollary 5.2. *Let ρ be a* log*-concave probability measure with bounded (convex) support $\mathcal{X} \subseteq \mathbb{R}^d$ and let $\mathcal{Y} \subseteq \mathbb{R}^d$ be compact. Given $\mu \in \mathcal{P}(\mathcal{Y})$, let $(\phi_\mu, \psi_\mu) \in \mathcal{C}(\mathcal{Y})$ be a pair of Kantorovich potentials for $\mathcal{T}'(\rho, \mu)$, let $\psi \in \mathcal{C}(\mathcal{Y})$ and $\phi = \psi'$, it holds*

$$\mathrm{Var}_\rho\left(\phi_\mu - \phi\right) \leq 2M_{\phi_\mu,\phi}\left[\mathcal{T}'(\rho, \mu) - (\langle \mu|\psi\rangle + \langle \rho, \phi\rangle)\right]. \tag{5.3}$$

Proof of Corollary 5.2. $\mathcal{K}'$ is a convex functional defined on $\mathcal{C}(\mathcal{Y})$ and therefore differentiable in duality with $\mathcal{M}(\mathcal{Y})$ the space of Radon measures. Moreover, one can easily check that for any $\mu \in \mathcal{P}_2(\mathcal{Y})$, if ψ_μ denotes its Kantorovich potential, then $\psi_\mu \in D(\partial\mathcal{K}')$ and $\partial\mathcal{K}'(\psi_\mu) = \mu$. One concludes by monotonicity of the slopes. □

We now turn to the proof of Theorem 5.1. It is known or easy to check that while strict convexity yields unicity in a minimization problem, strong convexity yields explicit stability estimates. Following that idea developed in [Del22], we aim at a suitable convexity property for the functional $\mathcal{K}'$, which we obtain first in the ε-entropic semi-discrete (i.e., for finite $\mathcal{Y}$) case then via an approximation argument. However, to highlight the differences (and the improvements) with respect to [Del22], we split the proof of Theorem 5.1 into three steps:

*Step 1 (*log*-concavity of $\mathcal{Z}$).* Arguing for fixed $\varepsilon > 0$ and $\mathcal{Y}$ finite, we establish log-concavity of the "partition function" $\mathcal{Z}$. Unlike [Del22, proposition 4.6], where this is obtained as a "black box" application of the Prékopa-Leindler inequality when ρ is the Lebesgue measure on a convex bounded $\mathcal{X}$, we obtain this property for a general log-concave measure ρ through an optimal transport argument (Proposition 5.3), which itself can be used to prove the Prékopa-Leindler inequality, but is more amenable to further generalizations.

Step 2 (modified convexity of $\mathcal{K}'^{,\varepsilon}$). By direct computation, we go from the log-concavity of $\mathcal{Z}$ to a strong convexity-like inequality on $\mathcal{K}'^{,\varepsilon}$, that reads for ψ_0, $\psi_1 \in \mathcal{C}(\mathcal{Y})$,

$$\mathrm{Var}_\rho(\psi_1'^{,\varepsilon} - \psi_0'^{,\varepsilon}) \leq C\langle \nabla\mathcal{K}'^{,\varepsilon}(\psi_1) - \nabla\mathcal{K}'^{,\varepsilon}(\psi_0)|\psi_1 - \psi_0\rangle, \tag{5.4}$$

for a suitable constant C. Let us stress that the inequality above provides quantitative convexity, but not strong convexity in the usual sense, for the dual potentials appear in the left hand side. Still, our use of integration against ρ in the first step allows us to yield a constant independent of bounds above and below on the density of ρ, going beyond the case covered in [Del22]. Moreover, when compared to [Del22], where the (stronger) inequality with $\mathrm{Var}_{(\mu+\nu)/2}(\psi_\mu - \psi_\nu)$ in the left hand side is established, our proof does not rely on linearity of the cost (while [Del22, lemma 2.3] apparently does).

Step 3 (approximation and scaling) This final step consists in moving from the ε-entropic regularization to the classical optimal transport by letting $\varepsilon \downarrow 0$, as briefly discussed in Remark 4.3, and then via an approximation argument to go from the discrete case, i.e., finite $\mathcal{Y}$ to the general case, i.e., $\mathcal{Y}$ compact, using Lemma 4.4. Finally, thanks to a scaling argument we improve the constant in the stability inequality: indeed, it is sufficient to introduce an "inverse temperature" parameter $\beta > 0$, defining

$$\mathcal{Z}_\beta(\psi) := \int_{\mathcal{X}} e^{\beta\psi',\varepsilon} d\rho \tag{5.5}$$

and, after repeating the first two steps, we optimize upon β. This step does not differ substantially from [Del22, section 4.5, section 2.4, proposition 3.3].The scaling argument, also in [Del22], is actually of great importance for the case $1 < p < 2$ and will be detailed in Section 5.2, while adapting it to the present setting is straightforward.

In view of the scheme detailed above, we establish the following proposition (settling *Step 1*).

theo:Iconcav **Proposition 5.3** (log-concavity of $\mathcal{Z}$). *Let ρ be a log-concave measure with (convex) bounded support $\mathcal{X}$, let $\mathcal{Y} \subseteq \mathbb{R}^d$ be finite, $\varepsilon > 0$ and $\beta > 0$. For every ψ_0, $\psi_1 \in \mathcal{C}(\mathcal{Y})$ and $t \in [0,1]$, it holds*

eq:logconcavI}
$$\log \mathcal{Z}_\beta((1-t)\psi^0 + t\psi^1) \geq (1-t)\log \mathcal{Z}_\beta(\psi^0) + t \log \mathcal{Z}_\beta(\psi^1). \tag{5.6}$$

Proof. Recall that we identify $\mathcal{C}(\mathcal{Y})$ with $\mathbb{R}^n$, since $\mathcal{Y} = \{y_i\}_{i=1}^n$, and write for brevity $\psi_i := \psi(y_i)$, $\mu_i := \mu(y_i)$. To simplify the notation, we also simply write ψ' instead of $\psi'^{,\varepsilon}$ and argue in the case $\beta = 1$ only. Thus, given ψ_0, $\psi_1 \in \mathbb{R}^n$ and $x_0, x_1 \in \mathcal{X}$ and setting $\psi_t = (1-t)\psi_0 + t\psi_1$, we have

$$\begin{aligned} -(\psi_t)'(x_t) &= \varepsilon \log\Big(\sum_{i=1}^n \exp\Big(\frac{\langle x_t, y_i\rangle + (\psi_t)_i}{\varepsilon}\Big)\sigma_i\Big) \\ &= \varepsilon \log\Big(\sum_{i=1}^n \exp\Big(\frac{(1-t)\langle x_0, y_i\rangle + t\langle x_1, y_i\rangle + (1-t)(\psi_0)_i + t(\psi_1)_i}{\varepsilon}\Big)\sigma_i\Big). \end{aligned} \tag{5.7}$$

Using Hölder's inequality, we get

$$-(\psi_t)'(x_t) \leq -(1-t)(\psi_0)'(x_0) - t(\psi_1)'(x_1) \tag{5.8}$$

Therefore, taking the exponential and setting $h_t(x) = \exp((\psi_t)'(x))$ we obtain the inequality

{eq:hypPL}
$$h_t(x_t) \geq h_0(x_0)^{1-t} h_1(x_1)^t \tag{5.9}$$

As mentioned earlier, one can now use as a black box the Prékopa-Leindler inequality for log-concave measures which yields $\|h_t\|_{L^1(\rho)} \geq \|h_0\|_{L^1(\rho)}^{1-t}\|h_1\|_{L^1(\rho)}^t$. For completeness we recall the optimal transport version of the proof. Up to renormalization, we can always assume $\|h_0\|_{L^1(\rho)} = \|h_1\|_{L^1(\rho)} = 1$, so that h_0 and h_1 are probability densities with respect to ρ.

By Lemma 4.6 applied with $p = 2$, $\rho_0 := h_0\rho$, $\rho_1 := h_1\rho$, using also the log-concavity of ρ and writing ρ_t the displacement interpolation between $h_0\rho$ and $h_1\rho$, we get for $h_0\rho$-a.e. $x \in \mathcal{X}$ the inequality

$$\begin{aligned} \rho_t(T_t(x)) &\leq (h_0(x)\rho(x))^{1-t}(h_1(T(x))\rho(T(x)))^t \\ &\leq h_0(x)^{1-t}(h_1(T(x)))^t \rho(x)^{1-t}\rho(T(x))^t \\ &\leq h_0(x)^{1-t}(h_1(T(x)))^t \rho(T_t(x)) \\ &\leq h_t(T_t(x))\rho(T_t(x)). \end{aligned} \tag{5.10}$$

Therefore, we deduce that

$$\rho_t(x) \leq h_t(x)\rho(x) \quad \text{holds } \rho_t \text{ a.e.} \tag{5.11}$$

Integrating thus yields $\|h_t\|_{L^1(\rho)} \geq 1$. Introducing back $\|h_0\|_{L^1(\rho)}$ and $\|h_1\|_{L^1(\rho)}$, we get the Prékopa-Leindler inequality :

$$\|h_t\|_{L^1(\rho)} \geq \|h_0\|_{L^1(\rho)}^{1-t} \|h_1\|_{L^1(\rho)}^{t}. \tag{5.12}$$

Taking the log and noticing that $\|h_t\|_{L^1(\rho)} = \mathcal{Z}(\psi_t)$ yields the thesis. □

In order to address *Step 2* with lighter notations, fixing ψ_0, $\psi_1 \in \mathcal{C}(\mathcal{Y})$ and $\psi_t = (1-t)\psi_0 + t\psi_1$, we write $\pi_t(.|x) := \pi'_{\varepsilon,\psi_t}(.|x)$, so that

{eq:hessI}

$$\langle v, \nabla^2 \log \mathcal{Z}(\psi_t) v\rangle = -\frac{1}{\varepsilon}\int_{\mathcal{X}} \mathrm{Var}_{\pi_t(.|x)}(v) d\rho^\varepsilon_\psi(x) + \mathrm{Var}_{\rho^\varepsilon_\psi}\left(\mathsf{m}_{\pi_t(.|x)}(v)\right), \tag{5.13}$$

op:Kstrongconv

Proposition 5.4. *Let ρ be a* log*-concave probability measure with compact support $\mathcal{X} \subset \mathbb{R}^d$, let $\mu \in \mathcal{P}(\mathcal{Y})$ with $\mathcal{Y} \subseteq \mathbb{R}^d$ finite and let $\varepsilon > 0$. Given ψ_0, $\psi_1 \in \mathcal{C}(\mathcal{Y})$, set*

$$\phi^\varepsilon_t := ((1-t)\psi_0 + t\psi_1)^{\prime,\varepsilon} \quad \textit{for } t \in [0,1], \tag{5.14}$$

Then,

strongconvex}

$$\mathrm{Var}_\rho(\phi^\varepsilon_1 - \phi^\varepsilon_0) \leq C\langle \nabla\mathcal{K}^{\prime,\varepsilon}(\psi_1) - \nabla\mathcal{K}^{\prime,\varepsilon}(\psi_0)|\psi_1 - \psi_0\rangle \tag{5.15}$$

with $C := \exp(2\sup_{t\in[0,1]} \mathrm{osc}(\phi^\varepsilon_t))$.

Proof. As usual, we identify $\mathcal{C}(\mathcal{Y})$ with $\mathbb{R}^n$. From Proposition 5.3, we get that $\nabla^2 \log \mathcal{Z}(\psi_t) \in \mathbb{R}^{n\times n}$ is a symmetric negative semi-definite matrix, which yields for every $v \in \mathbb{R}^n$,

$$\left\langle v, \nabla^2 \log \mathcal{Z}(\psi_t) v\right\rangle \leq 0 \quad \forall\, 0 \leq t \leq 1.$$

By the identity (5.13), we get

essian-I-used}

$$\mathrm{Var}_{\rho^\varepsilon_{\psi_t}}\left(\mathsf{m}_{\pi_t(.|x)}(v)\right) \leq \frac{1}{\varepsilon}\int_{\mathcal{X}} \mathrm{Var}_{\pi_t(.|x)}(v) d\rho^\varepsilon_{\psi_t} \quad \forall\, 0 \leq t \leq 1. \tag{5.16}$$

Our goal is now to change the integration against $\rho^\varepsilon_{\psi_t}$ to integration against ρ. Recalling the the definition of ρ^ε_ψ in (4.21), we deduce

ineq:boundrho}

$$e^{-\mathrm{osc}(\phi^\varepsilon_t)}\rho^\varepsilon_{\psi_t} \leq \rho \leq e^{\mathrm{osc}(\phi^\varepsilon_t)}\rho^\varepsilon_{\psi_t}. \tag{5.17}$$

We remark that here is where the choice of ρ as a "reference" measure in the definition of $\mathcal{Z}$, instead e.g. of the Lebesgue measure on $\mathcal{X}$, turns out to be useful, as it frees us from making assumptions on the density of ρ (besides its log-concavity). Using **??** in both sides of (5.16) thus find, for every $0 \leq t \leq 1$,

ariance-bound}

$$\mathrm{Var}_\rho\left(\mathsf{m}_{\pi_t(.|x)}(v)\right) \leq \frac{e^{2\,\mathrm{osc}(\phi^\varepsilon_t)}}{\varepsilon}\int_{\mathcal{X}} \mathrm{Var}_{\pi_t(.|y)}(v) d\rho = e^{2\,\mathrm{osc}(\phi^\varepsilon_t)}\langle v, \nabla^2\mathcal{K}^{\prime,\varepsilon}(\psi_t)v\rangle \tag{5.18}$$

where the second identity follows from (4.18).

Recalling that $\phi^\varepsilon_t := (\psi_t)^{\prime,\varepsilon}$, we have by the chain rule

{eq:derivphit}

$$\frac{d}{dt}\phi^\varepsilon_t(x) = \langle \nabla\psi^{\prime,\varepsilon}(y_i|x)|\psi_1 - \psi_0\rangle = -\mathsf{m}_{\pi_t(.|x)}(\psi_1 - \psi_0) \tag{5.19}$$

hence, by convexity of the variance,

$$\begin{aligned} \mathrm{Var}_\rho\left(\phi^\varepsilon_1 - \phi^\varepsilon_0\right) &= \mathrm{Var}_\rho\left(\int_0^1 \frac{d}{dt}\phi^\varepsilon_t dt\right) \\ &\leq \int_0^1 \mathrm{Var}_\rho\left(\frac{d}{dt}\phi^\varepsilon_t\right) dt \\ &\leq \int_0^1 \mathrm{Var}_\rho\left(\mathsf{m}_{\pi_t(.|x)}(\psi_1 - \psi_0)\right) dt. \end{aligned} \tag{5.20}$$

Using (5.18) with $v = \psi_1 - \psi_0$, we thus obtain

$$
\begin{aligned}
\mathrm{Var}_\rho(\phi_1^\varepsilon - \phi_0^\varepsilon) &\le \int_0^1 e^{2\,\mathrm{osc}(\phi_t^\varepsilon)} \langle \psi_1 - \psi_0, \nabla^2 \mathcal{K}'^{,\varepsilon}(\psi_t)(\psi_1 - \psi_0)\rangle dt \\
&\le C \int_0^1 \langle \psi_1 - \psi_0, \nabla^2 \mathcal{K}'^{,\varepsilon}(\psi_t)(\psi_1 - \psi_0)\rangle dt, \\
&\le C \langle \psi_1 - \psi_0, \int_0^1 \nabla^2 \mathcal{K}'^{,\varepsilon}(\psi_t)(\psi_1 - \psi_0) dt\rangle
\end{aligned} \tag{5.21}
$$

with $C = \exp(2 \sup_{t\in[0,1]} \mathrm{osc}(\phi_t^\varepsilon))$. Finally, using that

$$
\frac{d}{dt} \nabla \mathcal{K}'^{,\varepsilon}(\psi_t) = \nabla^2 \mathcal{K}'^{,\varepsilon}(\psi_t)(\psi_1 - \psi_0), \tag{5.22}
$$

we obtain the thesis. □

Let us recall here that we refer to Remark 4.3 and Lemma 4.4 for *Step 3* and to [Del22, section 4.5, section 2.4, proposition 3.3] for detailed proofs. delalande2022quantitativ

sec:p<2 5.2. **The case $1 < p < 2$.** We finally move to the most challenging case of exponents $p \in (1,2)$ for which in particular, the smoothness or even the concavity condition of the above remark fail. We prove the following inequality.

theo:finalestp **Theorem 5.5.** *Let $p \in (1,2)$, set $q = p/(p-1)$, let ρ be a* log*-concave probability measure with bounded (convex) support $\mathcal{X} \subseteq \mathbb{R}^d$ and let $\mathcal{Y} \subseteq \mathbb{R}^d$ be compact. Given μ_0, $\mu_1 \in \mathcal{P}(\mathcal{Y})$, let $(\phi_{\mu_0}, \psi_{\mu_0})$, $(\phi_{\mu_1}, \psi_{\mu_1}) \in \mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{Y})$ be pairs of Kantorovich potentials respectively for $\mathcal{W}_p(\rho, \mu_0)$ and $\mathcal{W}_p(\rho, \mu_1)$ then, there exists $C = C(p, \rho, \mathcal{Y}) < \infty$ such that*

$$
\mathrm{Var}_\rho(\phi_1 - \phi_0) \le C \langle \mu_1 - \mu_0 | \psi_1 - \psi_0 \rangle^{\frac{2}{q}}. \tag{5.23}
$$

We also obtain the following corollary.

cor:stabp<2 **Corollary 5.6.** *Let $p \in (1,2)$, set $q = p/(p-1)$, let ρ be a* log*-concave probability measure with bounded (convex) support $\mathcal{X} \subseteq \mathbb{R}^d$ and let $\mathcal{Y} \subseteq \mathbb{R}^d$ be compact. Given $\mu \in \mathcal{P}(\mathcal{Y})$, let $(\phi_\mu, \psi_\mu) \in \mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{Y})$ be Kantorovich potentials for $\mathcal{W}_p(\rho, \mu)$, let $\psi \in \mathcal{C}(\mathcal{Y})$ and write $\phi := \psi^c$. Then, there exists $C = C(p, \rho, \mathcal{Y}) < \infty$ such that it holds*

$$
\mathrm{Var}_\rho(\phi_\mu - \phi) \le C \left[\mathcal{W}_p(\rho, \mu)^p - (\langle \mu | \psi \rangle + \langle \rho | \phi \rangle)\right]^{\frac{2}{q}}. \tag{5.24}
$$

As already noticed, we cannot restrict ourselves to the previous $\mathcal{C}^2$ case as the cost does not have bounded derivatives on the diagonal. We will derive a weaker version of *Step 1*, before performing *Step 3* as the scaling argument will be necessary to control the additional term.

prop:Iconcavp **Proposition 5.7.** *Let ρ be* log*-concave with (convex) bounded support $\mathcal{X}$, $\mathcal{Y} \subseteq \mathbb{R}^d$ be finite, $\varepsilon > 0$ and $\beta > 0$. For every ψ_0, $\psi_1 \in \mathcal{C}(\mathcal{Y})$ and $t \in [0,1]$, it holds*

eq:logconcavI}

$$
\begin{aligned}
\log \mathcal{Z}_\beta((1-t)\psi^0 + t\psi^1) \ge{}& (1-t) \log \mathcal{Z}_\beta(\psi^0) + t \log \mathcal{Z}_\beta(\psi^1) \\
&- \beta t (1-t) \gamma \mathcal{W}_p(\rho^\epsilon_{\beta,\psi_0}, \rho^\epsilon_{\beta,\psi_1})^p.
\end{aligned} \tag{5.25}
$$

where $\gamma = \gamma(p) < \infty$ and $\rho^\epsilon_{\beta,\psi_0}$ is defined as in (4.21) (with ψ_0 instead of ψ).

Proof of Proposition 5.7. Recall that we identify $\mathcal{C}(\mathcal{Y})$ and $\mathbb{R}^n$ and again for simplicity we prove the result only for $\beta = 1$. Let us first recall that, for every x_0, $x_1 \in \mathbb{R}^d$, with the notation $x_t = (1-t)x_0 + t x_1$, there exists a constant $\gamma = \gamma(p)$ such that the following inequality holds:

{eq:pPC}

$$
|x_t|^p \ge (1-t)|x_0|^p + t|x_1|^p - \gamma t (1-t) |x_0 - x_1|^p \quad \text{for every } t \in [0,1]. \tag{5.26}
$$

Therefore, letting ψ_0, $\psi_1 \in \mathbb{R}^n$, it holds

$$
\begin{aligned}
-(\psi_t)^{c,\epsilon}(x_t) &= \epsilon \log\left[\sum_{i=1}^n \exp\left(\frac{(\psi_t)_i - |x_t - y_i|^p}{\epsilon}\right)\sigma_i\right] \\
&\leq \epsilon \log\left[\sum_{i=1}^n \exp\left(\frac{(1-t)(\psi_0)_i + t(\psi_1)_i - (1-t)|x_0 - y_i|^p - t|x_1 - y_i|^p}{\epsilon}\right)\cdot\right. \\
&\quad \left.\cdot \exp\left(\frac{t(1-t)|x_0 - x_1|^p}{\epsilon}\right)\sigma_i\right]
\end{aligned} \tag{5.27}
$$

and we thus get, applying Holder :

$$
(\psi_t)^{c,\epsilon}(x_t) \geq (1-t)(\psi_0)^{c,\epsilon}(x_0) + t(\psi_1)^{c,\epsilon}(x_1)) - t(1-t)\gamma|x_0 - x_1|^p. \tag{5.28}
$$

Therefore, taking the exponential and setting $h_t(x) = \exp((\psi_t)^{c,\varepsilon}(x))$ we get

$$
h_t(x_t) \geq h_0(x_0)^{1-t} h_1(x_1)^t \exp(-t(1-t)\gamma|x_0 - x_1|^p) \tag{5.29}
$$

We next obtain from this bound an integral inequality, in a similar fashion than the Prékopa-Leindler inequality. Up to renormalization, we can always assume $\|h_0\|_{L^1(\rho)} = \|h_1\|_{L^1(\rho)} = 1$, so that h_0 and h_1 are probability densities with respect to ρ.

By Lemma 4.6 applied to the p-cost, $\rho_0 := h_0\rho$, $\rho_1 := h_1\rho$, using also the log-concavity of ρ and writing ρ_t for the displacement interpolation between $h_0\rho$ and $h_1\rho$, we get, for $h_0\rho$-a.e. $x \in \mathcal{X}$, the following inequality :

$$
\begin{aligned}
\rho_t(T_t(x)) &\leq (h_0(x)\rho(x))^{1-t}(h_1(T(x))\rho(T(x)))^t \\
&\leq h_0(x)^{1-t}(h_1(T(x)))^t \rho(x)^{1-t}\rho(T(x))^t \\
&\leq h_0(x)^{1-t}(h_1(T(x)))^t \rho(T_t(x)) \\
&\leq h_t(T_t(x))\rho(T_t(x)) \exp(t(1-t)\gamma|x - T(x)|^p).
\end{aligned} \tag{5.30}
$$

so that

$$
\rho(x)\exp(-t(1-t)\gamma|(T_t)^{-1}(x) - T((T_t)^{-1}(x))|^p) \leq h_t(x)\rho(x) \tag{5.31}
$$

holds ρ_t a.e., where we the inverse of T_t also makes sense ρ_t a.e.. Integrating over $\mathcal{X}$ and using that $\rho_t = (T_t)_\sharp(h_0\rho)$,

$$
\begin{aligned}
\|h_t\|_1 &\geq \int_{\mathcal{X}} \exp(-\gamma t(1-t)|(T_t)^{-1}(x) - T((T_t)^{-1}(x))|^p)\rho_t(x)dx \\
&\geq \int_{\mathcal{X}} \exp(-\gamma t(1-t)|x - T(x)|^p) h_0(x) d\rho(x).
\end{aligned} \tag{5.32}
$$

Applying Jensen's inequality with the exponential function we get, introducing back the norms of h_0 and h_1,

$$
\|h\|_{L^1(\rho)} \geq \|h_0\|_{L^1(\rho)}^{1-t}\|h_1\|_{L^1(\rho)}^t \exp\left(-\gamma t(1-t)\int_{\mathcal{X}} |x - T(x)|^p h_0(x) d\rho(x)\right). \tag{5.33}
$$

Finally, taking the log yields the thesis:

$$
\begin{aligned}
\log \mathcal{Z}((1-t)\psi^0 + t\psi^1) &\geq (1-t)\log \mathcal{Z}(\psi^0) + t\log \mathcal{Z}(\psi^1) \\
&\quad - t(1-t)\gamma \mathcal{W}_p(\rho^\varepsilon_{\psi_0}, \rho^\varepsilon_{\psi_1})^p.
\end{aligned}
$$

□

Using the established bound for $\log \mathcal{Z}_\beta$, we prove the main result for this section.

Proof of Theorem 5.5. From Proposition 5.7, writing $I_\beta := \log \mathcal{Z}_\beta$ for brevity, one gets for all $t \in (0,1)$,

$$
\begin{aligned}
\beta\gamma \mathcal{W}_p(\rho^\varepsilon_{\beta,\psi_0}, \rho^\varepsilon_{\beta,\psi_1})^p &\geq \frac{(1-t)I_\beta(\psi^0) + tI_\beta(\psi^1) - I_\beta((1-t)\psi^0 + t\psi^1)}{t(1-t)} \\
&\geq -\int_0^1 \langle \nabla I_\beta(\psi_{st}) | \psi_1 - \psi_0\rangle ds + \int_0^1 \langle \nabla I_\beta(\psi_{1-(1-t)s}) | \psi_1 - \psi_0\rangle ds.
\end{aligned} \tag{5.34}
$$

Letting successively $t \to 0$ and $t \to 1$ we find

$$\beta\gamma\mathcal{W}_p(\rho^\varepsilon_{\beta,\psi_0},\rho^\varepsilon_{\beta,\psi_1})^p \geq -\langle \nabla I_\beta(\psi_0)|\psi_1-\psi_0\rangle + \int_0^1 \langle \nabla I_\beta(\psi_{1-s})|\psi_1-\psi_0\rangle ds \tag{5.35}$$

and

$$\beta\gamma\mathcal{W}_p(\rho^\varepsilon_{\beta,\psi_0},\rho^\varepsilon_{\beta,\psi_1})^p \geq -\int_0^1 \langle \nabla I_\beta(\psi_s)|\psi_1-\psi_0\rangle ds + \langle \nabla I_\beta(\psi_1)|\psi_1-\psi_0\rangle \tag{5.36}$$

so that, by summing the two inequalities,

$$2\beta\gamma\mathcal{W}_p(\rho^\varepsilon_{\beta,\psi_0},\rho^\varepsilon_{\beta,\psi_1})^p \geq \langle \nabla I_\beta(\psi_1)-\nabla I_\beta(\psi_0)|\psi_1-\psi_0\rangle \tag{5.37}$$

Following the proof of *Step 2* as proposed in Proposition 5.4 and letting for $\phi^\varepsilon_t = \psi^{c,\varepsilon}_t$, for $t\in[0,1]$ where $\psi_t = (1-t)\psi_0 + t\psi_1$, $M^\varepsilon := \sup_{t\in[0,1]}\{\mathrm{osc}(\phi^\varepsilon_t)\}$, and for all $\phi\in\mathcal{C}(\mathcal{X})$,

$$\rho_\phi := \frac{e^\phi}{\int_\mathcal{X} e^\phi d\rho}\rho, \tag{5.38}$$

we find

dadiscreteeps}

$$\beta e^{-\beta M^\varepsilon}\,\mathrm{Var}(\phi^\varepsilon_1-\phi^\varepsilon_0) \leq e^{\beta M^\varepsilon}\langle \nabla\mathcal{K}^{c,\varepsilon}(\psi_1)-\nabla\mathcal{K}^{c,\varepsilon}(\psi_0)|\psi_1-\psi_0\rangle + 2\gamma\mathcal{W}_p(\rho_{\beta\phi^\varepsilon_0},\rho_{\beta\phi^\varepsilon_1})^p. \tag{5.39}$$

If we then let $\varepsilon\downarrow 0$ minding Remark 4.3, by choosing for simplicity the potentials so that ϕ_k have zero ρ-mean, we get that for all $\beta>0$ the following bound holds:

nalestplambda}

$$\beta e^{-\beta M_{\phi_0,\phi_1}}\|\phi_1-\phi_0\|^2_{L^2(\rho)} \leq e^{\beta M_{\phi_0,\phi_1}}\langle \mu_0-\mu_1|\psi_1-\psi_0\rangle + 2\gamma\mathcal{W}_p(\rho_{\beta\phi_0},\rho_{\beta\phi_1})^p, \tag{5.40}$$

The next step is to derive an estimate of $\mathcal{W}_p(\rho_{\beta\phi_0},\rho_{\beta\phi_1})$. Again, for simplicity we argue in the case $\beta=1$. First, we notice that by the monotonicity of Wasserstein distances, we have $\mathcal{W}_p(\rho_{\phi_0},\rho_{\phi_1})\leq\mathcal{W}_2(\rho_{\phi_0},\rho_{\phi_1})$. Then, we apply Lemma 4.7, with $\mu_k=\rho_{\phi_k}$, $k\in\{0,1\}$, so that $f_k\geq e^{-M_{\phi_0,\phi_1}}$ and we get

{eq:W2boundL2}

$$\mathcal{W}_2^2(\rho_{\phi_1},\rho_{\phi_0}) \leq C_\rho e^{M_{\phi_0,\phi_1}}\left\|\frac{e^{\phi_1}}{\int_\mathcal{X} e^{\phi_1}d\rho}-\frac{e^{\phi_0}}{\int_\mathcal{X} e^{\phi_0}d\rho}\right\|^2_{L^2(\rho)}. \tag{5.41}$$

For any $x\in\mathcal{X}$, we have

$$\begin{aligned}\left|\frac{e^{\phi_1(x)}}{\int_\mathcal{X} e^{\phi_1}d\rho}-\frac{e^{\phi_0(x)}}{\int_\mathcal{X} e^{\phi_0}d\rho}\right| &\leq \left|\frac{e^{\phi_1(x)}-e^{\phi_0(x)}}{\int_\mathcal{X} e^{\phi_1}d\rho}\right| + e^{\phi_0(x)}\left|\frac{1}{\int_\mathcal{X} e^{\phi_1}d\rho}-\frac{1}{\int_\mathcal{X} e^{\phi_0}d\rho}\right| \\ &\leq e^{M_{\phi_0,\phi_1}}|e^{\phi_1(x)}-e^{\phi_0(x)}| + e^{2M_{\phi_0,\phi_1}}\left|\int_\mathcal{X}\left(e^{\phi_1}-e^{\phi_0}\right)d\rho\right| \\ &\leq e^{2M_{\phi_0,\phi_1}}\Big(|\phi_1(x)-\phi_0(x)| + \|\phi_1-\phi_0\|_{L^1(\rho)}\Big)\end{aligned} \tag{5.42}$$

where the zero ρ-mean property of ϕ_k assures that

$$-M_{\phi_0,\phi_1}\leq \min\phi_k\leq\max\phi_k\leq M_{\phi_0,\phi_1}.$$

Integrating with respect to ρ yields

$$\left\|\frac{e^{\phi_1}}{\int_\mathcal{X} e^{\phi_1}d\rho}-\frac{e^{\phi_0}}{\int_\mathcal{X} e^{\phi_0}d\rho}\right\|^2_{L^2(\rho)} \leq C_\mathcal{X} e^{4M_{\phi_0,\phi_1}}\|\phi_1-\phi_0\|^2_{L^2(\rho)}. \tag{5.43}$$

From this estimation, (5.40) and (5.41) we get

$$\beta e^{-\beta M_{\phi_0,\phi_1}}\|\phi_1-\phi_0\|^2_{L^2(\rho)} \leq e^{\beta M_{\phi_0,\phi_1}}\langle \mu_0-\mu_1|\psi_1-\psi_0\rangle + \gamma C_\rho e^{\frac{5p\beta}{2}M_{\phi_0,\phi_1}}\beta^p\|\phi_1-\phi_0\|^p_{L^2(\rho)} \tag{5.44}$$

so,

$$\beta e^{-2\beta M_{\phi_0,\phi_1}}\|\phi_1-\phi_0\|^2_{L^2(\rho)} - \gamma C_\rho e^{\frac{(5p-2)\beta}{2}M_{\phi_0,\phi_1}}\beta^p\|\phi_1-\phi_0\|^p_{L^2(\rho)} \leq \langle \mu_0-\mu_1|\psi_1-\psi_0\rangle. \tag{5.45}$$

Setting $C=\gamma C_\rho$, $\alpha = 2M_{\phi_0,\phi_1}$ and $\alpha' = \frac{(5p-2)}{2}M_{\phi_0,\phi_1}$ so that $C,\alpha,\alpha'>0$ and $h(\beta) := \beta e^{-\alpha\beta} - Ce^{\alpha'\beta}\beta^p$, finally optimizing in β we get

{eq:uglyccl}

$$\min(\|\phi_1-\phi_0\|^2_{L^2(\rho)}, \|\phi_1-\phi_0\|^{\frac{p}{p-1}}_{L^2(\rho)}) \leq \frac{1}{\sup\limits_{\beta\in\mathbb{R}_+} h(\beta)}\langle \mu_0-\mu_1|\psi_1-\psi_0\rangle. \tag{5.46}$$

Now, deriving the absolute bound

$$\begin{aligned}\|\phi_1-\phi_0\|_{L^2(\rho)} &\leq |\mathcal{X}|(\|\phi_0\|_{L^\infty}+\|\phi_1\|_{L^\infty})\\ &\leq 2|\mathcal{X}|M_{\phi_0,\phi_1} \leq C = C(\rho,p,\mathcal{Y})\end{aligned} \tag{5.47}$$

we get $\|\phi_1-\phi_0\|^{\frac{p}{p-1}}_{L^2(\rho)} \leq C\|\phi_1-\phi_0\|^2_{L^2(\rho)}$. From this last inequality and (5.46) we can conclude

$$\|\phi_1-\phi_0\|^{\frac{p}{p-1}}_{L^2(\rho)} \leq C\langle \mu_0-\mu_1|\psi_1-\psi_0\rangle \tag{5.48}$$

for some $C = C(p,\rho,\mathcal{X},\mathcal{Y}) > 0$. One final application of Lemma 4.4 extends the bounds from the semi-discrete case to the general case of $\mathcal{Y}$. □

References

AGS [AGS08] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Second edition. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2008. ISBN: 978-3-7643-8721-1 (cit. on pp. 2, 12).

mbrosio2019pde [AST19] L. Ambrosio, F. Stra, and D. Trevisan. "A PDE approach to a 2-dimensional matching problem". In: *Probability Theory and Related Fields* 173 (2019), pp. 433–477 (cit. on p. 12).

22quantitative [Del22] A. Delalande. "Quantitative Stability in Quadratic Optimal Transport". PhD thesis. Université Paris-Saclay, 2022 (cit. on pp. 9–14, 16).

21quantitative [GHO21] M. Goldman, M. Huesmann, and F. Otto. "Quantitative Linearization Results for the Monge-Ampère Equation". In: *Comm. Pure Appl. Math.* 74.12 (2021), pp. 2483–2560 (cit. on p. 12).

Le17 [Led17] M. Ledoux. "On optimal matching of Gaussian samples". In: *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* 457.Veroyatnost' i Statistika. 25 (2017), pp. 226–264. ISSN: 0373-2703 (cit. on p. 12).

McCann [McC94] R. McCann. "A convexity theory for interacting gases and equilibrium crystals". In: *PhD Thesis, Princeton Univ.* (1994) (cit. on p. 12).

ino2020optimal [MG20] S. D. Marino and A. Gerolin. "An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm". In: *Journal of Scientific Computing* 85.2 (2020), p. 27 (cit. on p. 10).

timaltransport [MT24] O. Mischler and D. Trevisan. *Quantitative stability in optimal transport for general power costs.* 2024. arXiv: 2407.19337 [math.FA]. URL: https://arxiv.org/abs/2407.19337 (cit. on pp. 2, 12).

tz2022entropic [NW22] M. Nutz and J. Wiesel. "Entropic optimal transport: Convergence of potentials". In: *Probability Theory and Related Fields* 184.1 (2022), pp. 401–424 (cit. on p. 10).

2018comparison [Pey18] R. Peyre. "Comparison between W_2 distance and $\dot{H}^{-1}$ norm, and localization of Wasserstein distance". English. In: *ESAIM, Control Optim. Calc. Var.* 24.4 (2018), pp. 1489–1501. ISSN: 1292-8119 (cit. on p. 12).

ani2009optimal [Vil+09] C. Villani et al. *Optimal transport: old and new.* Vol. 338. Springer, 2009 (cit. on p. 2).

(Octave Mischler) DMA, Ecole Normale Supérieure PSL, 45 rue d'Ulm, 75005 Paris 5, France
Email address: octave.mischler@ens.psl.eu