# Quantitative stability in optimal transport

Cyril Letrouit

# Foreword

After the foundational works of Gaspard Monge (1781), Leonid Kantorovich (1942), and many others, optimal transport has witnessed explosive growth since the 1980s. This owes much to the universality of the question it addresses: what is the most efficient way of moving "mass" from one distribution to another, given a cost of moving mass between points? This basic question lies at the crossroads of numerous mathematical fields nowadays, including analysis, probability, statistics, geometry and optimization, and optimal transport has led to powerful applications, notably in machine learning, economics, physics, computer vision and biology.

A fundamental object in optimal transport is the *optimal transport map* $T_{\rho\to\mu}$ connecting two probability measures $\rho$ and $\mu$. Brenier showed at the turn of the 1990s its existence and uniqueness under natural assumptions, for instance for the quadratic cost in $\mathbb{R}^d$, when $\rho, \mu$ have second moments and $\rho$ has a density with respect to the Lebesgue measure. Moreover, any optimal transport map is the gradient of a convex function, called a *Kantorovich potential.*

For many reasons, not only existence and uniqueness of optimal transport maps, but also their *stability* with respect to variations of the marginal distributions $\rho$ and $\mu$ is of fundamental importance: taken together, existence, uniqueness and stability form the three pillars of a well-posed problem. For instance, in numerical simulations, the source and the target measures are replaced by discrete approximations and it is desirable to quantify the discrepancy between the optimal transport map and its numerical approximation. In statistics, where the marginals come from real data and are only known through samples, one may estimate the optimal transport map using the samples, but it raises the question of the accuracy of this estimator compared to the "true" optimal transport map. Fundamentally, this is also a stability problem. In both numerical and statistical applications, the main question is thus the following one:

*for a fixed source density $\rho$, does a small modification of the target measure $\mu$ result in a small modification of the optimal transport map $T_{\rho\to\mu}$? If yes, is it possible to quantify the answer?*

*Qualitative* stability results have long been established: already Brenier in his seminal 1991 paper [20] had discovered that under the same assumptions on the source measure $\rho$ as in his existence and uniqueness theorem, the map $\mu \mapsto T_{\rho\to\mu}$ is continuous in the natural topologies. In other words, a small perturbation of the target measure results in a small perturbation of the optimal transport map. This continuity statement is however only qualitative and does not come with effective bounds, whereas for both numerical and theoretical purposes, *quantitative estimates* are often needed.

It is only recently that the first quantitative stability bounds have been proved, with key contributions by Gigli, Berman, Delalande, Mérigot, and others. The interest in the subject has also been strengthened by applications, notably to statistics, to the geometry of the Wasserstein space and to the convergence of the Sinkhorn algorithm, which is used to solve a regularized version of the optimal transport problem, called entropic optimal transport. It is our purpose in these lecture notes to review the theoretical advances on quantitative stability estimates, to point out what is not yet understood, and to highlight the applications of this emerging and flourishing field.

These notes are based on my Cours Peccot delivered at the Collège de France in May-June 2025. Preparing these lectures has been a great source of pleasure for me, and I would like to

thank all attendees, as well as the Collège de France for posting online the videos of the lectures[1]. I express my deepest thanks to Quentin Mérigot for this very fruitful collaboration, and to Jun Kitagawa for our successful collaboration on a second paper. I also greatly benefited from numerous discussions on optimal transport with many colleagues, in particular with Yann Brenier, Guillaume Carlier, Vincent Divol, William Ford, Michael Goldman, Marc Hallin, Antoine Julia, Tudor Manole, Jonathan Niles-Weed, Pierre Pansu, Gabriel Peyré, Aram Pooladian, Philippe Rigollet and François-Xavier Vialard. I finally thank my wife Claire for everything (and for attending one of my lectures!).

The manuscript is organized into five chapters, and Chapters 4 and 5 may be read independently of the rest. To lighten the presentation, most references are collected at the end of each chapter in a bibliographical paragraph. Chapter 1 introduces the problem, and recalls fundamental facts of optimal transport theory, in particular Brenier's theorem. It describes some of the initial motivations for developing a quantitative theory of stability, notably the so-called *linearized optimal transport* framework, which draws upon our knowledge of the geometry of the Wasserstein space. It also presents two important quantitative stability results, whose proofs are the main thread of Chapters 2 and 3. Although relatively recent, the tools developed in Chapter 2 will sound familiar to most people working in optimal transport. The *Kantorovich functional* $\int \psi^* \mathrm{d}\rho$ is the main character of this chapter, and its strong convexity, shown using functional inequalities, is one of the key ingredients to prove quantitative stability estimates. The mathematics of Chapter 3 might sound more exotic: they find their roots in the proofs of Sobolev–Poincaré inequalities developed in the 1980s, as well as in spectral graph theory. However, they provide natural and very efficient tools to extend the stability estimates of Chapter 2. To go beyond quadratic optimal transport in Euclidean spaces, we add in Chapter 4 yet another ingredient: *entropic optimal transport*. This penalized version of the original optimal transport problem, which is a very efficient computational tool, gives us a regularized analogue of the Kantorovich functional whose strong concavity makes it possible to handle more general costs. We conclude by presenting three important applications of optimal transport stability bounds in Chapter 5: to statistical optimal transport, Wasserstein barycenters, and to the convergence of the Sinkhorn algorithm. They illustrate the theoretical and the numerical relevance of this theory.

Although this manuscript is mainly devoted to quantitative stability bounds in optimal transport, it also provides short and self-contained introductions to several topics such as statistical optimal transport, entropic optimal transport, Wasserstein barycenters, and spectral graph theory, which, we hope, will stimulate the readers' curiosity and inspire them to delve into these nice research areas.

All comments, suggestions and bug reports are very welcome and can be sent to my email address *cyril.letrouit@universite-paris-saclay.fr.*

Cyril Letrouit
Orsay, September 2025.

[1] Available at `https://www.college-de-france.fr/fr/agenda/conferencier-invite/stabilite-quantitative-du-transport-optimal`

# Contents

# 1 Introduction: optimal transport and stability

## 1.1 The optimal transport problem

The nearly 250 years old Monge transportation problem consists in finding the optimal way to transport mass from a given source to a given target probability measure, while minimizing an integrated cost.

Let $\rho$ be a probability measure on a Polish (i.e., complete, separable metric) space $\mathcal{X}$ and $\mu$ be a probability measure on a Polish space $\mathcal{Y}$. Let $c$ be a non-negative measurable function on $\mathcal{X} \times \mathcal{Y}$. An *admissible mass transport plan* is an element $\gamma$ of the space $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ of probability measures over $\mathcal{X} \times \mathcal{Y}$ whose marginals coincide with $\rho$ and $\mu$, i.e., for all measurable sets $A \subset \mathcal{X}$, $B \subset \mathcal{Y}$,

$$\gamma(A \times \mathcal{Y}) = \rho(A) \qquad \text{and} \qquad \gamma(\mathcal{X} \times B) = \mu(B). \tag{1.1}$$

These conditions mean that for any $x \in \mathcal{X}$, $y \in \mathcal{Y}$, the amount of mass taken from $x$ coincides with $\mathrm{d}\rho(x)$, and the amount of mass arriving at $y$ coincides with $\mathrm{d}\mu(y)$. The set of all admissible transport plans is

$$\Pi(\rho, \mu) = \{\gamma \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \mid (1.1) \text{ holds}\}.$$

It is non-empty and convex. The optimal transport problem with cost $c$ is the minimization problem

$$\inf_{\gamma \in \Pi(\rho,\mu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \mathrm{d}\gamma(x, y). \tag{1.2}$$

A solution to (1.2), i.e., an admissible transport plan $\gamma$ which attains the infimum, is called an optimal transport plan. In the particular case where $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^d$ and $c(x, y) = |x - y|^2$, one finds the *quadratic optimal transport problem*

$$\inf_{\gamma \in \Pi(\rho,\mu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \mathrm{d}\gamma(x, y), \tag{1.3}$$

which will be our main focus in the sequel. The case of $p$-costs $c(x, y) = |x - y|^p$ with $p \geq 1$ is also of interest, and gives rise to the $p$-Wasserstein distance defined as

$$W_p(\rho, \mu) = \left( \inf_{\gamma \in \Pi(\rho,\mu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \mathrm{d}\gamma(x, y) \right)^{1/p}.$$

Wasserstein distances are indeed distances, and they verify $W_p \leq W_q$ for $1 \leq p \leq q$.

Optimal transport and Wasserstein distances are used in an incredible number of fields, among which:

- Engineering. Monge formulated in 1781 the problem of moving a pile of sand from one place to another while minimizing a total transportation cost.
- Economics. Around 1942, Kantorovich was interested in the optimal allocation of resources between $m$ production stations and $n$ consumption stations. He used his duality theory to solve this problem, which is an instance of the optimal transport problem.
- Mathematical physics, modelling and PDEs. Arnold (1966) and then Brenier (1989) studied the Euler equation of fluid mechanics via a least action principle in the space of diffeomorphisms. Otto interpreted in the 1990s the heat equation as a gradient descent of entropy in the geometry of mass transport. In another direction, Tanaka used Wasserstein distances in the 1970s to prove relaxation to equilibrium for the spatially homogeneous

Boltzmann equation. And Caffarelli analyzed in the 1990s the Monge-Ampère partial differential equation $\det(D^2 f) = g$, resulting in a regularity theory for optimal transport maps.

- Geometry and related areas. Optimal transport has also led to the discovery of new functional inequalities with geometric content, and to new proofs of old ones. Among them, we can mention the Brunn-Minkowski, Brascamp-Lieb and Prékopa-Leindler inequalities, which we review in Section 2.4). In metric geometry, optimal transport has been used to give a meaning to Ricci curvature in non-smooth spaces $\mathcal{X}$, relying on displacement convexity of certain functions on the Wasserstein space $\mathcal{P}_2(\mathcal{X})$ (Lott-Sturm-Villani 2004-2009).

- Image processing. Relevant features of images (colors, contours, orientations, textures) may be represented as histograms, or densities. Optimal transport is then used as a powerful tool to compare images, interpolate between them, transfer colors, or segment images.

- Statistics and machine learning. Wasserstein distances are widely used to measure distances between probability distributions, since the work of Dudley in 1969 who analyzed the rate of convergence of empirical probability measures $\mu_n$ to their limit $\mu$. More recently, optimal transport has been used to interpolate multiple data distributions (e.g. samples, images, domains, etc) using Wasserstein barycenters; to analyze the training dynamics of neural networks, and sampling algorithms such as the Langevin Monte Carlo algorithm.

Many other fields, applications and names are of course missing: our purpose is only to illustrate the diversity of fields in which one may encounter optimal transport and Wasserstein distances. Although the cost function $c(x, y) = |x - y|$ might seem more physical at first glance than the quadratic cost – for instance, in problems like moving materials at minimal cost on a construction site –, the quadratic cost $c(x, y) = |x - y|^2$ is actually the one which is most useful in the above applications. This is partly due to the fact that the $W_2$ distance gives a Riemannian structure to the space of probability measures (with notions of geodesics, interpolation, etc), but also to very convenient existence and uniqueness properties.

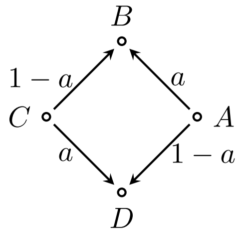
**Existence and uniqueness of optimal transport maps.** Recall that Polish spaces are complete, separable, metric spaces. A solution to (1.3) (or (1.2)) exists under mild assumptions:

**Proposition 1.1.** *If $\mathcal{X}, \mathcal{Y}$ are Polish spaces and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+ \cup \{+\infty\}$ is lower semi-continuous, then there exists a solution to* (1.2).

We shall not prove this proposition, see [105, Theorem 1.3] for a proof. The solution to (1.3) (or (1.2)) is not unique in general. For instance, if $A = (1, 0)$, $B = (0, 1)$, $C = (-1, 0)$, and $D = (0, -1)$ are the vertices of a square in $\mathbb{R}^2$ (endowed with the quadratic cost), there is an infinite number of solutions to (1.3) when $\rho = \frac{1}{2}(\delta_A + \delta_C)$ and $\mu = \frac{1}{2}(\delta_B + \delta_D)$: for any $a \in [0, 1]$,

$$\gamma = \frac{1}{2}\left(a\delta_{(A,B)} + (1-a)\delta_{(A,D)} + (1-a)\delta_{(C,B)} + a\delta_{(C,D)}\right)$$

is an admissible transport plan which is a solution of (1.3). Notice that in this example, the mass leaving $A$ is split into one part going to $B$ and one part going to $D$.




Let us pause for a moment and ask what would happen if we did not allow mass-splitting, i.e., if we replace the infimum in (1.3) by a minimization over the admissible transport plans $\gamma \in \Pi(\rho, \mu)$ which are supported on the graph of a univalued map $T : \mathcal{X} \to \mathcal{Y}$: in other words, all the mass at $x \in \mathcal{X}$ will be sent into $T(x) \in \mathcal{Y}$. The condition $\gamma \in \Pi(\rho, \mu)$ then turns into the set of equalities:

$$\text{for any measurable } U \subset \mathcal{Y}, \quad \rho(T^{-1}(U)) = \mu(U)$$

which may be simply stated in terms of the pushforward operation $\#$ as $T_{\#}\rho = \mu$. The associated admissible transport plan is $\gamma = (\mathrm{Id}, T)_{\#}\rho$. We obtain the so-called Monge problem:

$$\inf_{\substack{S:\mathcal{X}\to\mathcal{Y} \\ S_{\#}\rho=\mu}} \int_{\mathbb{R}^d} |x - S(x)|^2 \mathrm{d}\rho(x). \tag{1.4}$$

A solution to (1.4) is called an optimal transport map. Note that the Monge problem does not necessarily have a solution. If $\rho$ is a sum of Dirac masses but $\mu$ is not, then there does not exist any $S : \mathcal{X} \to \mathcal{Y}$ such that $S_{\#}\rho = \mu$.

To avoid this issue, there exists a simple assumption which guarantees that the solution to (1.3) is unique: Brenier showed that the absolute continuity of the source measure $\rho$ together with moments assumptions on $\rho$ and $\mu$ is a sufficient condition for a unique solution to (1.3) to exist. And even more: he shows that in this case, the Monge problem (1.4) has a unique solution $T$, and that these solutions to the two problems are related by $\gamma = (\mathrm{Id}, T)_{\#}\rho$.

In the sequel, $\mathcal{P}(\mathcal{X})$ denotes the set of probability measures on $\mathcal{X} \subset \mathbb{R}^d$, and $\mathcal{P}_p(\mathcal{X})$ is the set of probability measures on $\mathcal{X}$ with finite $p$-th moment:

$$\mathcal{P}_p(\mathcal{X}) = \left\{ \rho \in \mathcal{P}(\mathcal{X}) \mid \int_{\mathcal{X}} |x|^p \mathrm{d}\rho(x) < +\infty \right\}.$$

The weak topology on $\mathcal{P}(\mathcal{X})$ (or topology of weak convergence, or narrow topology) is induced by convergence against $C_b(\mathcal{X})$, i.e., bounded continuous functions.

**Theorem 1.2** (Existence and uniqueness of optimal transport maps, Brenier)**.** *Let $\rho, \mu \in \mathcal{P}_2(\mathbb{R}^d)$ and $c(x, y) = |x - y|^2$ be the quadratic cost on $\mathbb{R}^d$. Assume that $\rho$ is absolutely continuous with respect to the Lebesgue measure. Then there exists between $\rho$ and $\mu$ a $\rho$-almost everywhere unique optimal transport map $T$ and a unique optimal transport plan $\gamma$, and these solutions are related by $\gamma = (\mathrm{Id}, T)_{\#}\rho$. Furthermore, the map $T$ is the gradient of a convex function $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, and if $(\nabla f)_{\#}\rho = \mu$ for some other convex function $f$, then $\nabla f = \nabla \phi$ $\rho$-almost everywhere.*

If the support $\mathcal{X}$ of $\rho$ is the closure of a bounded connected open set, $\phi$ is uniquely determined on $\mathcal{X}$ up to additive constants. As a consequence of Brenier's theorem, for any convex function $\phi$ and any absolutely continuous $\rho \in \mathcal{P}_2(\mathbb{R}^d)$, the map $\nabla\phi$ is the optimal transport map from $\rho$ to $(\nabla\phi)_{\#}\rho$.

To turn (1.3) (or (1.4)) into a well-posed problem in the sense of Hadamard, there only remains to show *stability of the solution $T$* with respect to perturbations of $\rho$ and $\mu$. The question of stability is fundamental both from the theoretical and the numerical point of view. Soft (compactness) arguments provide without any difficulty a qualitative stability result presented in Section 1.2. However, *quantitative results* are needed in most applications, and for this more difficult problem, tools have started to emerge only recently. The purpose of these notes is to review the theoretical advances in this now rapidly developing field, and to discuss applications to various problems.

## 1.2 Stability of optimal transport

The following general qualitative stability result holds:

**Proposition 1.3** (Qualitative stability of plans)**.** *Let $(\rho_k)_{k\in\mathbb{N}}$ converge weakly to $\rho$ and $(\mu_k)_{k\in\mathbb{N}}$ converge weakly to $\mu$. For each $k \in \mathbb{N}$, let $\gamma_k$ be an optimal transport plan between $\rho_k$ and $\mu_k$, and assume that*

$$\liminf_{k\in\mathbb{N}} \int_{\mathcal{X}\times\mathcal{Y}} |x-y|^2 \mathrm{d}\gamma_k(x,y) < +\infty.$$

*Then the optimal transport cost between $\rho$ and $\mu$ is finite and, up to extraction of a subsequence, $\gamma_k$ converges weakly to some optimal transport plan $\gamma$ between $\rho$ and $\mu$.*

We will not give the proof of this proposition, written in detail in [106, Theorem 5.20]. Proposition 1.3 actually holds in general Polish spaces $\mathcal{X}$ and $\mathcal{Y}$, with a continuous cost function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ such that $\inf c > -\infty$. The proof relies on the Prokhorov theorem, used to extract a converging subsequence of plans, and on a characterization of optimal transport plans as cyclically monotone sets. We provide a taste of the latter argument in the 1 dimensional case in Proposition 1.6.

In these lectures, we will fix the source measure $\rho$ and consider stability with respect to the target measure only. We denote by $T_{\rho\to\mu}$ the optimal transport map from $\rho$ to $\mu$ given by Brenier's theorem. The problem we are interested in reads:

*if $\mu$ and $\nu$ are quantitatively close, prove that $T_{\rho\to\mu}$ and $T_{\rho\to\nu}$ are quantitatively close.*

There are several reasons for this choice of fixing the source measure:

- first, because the mapping $\mu \mapsto T_{\rho\to\mu}$ may be used to embed the Wasserstein space (or part of it) into the Hilbert space $L^2(\rho)$ with a controlled distortion. We shall come back to this important idea in Section 1.4.
- Second, because $T_{\rho\to\mu}$ and $T_{\rho\to\nu}$ are in $L^2(\rho)$ according to Brenier's Theorem 1.2, and thus we may measure their distance simply in $L^2(\rho)$. If we had $\rho$ and $\rho'$ as source measures, measuring distances between the maps would be less easy: instead, one would probably measure the Wasserstein distance between optimal transport plans.
- Third, because in some applications, $\rho$ is a perfectly known probability density, e.g. a standard Gaussian.
- Finally, it is sometimes a first step towards a proof of stability with respect to perturbations of both marginals.

To summarize, in these lecture notes, some $\rho \in \mathcal{P}_2(\mathbb{R}^d)$, assumed to be absolutely continuous with respect to the Lebesgue measure, is fixed. Therefore, we shall drop in the notation the reference to this source measure. Given $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ we call

- the *optimal transport map* and denote by $T_\mu \in L^2(\rho)$ the unique solution to (1.4);

- the *Kantorovich potential* the unique convex function $\phi_\mu \in L^2(\rho)$ such that $T_\mu = \nabla\phi_\mu$ and $\int_{\mathcal{X}} \phi_\mu \mathrm{d}\rho = 0$. In these lectures, we focus on the case where the support of $\rho$ is the closure of a connected open set, and thus the Kantorovich potential is always uniquely defined since any two functions whose gradients coincide almost everywhere differ by a constant. Uniqueness may fail, however, if the support of $\rho$ consists of multiple connected components. Nevertheless, most of our results can be adapted to this case at the expense of shifting Kantorovich potentials by an appropriate constant on each connected component. We shall not pursue this here.

The source measure $\rho$ being now fixed, we formulate the qualitative stability of optimal transport maps as follows:

**Proposition 1.4** (Qualitative stability of maps)**.** *The map* $\mu \mapsto T_\mu$ *from* $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ *to* $L^2(\rho)$ *is continuous.*

This result is already contained in the paper by Brenier [20] where he proves Theorem 1.2. We give here a different (self-contained) proof.

*Proof of Proposition 1.4.* Let $(\mu_n)_{n\in\mathbb{N}}$ and $\mu$ be in $\mathcal{P}_2(\mathbb{R}^d)$ such that $W_2(\mu_n, \mu) \to 0$ as $n \to +\infty$. Then $W_2(\rho, \mu_n) \to W_2(\rho, \mu)$ by the triangle inequality, hence

$$\int_{\mathbb{R}^d} |x - T_{\mu_n}(x)|^2 \mathrm{d}\rho(x) \underset{n\to+\infty}{\longrightarrow} \int_{\mathbb{R}^d} |x - T_\mu(x)|^2 \mathrm{d}\rho(x). \tag{1.5}$$

Therefore, $(T_{\mu_n})_{n\in\mathbb{N}}$ is bounded in $L^2(\rho)$.

Let $\Omega = \mathrm{supp}(\rho)$ and

$$K_\varepsilon = \{x \in \mathbb{R}^d \mid |x| \leq \varepsilon^{-1},\ \rho(x) \geq \varepsilon,\ \mathrm{dist}(x, \partial\Omega) \geq \varepsilon\}.$$

Here and in the sequel, we identify the absolutely continuous measure $\rho$ to its density with respect to the Lebesgue measure $\lambda$ on $\mathbb{R}^d$. Let us prove that for any $\varepsilon > 0$,

$$\sup_{n\in\mathbb{N}} \|T_{\mu_n}\|_{L^\infty(K_\varepsilon)} < +\infty. \tag{1.6}$$

For this we rely on the fact that for any convex function $f$ over $\mathbb{R}^d$, any $x \in \mathbb{R}^d$ and $\eta > 0$,

$$\|\partial f\|_{L^\infty(B(x,\eta))} \leq \frac{6}{\omega_d \eta^d} \int_{B(x,4\eta)} |\nabla f| \mathrm{d}\lambda \tag{1.7}$$

where $\omega_d$ denotes the volume of the unit ball in $\mathbb{R}^d$ and

$$\|\partial f\|_{L^\infty(B(x,\eta))} = \sup_{y\in B(x,\eta)} \sup_{g\in\partial f(y)} |g|.$$

The proof of (1.7) is postponed to the end of the proof of Proposition 1.4. Let us deduce (1.6) from (1.7). For any $x$ such that $B(x, 4\varepsilon) \subset K_\varepsilon$,

$$\frac{\omega_d \varepsilon^d}{6} \|T_{\mu_n}\|_{L^\infty(B(x,\varepsilon))} \leq \int_{B(x,4\varepsilon)} |T_{\mu_n}| \mathrm{d}\lambda \leq \frac{1}{\varepsilon} \Big( \int_{B(x,4\varepsilon)} |T_{\mu_n}|^2 \mathrm{d}\rho \Big)^{1/2} \tag{1.8}$$

by applying (1.7) to $\phi_{\mu_n}$, using that $\rho(x) \geq \varepsilon$ on $K_\varepsilon$, and finally applying the Cauchy-Schwarz inequality. Since $(T_{\mu_n})$ is bounded in $L^2(\rho)$, the right-hand side in (1.8) for fixed $\varepsilon > 0$ is uniformly bounded in $n$. Therefore, setting

$$K'_\varepsilon = \{x \in K_\varepsilon \mid B(x, 4\varepsilon) \subset K_\varepsilon\}$$

we obtain that $\|T_{\mu_n}\|_{L^\infty(K'_\varepsilon)}$ is uniformly bounded in $n$. Sending $\varepsilon$ to 0, this implies that $\sup_n \|T_{\mu_n}\|_{L^\infty(K)} < +\infty$ for any compact set $K$ included in the interior of the support of $\rho$. In particular, this implies (1.6).

Let $x_0$ such that $\rho(x_0) > 0$. Then $\phi_{\mu_n}(x_0) < +\infty$ for any $n$, and we may normalize $\phi_{\mu_n}$ in a way that $\phi_{\mu_n}(x_0) = 0$. By Arzelà-Ascoli, up to extraction of a subsequence omitted in the notation, $(\phi_{\mu_n})$ converges toward some $\phi$ uniformly over any $K_\varepsilon$. Of course, $\phi$ is convex, and we denote its gradient by $T$. We also denote by $\langle \cdot, \cdot \rangle$ the Euclidean scalar product. Passing to the limit $n \to +\infty$ in the inequality

$$\phi_{\mu_n}(y) \geq \phi_{\mu_n}(x) + \langle y - x, \nabla \phi_{\mu_n}(x) \rangle$$

yields that any limit point of $(\nabla \phi_{\mu_n}(x))$ is in $\partial \phi(x)$. This proves that at any point $x$ of differentiability of $\phi$, $(\nabla \phi_{\mu_n}(x))$ converges to $T(x) = \nabla \phi(x)$. Since $\phi$ is convex, it is differentiable almost everywhere, thus

$$T_{\mu_n}(x) \to T(x) \text{ for } \rho\text{-almost every } x. \tag{1.9}$$

We deduce using (1.6) and Lebesgue's dominated convergence theorem that

$$(T_{\mu_n})_{n \in \mathbb{N}} \text{ converges strongly to } T \text{ in } L^2(\rho, K_\varepsilon) \text{ for any } \varepsilon > 0. \tag{1.10}$$

Also, since $(T_{\mu_n})_{n\in\mathbb{N}}$ is bounded in $L^2(\rho)$, it converges weakly to some $T' \in L^2(\rho)$ (up to extraction of a subsequence, omitted in the notation), and we deduce from (1.10) that $T' = T$, $\rho$-almost everywhere. Since $\mathrm{Id} \in L^2(\rho)$, $\langle \mathrm{Id}, T_{\mu_n} \rangle_{L^2(\rho)} \to \langle \mathrm{Id}, T \rangle_{L^2(\rho)}$, and plugging into (1.5) we obtain that $\|T_{\mu_n}\|_{L^2(\rho)} \to \|T\|_{L^2(\rho)}$. By a classical argument, strong convergence follows from weak convergence together with convergence in norm; this proves that $(T_{\mu_n})$ converges *strongly* to $T$ in $L^2(\rho)$.

Finally, let us observe that $T_\# \rho = \mu$: indeed, for any open set $A$, using (1.9) and Lebesgue's dominated convergence theorem,

$$\mu_n(A) = \rho(T_{\mu_n}^{-1}(A)) = \int_{\mathbb{R}^d} \mathbf{1}_{\{x | T_{\mu_n}(x) \in A\}} \mathrm{d}\rho \underset{n \to +\infty}{\longrightarrow} \int_{\mathbb{R}^d} \mathbf{1}_{\{x | T(x) \in A\}} \mathrm{d}\rho = \rho(T^{-1}(A))$$

and since the left-hand side converges to $\mu(A)$, we get the result. Since $T$ is the gradient of a convex function, Brenier's theorem implies that $T = T_\mu$ is the optimal transport map from $\rho$ to $\mu$. We conclude that the full sequence $(T_{\mu_n})$ converges strongly to $T$ in $L^2(\rho)$.

There remains to prove (1.7). We follow the proof of [25, Lemma 2.3]. There are two main steps, the first one being to prove the inequality

$$\|\partial f\|_{L^\infty(B(x,\eta))} \leq \frac{1}{\eta} \mathrm{osc}_{B(x,2\eta)}(f) \tag{1.11}$$

where $\mathrm{osc}_A(f) = \sup_{u,v \in A} |f(u) - f(v)|$. Let $y \in B(x, \eta)$ and $g \in \partial f(y)$. We assume $|g| \neq 0$. For any $z \in B(x, 2\eta)$,

$$\langle g, z - y \rangle \leq f(z) - f(y) \leq \mathrm{osc}_{B(x,2\eta)}(f).$$

Choosing $z = y + \eta \frac{g}{|g|}$, we get $|g| \leq \frac{1}{\eta} \mathrm{osc}_{B(x,2\eta)}(f)$. This inequality also holds if $|g| = 0$. Taking the sup over $y \in B(x, \eta)$ and $g \in \partial f(y)$ we get (1.11).

The second step is to prove

$$\mathrm{osc}_{B(x,2\eta)}(f) \leq \frac{6}{\omega_d \eta^{d-1}} \int_{B(x,4\eta)} |\nabla f| \mathrm{d}\lambda. \tag{1.12}$$

Together with (1.11), this immediately yields (1.7). There remains to show (1.12). We consider $y_0 \in \operatorname{argmin}_{B(x,2\eta)} f$, $y_1 \in \operatorname{argmax}_{B(x,2\eta)} f$, and $g_1 \in \partial f(y_1)$. For any $y \in \mathbb{R}^d$ and $g \in \partial f(y)$,

$$|g||y - y_0| \geq \langle g, y - y_0 \rangle \geq f(y) - f(y_0) \geq f(y_1) + \langle g_1, y - y_1 \rangle - f(y_0)$$

and we deduce

$$|g| \geq \frac{\operatorname{osc}_{B(x,2\eta)}(f) + \langle g_1, y - y_1 \rangle}{|y - y_0|}.$$

Let us assume that $g_1 \neq 0$. Then for $y \in B_1 := B(y_1 + \eta \frac{g_1}{|g_1|}, \eta) \subset B(x, 4\eta)$, there holds $\langle g_1, y - y_1 \rangle \geq 0$. It is also the case if $g_1 = 0$. Therefore

$$\int_{B(x,4\eta)} |\nabla f| \mathrm{d}\lambda \geq \int_{B_1} \frac{\operatorname{osc}_{B(x,2\eta)}(f)}{|y - y_0|} \mathrm{d}\lambda \geq \frac{\lambda(B_1)}{6\eta} \operatorname{osc}_{B(x,2\eta)}(f)$$

since $|y - y_0| \leq |y - y_1| + |y_1 - y_0| \leq 6\eta$. Using $\lambda(B_1) = \omega_d \eta^d$, we get (1.12), which concludes the proof. □

The main problem under consideration in these notes will (almost) be the following one: for a given absolutely continuous $\rho \in \mathcal{P}_2(\mathbb{R}^d)$, do there exist constants $C, \alpha > 0$ such that for all $\mu, \nu$ with finite second moment,

$$\|T_\mu - T_\nu\|_{L^2(\rho)} \leq C W_2(\mu, \nu)^\alpha \tag{1.13}$$

holds? More generally, replacing $W_2$ by $W_p$ for some $p \geq 1$, we will consider inequalities of the type

$$\|T_\mu - T_\nu\|_{L^2(\rho)} \leq C W_p(\mu, \nu)^\alpha, \tag{1.14}$$

the strongest one being for $p = 1$ (since $W_p \leq W_q$ for $p \leq q$).

An important observation is that the reverse inequality

$$\|T_\mu - T_\nu\|_{L^2(\rho)} \geq W_2(\mu, \nu) \tag{1.15}$$

always holds: indeed, $\gamma = (T_\mu, T_\nu)_{\#}\rho$ is an admissible transport plan between $\mu$ and $\nu$ (since its marginals are $\mu$ and $\nu$), and its cost

$$\left( \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \mathrm{d}\gamma(x, y) \right)^{1/2} = \left( \int_{\mathbb{R}^d} |T_\mu(x) - T_\nu(x)|^2 \mathrm{d}\rho(x) \right)^{1/2} = \|T_\mu - T_\nu\|_{L^2(\rho)}$$

is by definition not lower than the cost $W_2(\mu, \nu)$ of an optimal transport plan between $\mu$ and $\nu$.

Put together, the inequalities (1.13) and (1.15) imply that the mapping $\mu \mapsto T_\mu$ is a bi-Hölder embedding of the Wasserstein space into $L^2(\rho)$. However, as we shall discuss in more details in Section 1.4, it is known that if $d \geq 3$, then (1.13) *cannot hold uniformly over all probability measures* $\mu$ and $\nu$ on $\mathbb{R}^d$ with finite second moment (the case $d = 2$ seems open). In fact, it is not possible to embed the Wasserstein space into any $L^p$ space, even in a very coarse sense. Nevertheless, what we will discuss in depth in these lectures is that a stability bound such as (1.13) can hold *if one restricts to slightly smaller families of measures* $\mu, \nu$. For instance, we will show that under some assumptions on $\rho$, for any compact set $\mathcal{Y} \subset \mathbb{R}^d$, there exist $C, \alpha > 0$ such that (1.13) holds for any $\mu, \nu$ supported in $\mathcal{Y}$. In this case, since $\mathcal{Y}$ is assumed compact, one seeks for $\alpha$ as large as possible. The largest possible $\alpha$ is sometimes called the stability exponent (associated to $\rho$ and $\mathcal{Y}$) in the sequel.

In these notes, we will also be interested in quantitative stability estimates for Kantorovich potentials, which take the form

$$\|\phi_\mu - \phi_\nu\|_{L^2(\rho)} \leq C' W_{p'}(\mu,\nu)^{\alpha'}. \tag{1.16}$$

Kantorovich potentials are interesting objects on their own, for many reasons. First, many numerical methods used to solve optimal transport problem, for instance semi-discrete optimal transport and dual gradient methods, rely on solving first the dual formulation of the problem, discussed in Section 2.1. In these methods, one computes the Kantorovich potentials first, before taking the gradient to obtain the optimal transport map. Also, the Sinkhorn algorithm, which is one of the best ways to compute solutions to (regularized) optimal transport problems, computes the entropic version of the Kantorovich potentials (see Section 5.3 for more details). Finally, Kantorovich potentials have an economic interpretation which may help understand their meaning (see [106, beginning of Chapter 5]).

Let us already mention that although the first three chapters are focused on the quadratic cost in $\mathbb{R}^d$ given by $c(x,y) = |x-y|^2$, most results remain valid for more general costs, for instance $p$-costs $c(x,y) = |x-y|^p$, $p > 1$ (see Section 4.3) and the quadratic cost $c(x,y) = \frac{1}{2}\mathrm{dist}(x,y)^2$ on Riemannian manifolds (see Section 4.4).

It is clear that inequalities like (1.14) and (1.16) are useful to justify the theoretical consistance of "plugin methods" to compute optimal transport: if we want to compute the optimal transport map or the Kantorovich potential from $\rho$ to $\mu$ but do not know exactly $\mu$ (due to some noise for instance) and have only access to some approximation $\nu$ of $\mu$, these inequalities tell us how close we may expect $T_\nu$ to be from $T_\mu$ (and $\phi_\nu$ from $\phi_\mu$), depending on some Wasserstein distance between $\mu$ and $\nu$.

## 1.3 Stability in 1 dimension

The case where $d = 1$, i.e., $\rho, \mu, \nu$ are probability measures on $\mathbb{R}$, is particularly simple. Indeed, as soon as $\rho$ is absolutely continuous on $\mathbb{R}$, the mapping $\mu \mapsto T_\mu$ is an isometric embedding:

**Proposition 1.5.** *For any absolutely continuous $\rho \in \mathcal{P}(\mathbb{R})$, any $\mu, \nu \in \mathcal{P}(\mathbb{R})$, and any $p \geq 1$, there holds*

$$\|T_\mu - T_\nu\|_{L^p(\rho)} = W_p(\mu,\nu) \tag{1.17}$$

The stability problem is thus completely solvable in this case: the bound (1.13) holds with $C = \alpha = 1$.

We turn to the proof of Proposition 1.5. The main ingredient is to show that

$$\gamma_{\mathrm{opt}} = (T_\mu, T_\nu)_\# \rho \tag{1.18}$$

is an optimal transport plan between $\mu$ and $\nu$. Indeed, (1.17) then follows immediately:

$$W_p^p(\mu,\nu) = \int_{\mathbb{R}^2} |x-y|^p \mathrm{d}\gamma_{\mathrm{opt}}(x,y) = \int_{\mathbb{R}} |T_\mu(x) - T_\nu(x)|^p \mathrm{d}\rho(x) = \|T_\mu - T_\nu\|^p_{L^p(\rho)}.$$

It is clear that $\gamma_{\mathrm{opt}}$ is an admissible transport plan between $\mu$ and $\nu$ since $(T_\mu)_\#\rho = \mu$ and $(T_\nu)_\#\rho = \nu$. The difficulty is to show that it is optimal. If $p = 1$, it is not difficult to check that any transport plan, and in particular $\gamma_{\mathrm{opt}}$, is optimal: all transport plans have the same cost. Therefore we focus on the case $p > 1$ in the sequel.

Given $\rho, \mu \in \mathcal{P}(\mathbb{R})$, we call *monotone transport plan* any transport plan $\gamma \in \Pi(\rho,\mu)$ such that

$$\forall (x,y), (x',y') \in \mathrm{supp}(\gamma),\ x < x' \Rightarrow y \leq y'.$$

**Proposition 1.6.** *Let $p > 1$. For any $\rho, \mu \in \mathcal{P}(\mathbb{R})$, there exists a unique optimal transport plan between $\rho$ and $\mu$, a unique monotone transport plan between $\rho$ and $\mu$, and these two plans coincide.*

*Proof.* We only need to prove that

(i) any optimal transport plan is monotone, and

(ii) there exists a unique monotone transport plan.

We first show (i). The proof is based on the convexity of the $p$-cost: for any $x < x'$ and $y \leq y'$, the inequality

$$|x - y|^p + |x' - y'|^p \leq |x - y'|^p + |x' - y|^p$$

means that it is always less costly to transport mass from $x$ to $y$ and from $x'$ to $y'$ than to "cross trajectories" and transport mass from $x$ to $y'$ and from $x'$ to $y$. Let $\gamma$ be an optimal transport plan (which exists, according to Proposition 1.1). For the sake of a contradiction, assume that it is not monotone, and let $(x, y)$, $(x', y')$ in the support of $\gamma$ such that $x < x'$ and $y > y'$. By the strict convexity of the $p$-cost, there exists $\delta > 0$ such that

$$|x - y|^p + |x' - y'|^p - |x - y'|^p - |x' - y|^p + 4\delta < 0. \tag{1.19}$$

Let $r > 0$ such that $\gamma(B(x, r) \times B(y, r)) > 0$, $\gamma(B(x', r) \times B(y', r)) > 0$, small enough so that $B(x, r) \cap B(x', r) = \emptyset$, $B(y, r) \cap B(y', r) = \emptyset$, and for any $\widetilde{x} \in B(x, r)$, $\widetilde{x}' \in B(x', r)$, $\widetilde{y} \in B(y, r)$, $\widetilde{y}' \in B(y', r)$,

$$\begin{aligned} \left||\widetilde{x} - \widetilde{y}|^p - |x - y|^p\right| &\leq \delta, & \left||\widetilde{x}' - \widetilde{y}'|^p - |x' - y'|^p\right| &\leq \delta \\ \left||\widetilde{x} - \widetilde{y}'|^p - |x - y'|^p\right| &\leq \delta, & \left||\widetilde{x}' - \widetilde{y}|^p - |x' - y|^p\right| &\leq \delta. \end{aligned} \tag{1.20}$$

Let $\gamma_1, \gamma_2$ be two couplings, whose marginals have mass $\varepsilon > 0$ each, such that $\gamma_1$ is supported in $B(x, r) \times B(y, r)$, $\gamma_2$ is supported in $B(x', r) \times B(y', r)$, and $\gamma_i \leq \gamma$ for $i = 1, 2$ (in the sense that $\int f \mathrm{d}\gamma_i \leq \int f \mathrm{d}\gamma$ for any $f \geq 0$). Let $\rho_i$ be the first marginal of $\gamma_i$, and $\mu_i$ be the second marginal of $\gamma_i$. We consider $\sigma_1$ an arbitrary coupling between $\rho_1$ and $\mu_2$, and $\sigma_2$ an arbitrary coupling between $\rho_2$ and $\mu_1$. Then

$$\gamma' = \gamma - \gamma_1 - \gamma_2 + \sigma_1 + \sigma_2$$

is a transport plan between $\rho$ and $\mu$: its marginals coincide with those of $\gamma$ by definition, and it is non-negative since $\gamma_i \leq \gamma$, and the supports of $\gamma_1$ and $\gamma_2$ are disjoint. We observe that $\gamma'$ has strictly lower cost than $\gamma$: indeed,

$$\begin{aligned} \int c \,\mathrm{d}\gamma' - \int c \,\mathrm{d}\gamma &= -\int c \,\mathrm{d}\gamma_1 - \int c \,\mathrm{d}\gamma_2 + \int c \,\mathrm{d}\sigma_1 + \int c \,\mathrm{d}\sigma_2 \\ &\leq \varepsilon(|x - y|^p + |x' - y'|^p - |x - y'|^p - |x' - y|^p + 4\delta) < 0 \end{aligned}$$

due to (1.20) and (1.19). This is in contradiction with the optimality of $\gamma$. Therefore, $\gamma$ is monotone.

To prove (ii), we fix a monotone transport plan $\gamma$ and show that

$$\gamma((-\infty, a] \times (-\infty, b]) = \min(\mu((-\infty, a]), \nu((-\infty, b])). \tag{1.21}$$

This implies that the mass of $\gamma$ on $(-\infty, a] \times (-\infty, b]$ is uniquely determined by $\mu$ and $\nu$, for any $(a, b) \in \mathbb{R}^2$. Since $\gamma$ is a transport plan, the mass of $\gamma$ on $(-\infty, a) \times (b, +\infty)$ and on $(a, +\infty) \times (-\infty, b]$ is also uniquely determined by $\mu$ and $\nu$. But these sets generate all Borel sets of $\mathbb{R}^2$, hence we conclude from (1.21) that there exists a unique monotone transport plan $\gamma$. There remains to prove (1.21). For this, we set $A = (-\infty, a) \times (b, +\infty)$ and $B = (a, +\infty) \times (-\infty, b)$, and we observe that since $\gamma$ is monotone, either $\gamma(A) = 0$ or $\gamma(B) = 0$. Therefore

$$\begin{aligned}\gamma((-\infty, a] \times (-\infty, b]) &= \min(\gamma((-\infty, a] \times (-\infty, b] \cup A), \gamma((-\infty, a] \times (-\infty, b] \cup B)) \\ &= \min(\mu((-\infty, a]), \nu((-\infty, b]))\end{aligned}$$

i.e., (1.21) holds. □

We finally conclude the proof of (1.17). It follows from Proposition 1.6 that the optimal transport maps $T_\mu$ and $T_\nu$ are non-decreasing, due to the fact that the corresponding optimal transport plans are concentrated on the graphs of $T_\mu$ and $T_\nu$. From this, it is immediate to check that $\gamma_{\text{opt}}$ given by (1.18) is also monotone. Using again Proposition 1.6, we get that $\gamma_{\text{opt}}$ is the unique optimal transport plan between $\mu$ and $\nu$.

Finally, what can be said about stability of *Kantorovich potentials* in 1 dimension? If $\rho$ satisfies the Poincaré inequality, i.e., if there exists $C > 0$ such that

$$\int_{\mathcal{X}} f \, \mathrm{d}\rho = 0 \Rightarrow \int_{\mathcal{X}} f^2 \, \mathrm{d}\rho \leq C \int_{\mathcal{X}} |\nabla f|^2 \, \mathrm{d}\rho,$$

then it follows from (1.17) (with $p = 2$) that $\|\phi_\mu - \phi_\nu\|_{L^2(\rho)} \leq CW_2(\mu, \nu)$. As we shall see in Section 3.7.3, in particular in Remark 3.12, this stability inequality for Kantorovich potentials is no longer guaranteed if $\rho$ does not satisfy the Poincaré inequality, even if the support of $\rho$ is an interval (in which case Kantorovich potentials are unique): there exists $\rho$ supported on an interval such that any inequality of the form $\|\phi_\mu - \phi_\nu\|_{L^2(\rho)} \leq CW_2(\mu, \nu)^\alpha$ for $\alpha \in (0, 1)$ fails.

## 1.4 Embeddings of the Wasserstein space and linearized optimal transport

The geometry of the Wasserstein space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ is known to be curved, and complicated. Using optimal transport maps, one may hope to embed the Wasserstein space into simpler spaces (e.g., Hilbert or Banach spaces), at the cost of introducing a small distortion which one needs to quantify. As we will see, this boils down to proving quantitative stability estimates for optimal transport maps, and actually this observation was one of the initial motivations for developing this theory.

**The Wasserstein space as a (pseudo) Riemannian manifold.** The Wasserstein space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ being a metric space, it is natural to look for its (constant-speed) geodesics, i.e., the curves $(\mu_t)_{t\in[0,1]}$ for which there exists $c > 0$ such that for any $0 \leq s, t \leq 1$,

$$W_2(\mu_s, \mu_t) = c|s - t|.$$

In $(\mathcal{P}_2(\mathbb{R}^d), W_2)$, any geodesic is of the following form (see for instance [96, Theorem 5.27]): there exists an optimal transport plan $\gamma$ for the quadratic cost between $\mu_0$ and $\mu_1$ such that $\mu_t = \pi_{t\#}\gamma$ for any $t \in [0, 1]$, where $\pi_t : (x, y) \mapsto (1 - t)x + ty$. Conversely, any curve of this form is a geodesic. In particular, if $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ is absolutely continuous and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, then there exists a unique geodesic between $\rho$ and $\mu$, and it is given by

$$t \mapsto ((1 - t)\mathrm{Id} + tT_\mu)_\# \, \rho \in \mathcal{P}_2(\mathbb{R}^d) \tag{1.22}$$

where $T_\mu$ denotes the optimal transport map from $\rho$ to $\mu$. Differentiating (1.22) with respect to $t$, we obtain that $T_\mu - \mathrm{Id}$ can be regarded as an element of the tangent space at $\rho$, namely the initial tangent vector of the Wasserstein geodesic from $\rho$ to $\mu$. This is part of a more general interpretation of $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ as a (pseudo) Riemannian manifold, a viewpoint initiated by Otto [87] and systematically investigated notably by Ambrosio, Gigli and Savaré [2].

**Embeddings of the Wasserstein space.** On this (pseudo) Riemannian manifold, the exponential map with base-point $\rho$ is nothing else than

$$\exp_\rho : \mathrm{Tan}_\rho \mathcal{P}_2(\mathbb{R}^d) \to \mathcal{P}_2(\mathbb{R}^d), \qquad T \mapsto T_\# \rho$$

and the inverse mapping

$$\log_\rho : \mathcal{P}_2(\mathbb{R}^d) \to L^2(\rho), \qquad \mu \mapsto T_\mu - \mathrm{Id}$$

is the analog of the Riemannian logarithm. The map $\log_\rho$ is injective, and it provides an embedding of the Wasserstein space into the Hilbert space $L^2(\rho)$. One may ask whether it is a "good" embedding, i.e., whether the geometry of the Wasserstein space is preserved, at least coarsely, under this embedding. Working with this embedding is equivalent to endow $\mathcal{P}_2(\mathbb{R}^d)$ with the "$\rho$-based" distance

$$W_{2,\rho}(\mu, \nu) = \|T_\mu - T_\nu\|_{L^2(\rho)}. \tag{1.23}$$

Then our question may be reformulated as: how do the distances $W_2$ and $W_{2,\rho}$ compare to each other?

If the stability inequality (1.13) holds for any $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, this means that $\mu \mapsto T_\mu$ is a bi-Hölder embedding from $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ to the Hilbert space $L^2(\rho)$ (using the reverse inequality (1.15)), and the two distances would be comparable: $W_2 \lesssim W_{2,\rho} \lesssim W_2^\alpha$. For instance, the previous section showed that for $d = 1$, $\mu \mapsto T_\mu$ is an *isometric* embedding, and $W_2 = W_{2,\rho}$. However, in dimension $d \geq 2$, it is known that Wasserstein spaces do not embed into large families of Banach spaces (including Hilbert spaces), even for much coarser notions of embedding. Therefore, we will aim at establishing (1.13) for strict subsets of $\mathcal{P}_2(\mathbb{R}^d)$, for instance for target probability measures $\mu, \nu$ supported in a fixed compact set, or with bounds on some moments. In other words, we embed only some *strict subset of the Wasserstein space* with a bi-Hölder embedding into $L^2(\mathbb{R}^d)$.

We will only state one result to illustrate the impossibility to embed Wasserstein spaces into Banach spaces. First, recall that a Banach space $Y$ is said to have type $p \in [1, 2]$ if there exists a constant $C < +\infty$ such that for every finite sequence $(y_i)_{i=1}^n \subset Y$,

$$\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i x_i\right\|^p \leq C^p \sum_{i=1}^n \|x_i\|^p,$$

where $(\varepsilon_i)_{i=1}^n$ are independent Rademacher random variables ($\mathbb{P}(\varepsilon_i = \pm 1) = 1/2$). Every Banach space has type 1 trivially. A space has type 2 if it behaves, in some sense, like a Hilbert space regarding averages of random sums; in particular, Hilbert spaces have type 2. The space $L^p$ has type $p$ for $1 \leq p < 2$, type 2 for $2 \leq p < +\infty$, and type 1 for $p = +\infty$.

To state the result, we also need the following definition:

**Definition 1.7** (Coarse embedding)**.** *Given two metric spaces $X$ and $Y$, a coarse embedding is a map $f : X \to Y$ such that there exist two non-decreasing functions $\rho_-, \rho_+ : [0, +\infty) \to [0, +\infty)$ with $\lim_{t\to+\infty} \rho_-(t) = +\infty$ and such that for all $x, y \in X$,*

$$\rho_-(d_X(x, y)) \leq d_Y(f(x), f(y)) \leq \rho_+(d_X(x, y)).$$

Any bi-Lipschitz or bi-Hölder embedding is coarse, but this notion is actually much weaker, since it allows for instance for non-continuous $f$. However, the following result, due to Andoni, Naor and Neiman in [5], and which we do not prove, shows the impossibility to embed coarsely Wasserstein spaces.

**Theorem 1.8.** *[5] For every $p > 1$ and $d \geq 3$, the space $(\mathcal{P}_p(\mathbb{R}^d), W_p)$ does not admit a coarse embedding into any Banach space of type $> 1$ (in particular, into Hilbert spaces and $L^q$ spaces, $1 < q < +\infty$).*

This result is not known in dimension $d = 2$, but other non-embeddability results valid also in dimension 2 have been proved, see [93] for an account on such results.

**Linearized optimal transport.** Due to the linear structure of the Hilbert space $L^2(\rho)$, the logarithm map $\mu \mapsto T_\mu$ is also used as a way to "linearize" optimal transport. For instance, to compute an "average" between two measures $\mu$ and $\nu$ in the Wasserstein space, one usually resorts to the notion of Wasserstein barycenter (or McCann interpolation), defined as a minimizer of

$$\inf_{\chi \in \mathcal{P}_2(\mathbb{R}^d)} (1-t) W_2(\mu, \chi)^2 + t W_2(\nu, \chi)^2 \tag{1.24}$$

for some $t \in [0, 1]$. Solving this optimization problem is often complicated. One may get another notion of average by first fixing an absolutely continuous $\rho \in \mathcal{P}_2(\mathbb{R}^d)$, then computing $T_\mu, T_\nu$ and their average $(1-t)T_\mu + tT_\nu)$, and finally considering

$$\widehat{\chi} = ((1-t)T_\mu + tT_\nu)_\# \rho. \tag{1.25}$$

Notice that $(1-t)T_\mu + tT_\nu$ is simply the weighted average of the initial tangent vectors giving rise to the geodesics from $\rho$ to $\mu$ and $\rho$ to $\nu$. The measure $\widehat{\chi}$, which is the endpoint of the geodesic with this tangent vector, is in general different from the solution to (1.24). It is actually located on the *generalized Wasserstein geodesic* (in the terminology of Ambrosio-Gigli-Savaré) between $\mu$ and $\nu$, defined as the curve

$$[0, 1] \ni t \mapsto ((1-t)T_\mu + tT_\nu)_\# \rho \in \mathcal{P}_2(\mathbb{R}^d).$$

In case $\mu = \rho$, the generalized geodesic between $\mu$ and $\nu$ coincides with the Wasserstein geodesic between $\mu$ and $\nu$. The advantage of (1.25) over (1.24) is that once $T_\mu$ and $T_\nu$ are known, $\widehat{\chi}$ can be computed quickly for any $t$ (at least, if it is possible to compute quickly the pushforward under an optimal transport map, a question addressed in [25]), whereas (1.24) requires to solve a new minimization problem every time $t$ is changed.

More generally, since it is often difficult to perform computations in Wasserstein spaces, which are curved (and infinite dimensional), it is a current practice in applications to first make computations in the Hilbert space $L^2(\rho)$, i.e., on the side of $T_\mu$, before pushing forward $\rho$ by the result of the computations in $L^2(\rho)$.

## 1.5 Stability around regular optimal transport maps

The earliest quantitative stability result for optimal transport maps, due to Gigli around 2010, addressed stability in the vicinity of a sufficiently regular map.

**Theorem 1.9** (Stability near regular optimal transport maps)**.** *Let $\rho$ be a probability measure on $\mathbb{R}^d$, absolutely continuous with respect to the Lebesgue measure, and with compact support. Let $\mathcal{Y} \subset \mathbb{R}^d$ be compact, and $K > 0$. Let $\mu, \nu \in \mathcal{P}(\mathcal{Y})$. If the optimal transport map $T_\mu$ from $\rho$ to $\mu$ is $K$-Lipschitz, then*

$$\|T_\mu - T_\nu\|_{L^2(\rho)} \leq C W_1(\mu, \nu)^{1/2}$$

*where $C = (2K \mathrm{diam}(\mathrm{supp}(\rho)))^{1/2}$.*

We provide a complete proof of Theorem 1.9, due to Chazal, Delalande and Mérigot, in Section 2.2. It is not the original proof of Gigli, it is closer in spirit to the other proofs presented in these notes.

**Comments on the regularity assumption in Theorem 1.9.** The important weakness of Theorem 1.9 is that the assumption that $T_\mu$ is $K$-Lipschitz is very strong. First, it implies that the support of $\mu$ is connected. Second, to prove that $T_\mu$ is Lipschitz one has to invoke the regularity theory for optimal transport maps, which requires strong assumptions on $\mu$ (stronger than having a connected support). The Lipschitz regularity of the optimal transport map, studied by many authors starting with Caffarelli, has been established only under restrictive assumptions: Caffarelli proved this property under the assumption that the source and target measure have bounded support, are bounded above and below by positive constants on their support, and that the support of the target is convex. Since this seminal result, some improvements and extensions have been obtained, but the spirit remains the same. And it is also known that continuity of the optimal transport map fails in some cases, even when the target has connected support: Caffarelli gave the example of a source measure $\rho$ supported on a two-dimensional domain $\mathcal{X}$ obtained by connecting two half disks by a thin corridor, and for which the optimal transport map is not continuous.

There is a whole line of research, notably in the statistical optimal transport community, working under this kind of Lipschitz regularity assumptions on $T_\mu$. Strong stability results (in terms of exponents) have been established: for instance, under the assumption that $T_\mu$ is bi-Lipschitz, Theorem 5.2 shows that $\|T_\mu - T_\nu\|_{L^2(\rho)} \lesssim W_2(\mu,\nu)$, where the hidden constant depends on the Lipschitz constants of $T_\mu$ and $T_\mu^{-1}$. We shall prove this result in Section 5.1.

## 1.6 Main results

The discussion of the previous paragraph motivates us to look for results in which much weaker assumptions are made on the measures, than those ensuring regularity of the optimal transport map. The results presented below state various assumptions on $\rho$ under which we are able to prove quantitative stability inequalities of the form (1.14)-(1.16), with nearly no assumption on the target measures $\mu$ and $\nu$. The discussion about the sharpness of these assumptions and about the resulting stability inequalities is postponed Section 3.7. In a nutshell, let us already mention that

*the results presented in this manuscript are (almost) sharp for Kantorovich potentials, whereas our understanding of the stability of optimal transport maps is still incomplete.*

The field is progressing fast. Our understanding so far is that stability of Kantorovich potentials is related to some Poincaré inequality on $\rho$, while stability of optimal transport maps should hold under weaker (but still mysterious) assumptions. If the Poincaré inequality holds for $\rho$, then

$$\|\phi_\mu - \phi_\nu\|_{L^2(\rho)} \leq C\|T_\mu - T_\nu\|_{L^2(\rho)}, \tag{1.26}$$

hence any stability inequality on optimal transport maps immediately implies a stability inequality for Kantorovich potentials! However, with our current knowledge, we are not able to prove stability inequalities on optimal transport maps directly, except under regularity assumptions as in Theorem 1.9. The proofs are indirect: first we prove the stability of Kantorovich potentials, then we deduce the stability of optimal transport maps thanks to some "reverse Poincaré inequality". We shall come back to this several times.

In the sequel, absolutely continuous measures (with respect to the Lebesgue measure in $\mathbb{R}^d$) are identified to their density. The first main result is the following:

**Theorem 1.10** (Quantitative stability – Log-concave source densities)**.** *Let $\rho = e^{-U-F}$ be a probability density on $\mathbb{R}^d$, with $F \in L^\infty(\mathbb{R}^d)$ and $D^2U \geq \kappa\mathrm{Id}$ for some $\kappa > 0$. Then for any* *compact set $\mathcal{Y}$, there exists $C > 0$ such that for any $\mu, \nu$ supported in $\mathcal{Y}$,*

$$\|\phi_\mu - \phi_\nu\|_{L^2(\rho)} \leq C W_1(\mu,\nu)^{1/2}(1 + |\log W_1(\mu,\nu)|^{1/2}). \tag{1.27}$$

*If moreover $D^2 U \leq \kappa' \mathrm{Id}$, then for any $\varepsilon > 0$ there exists $C > 0$ such that for any $\mu, \nu$ supported in $\mathcal{Y}$,*

$$\|T_\mu - T_\nu\|_{L^2(\rho)} \leq C W_1(\mu,\nu)^{\frac{1}{8}+\varepsilon}. \tag{1.28}$$

We show in Section 3.7.2 that the inequality (1.27) is sharp, up to the logarithmic loss. The additional assumption $D^2 U \leq \kappa' \mathrm{Id}$ made to prove (1.28) is probably only technical, but we have not been able to avoid it.

The second main result handles the case of source measures $\rho$ with bounded support. To state it, we introduce John domains, a vast family of domains introduced by Fritz John in the 1960s, containing in particular all bounded connected Lipschitz domains, but also some fractal domains like the Koch snowflake.

**Definition 1.11.** *A bounded open subset $\mathcal{X}$ of a metric space is called a John domain if there exist $x_0 \in \mathcal{X}$ and a constant $\eta > 0$ such that, for every $x \in \mathcal{X}$, there is $T > 0$ and a rectifiable curve $\gamma : [0,T] \to \mathcal{X}$ parametrized by the arclength (and whose length $T$ depends on $x$) such that $\gamma(0) = x$, $\gamma(T) = x_0$, and for any $t \in [0,T]$,*

$$\mathrm{dist}(\gamma(t), \mathcal{X}^c) \geq \eta t \tag{1.29}$$

*where $\mathcal{X}^c$ denotes the complement of $\mathcal{X}$.*

**Theorem 1.12** (Quantitative stability – Non-degenerate source densities on bounded domains)**.** *Let $\rho$ be a probability density on a John domain $\mathcal{X} \subset \mathbb{R}^d$, and assume that $\rho$ is bounded above and below on $\mathcal{X}$ by positive constants. Then for any compact set $\mathcal{Y} \subset \mathbb{R}^d$, there exists $C > 0$ such that for any $\mu, \nu \in \mathcal{P}(\mathcal{Y})$,*

$$\|\phi_\mu - \phi_\nu\|_{L^2(\rho)} \leq C W_1(\mu,\nu)^{1/2}. \tag{1.30}$$

*If moreover $\partial\mathcal{X}$ has a finite $(d-1)$-dimensional Hausdorff measure, then there exists $C > 0$ such that for any $\mu, \nu \in \mathcal{P}(\mathcal{Y})$,*

$$\|T_\mu - T_\nu\|_{L^2(\rho)} \leq C W_1(\mu,\nu)^{1/6}. \tag{1.31}$$

Theorem 1.12 also holds when $\mathbb{R}^d$ is replaced by an arbitrary smooth connected Riemannian manifold $M$, and optimal transport is considered with respect to the quadratic cost $c(x,y) = \frac{1}{2}\mathrm{dist}(x,y)^2$ where dist denotes the Riemannian distance on $M$. In case $M$ is compact without boundary (e.g., the sphere), then we may choose $\mathcal{X} = \mathcal{Y} = M$. We shall present in detail this generalization to Riemannian manifolds in Section 4.4.

## 1.7 Comments

Many comments can be made about Theorems 1.10 and 1.12. First, regarding the fact that the targets are assumed to be compactly supported in both results, we do not believe that this is a fundamental assumption. In [42], the assumption that was used is that target measures have $p$-th moment for some $p > d$ (for $p < d$, there exist unbounded Brenier potentials). We guess that our proof techniques may also cover this case, but shall not pursue this here.

We do not know whether the assumption that $\partial\mathcal{X}$ has finite $(d-1)$-dimensional Hausdorff measure in Theorem 1.12 is technical or not.

As we explain in Section 3.6, the strategy we use to prove Theorem 1.10, Theorem 1.12 and point (i) below allows us to recover the known fact that for any $\rho$ satisfying the assumptions of one of these results, the Poincaré inequality holds:

$$\int_{\mathcal{X}} \left( f - \int_{\mathcal{X}} f \mathrm{d}\rho \right)^2 \mathrm{d}\rho \leq C \int_{\mathcal{X}} |\nabla f|^2 \mathrm{d}\rho \tag{1.32}$$

(for $\rho$ as in Theorem 1.10, $\mathcal{X}$ has to be replaced by $\mathbb{R}^d$). The examples and counterexamples in Section 3.7 show an analogy, but not an equivalence, between the fact that the Poincaré inequality holds for $\rho$ and the fact that a stability inequality for Kantorovich potentials holds.

Finally, there are two important directions to improve and generalize Theorems 1.10 and 1.12:

→ proving stability inequalities for a wider range of probability densities $\rho$;

→ improving the stability exponents ($1/8 + \varepsilon$ in (1.28), $1/6$ in (1.31)) for the source measures $\rho$ considered in our main results.

Progress on the second direction is currently stalled, and new ideas are required to move it forward. Therefore, we comment only on the first direction. Indeed, our proof strategy is robust enough to handle other cases of interest. In all the following cases the methods presented in the next chapters are sufficient to prove stability inequalities for Kantorovich potentials and optimal transport maps (we do not discuss stability exponents here, they are all dimension-free except for (1.33)):

(i) Degenerate densities $\rho$ in bounded domains. The assumption in Theorem 1.12 that $\rho$ is bounded above and below on $\mathcal{X}$ is not always necessary. We illustrate this on two examples which we find particularly relevant in applications. The first example is given by source probability densities satisfying

$$c_1 \mathrm{dist}(x, \partial\mathcal{X})^\delta \leq \rho(x) \leq c_2 \mathrm{dist}(x, \partial\mathcal{X})^\delta$$

for some $\delta > -1$ and $c_1, c_2 > 0$, when $\mathcal{X}$ is a bounded Lipschitz domain. These densities blow-up ($\delta < 0$) or decay ($\delta > 0$) near $\partial\mathcal{X}$. The second example is the source probability density

$$\rho(x) = \frac{c_d}{|x|^{d-1}} \mathbf{1}_{B_d(0,1)\setminus\{0\}} \tag{1.33}$$

on $\mathbb{R}^d$, with $c_d$ a normalising constant and $\mathbf{1}_A$ the characteristic function of a set $A$. This probability density is sometimes called the spherical uniform distribution, and has been used in the literature to define multivariate quantiles.

(ii) Source measures $\rho$ on $\mathbb{R}^d$ which decay polynomially at infinity:

$$\rho(x) = f(x)(1 + |x|)^{-\beta} \tag{1.34}$$

with $f$ bounded above and below by positive constants uniformly over $x \in \mathbb{R}^d$, and $\beta > d+2$ so that $\rho$ has finite second moment. We shall see later that this family of source probability measures is interesting because one cannot use the same proof strategy as for the families of probability measures covered by Theorems 1.10 and 1.12, see Section 3.5.

(iii) Source measures with disconnected support. If we replace the beginning of the statement of Theorem 1.12 by "Let $\rho$ be a probability density on a finite union of John domains", then (1.31) still holds. Some modified version of (1.30) also holds, but one needs to be careful since Kantorovich potentials are not unique when the support of $\rho$ is not connected.

## 1.8 An elementary example

In this paragraph, we show on an example that one cannot hope in general to have an exponent strictly larger than 1/2 in (1.31). Let $\rho = \rho(x)\mathrm{d}x = \frac{1}{\pi}\mathbf{1}_{\mathbb{D}}(x)\mathrm{d}x$ be the uniform probability on the unit disk $\mathbb{D} \subset \mathbb{R}^2$. For $\theta \in \mathbb{R}/2\pi\mathbb{Z}$, we set $x_\theta = (\cos(\theta), \sin(\theta)) \in \mathbb{R}^2$ and define the probability measure

$$\mu_\theta = \frac{1}{2}(\delta_{x_\theta} + \delta_{x_{\theta+\pi}}). \tag{1.35}$$

The $\rho$-a.e. unique optimal transport map $T_{\mu_\theta}$ from $\rho$ to $\mu_\theta$ for the quadratic cost is explicit:

$$T_{\mu_\theta}(x) = \begin{cases} x_\theta & \text{if } \langle x, x_\theta \rangle \geq 0 \\ x_{\theta+\pi} & \text{if } \langle x, x_\theta \rangle < 0 \end{cases}$$

for $x \in \mathbb{D}$. In other words, each point $x \in \mathbb{D}$ is sent to the closest point among $x_\theta$ and $x_{\theta+\pi}$. This cuts the disk into two (equal) halves, see Figure 1.

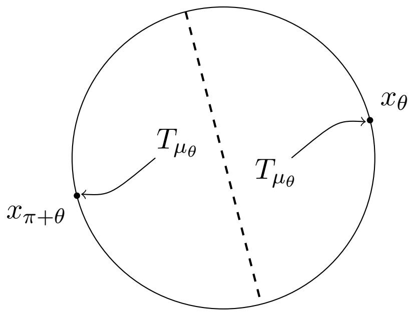



Figure 1: The optimal transport $T_{\mu_\theta}$ from $\rho$ to $\mu_\theta$.

Fix $\theta \in \mathbb{R}/2\pi\mathbb{Z}$, close to 0. Then, $\mathbb{D}$ may be written as $\mathbb{D} = A \sqcup (\mathbb{D} \setminus A)$ where

$$A = \{(r\cos(\theta), r\sin(\theta)) \mid 0 < r < 1,\ \theta \in (-\frac{\pi}{2} + \theta, \frac{\pi}{2} - \theta) \cup (\frac{\pi}{2} + \theta, -\frac{\pi}{2} - \theta)\}$$

is the set of points whose images under $T_{\mu_0}$ and $T_{\mu_\theta}$ are $x_0$, $x_\theta$, or $x_\pi$, $x_{\pi+\theta}$, i.e., the images are at angular distance $\theta$. Of course, $\mathbb{D} \setminus A$ is the set of points whose images under $T_{\mu_0}$ and $T_{\mu_\theta}$ are $x_0$, $x_{\pi+\theta}$, or $x_\pi$, $x_\theta$, i.e., far apart, at angular distance $\pi - \theta$. We find $\rho(A) = 1 - \frac{\theta}{\pi}$ and $\rho(\mathbb{D} \setminus A) = \frac{\theta}{\pi}$, hence as $\theta \to 0$,

$$\|T_{\mu_\theta} - T_{\mu_0}\|^2_{L^2(\rho)} = |2\sin(\theta/2)|^2 \rho(A) + |2\sin((\pi - \theta)/2)|^2 \rho(\mathbb{D} \setminus A) \underset{\theta \to 0}{\sim} \frac{4|\theta|}{\pi}. \tag{1.36}$$

On the other hand, for $\theta$ close enough to 0 and $p \geq 1$ arbitrary, the $W_p$ distance between $\mu_0$ and $\mu_\theta$ is obviously achieved by the map which sends $x_0$ to $x_\theta$ and $x_\pi$ to $x_{\theta+\pi}$. Its $p$-cost is

$$W_p(\mu_0, \mu_\theta) = |2\sin(\theta/2)| \underset{\theta \to 0}{\sim} |\theta|. \tag{1.37}$$

Putting together (1.36) and (1.37) for $p = 2$, we see that we cannot hope in this case to have an exponent strictly larger than 1/2 in (1.31).

In this example, it is also possible to compute the difference in $L^2$-norm between Kantorovich potentials. For this, we denote by $D_\theta \subset \mathbb{R}^2$ the line through the origin which is perpendicular to the segment $[x_\theta, x_{\theta+\pi}]$ (the dashed line on Figure 1) and observe that

$$\phi_{\mu_\theta}(x) = \mathrm{dist}(x, D_\theta) - C$$

for some constant $C$ independent of $\theta$ (simply equal to the integral of the function $x \mapsto \mathrm{dist}(x, D_\theta)$ on $\mathbb{D}$, which does not depend on $\theta$). Then

$$\|\phi_{\mu_\theta} - \phi_{\mu_0}\|^2_{L^2(\rho)} = \int_{\mathbb{D}} (|x_1 \cos(\theta) + x_2 \sin(\theta)| - |x_1|)^2 \mathrm{d}x = \theta^2 \int_{\mathbb{D}} x_2^2 \mathrm{d}x + O(\theta^3)$$

where $x = (x_1, x_2)$. Therefore, one cannot hope in this case to have a better exponent than 1 in (1.30).

The computations presented above can easily be generalized to any dimension and more general sources than the uniform probability on the disk. Actually, Gigli derived in [53, Theorem 5.1][2] more general conditions under which the maps $\mu \mapsto T_\mu$ is not better than $\frac{1}{2}$-Hölder.

## 1.9 Bibliographical notes

**§**1.1: There are many great books about optimal transport. We refer for instance to the two books by Villani [105] and [106], the one by Santambrogio for "applied mathematicians" [96], the book by Peyré-Cuturi about computational aspects of optimal transport [89], and the very recent book by Chewi-Niles Weed-Rigollet about statistical optimal transport [32]. To write the present text, we also took inspiration from the lecture notes by Quentin Mérigot at Institut Henri Poincaré, available on his webpage, and from the PhD thesis of Delalande [40]. Brenier presented his theorem in a short note [19] and gave details in an extended paper [20].

**§**1.2: The proof of Proposition 1.3 can be found in [106, Theorem 5.20]. Proposition 1.4 is a consequence of [20, Theorem 1.3] together with [106, Theorem 6.9]. The idea of the proof we provide was kindly communicated to us by Guillaume Carlier.

**§**1.3: For a more complete treatment of the one-dimensional case, see Chapter 2 in Santambrogio's book [96].

**§**1.4: Wasserstein geodesics are the main subject of the book by Ambrosio-Gigli-Savaré [2]. For a quick view on the subject, see [96, Chapter 5.4]. The interpretation of $W_2$ as a (pseudo) Riemannian manifold is due to Otto [87], who used it to study the long-time behavior of the porous medium equation. McCann introduced the concept of displacement interpolation in [77].

The paper [107] introduced the linearized optimal transport distance $W_{2,\rho}$ defined in (1.23) and used it for pattern recognition in images. Since then, this distance has been used for instance to perform super-resolution of highly corrupted images [67] and to detect and visualize phenotypic differences between classes of cells [9].

Wasserstein barycenters have been introduced in [1], generalizing the concept of displacement interpolation of McCann. This notion of barycenter has found many successful applications, for instance in image processing [94], geometry processing [101], statistics [100] or machine learning [39]. Section 5.2 of this manuscript contains a brief introduction to the topic, see the book chapters [89, Chapter 9.2], [32, Chapter 8] for more detailed treatments.

§1.5: Theorem 1.9 was first proven by [53]. The statement and the proof we give are borrowed from [79]. In [73], these ideas were used to prove error estimates for semi-discrete and fully-discrete algorithms to compute optimal transport plans and maps. The regularity theory of the Monge-Ampère equation and its link to regularity of optimal transport maps is explained in the survey [44]. The counterexample to the continuity of the optimal transport map is due to Caffarelli, see [23].

[2]be careful, a preliminary version of the paper, available online, contained a mistake, which has been corrected in the published version.

§1.6: Motivated by *semi-discrete* optimal transport, in which the target measure is discrete (or discretized) but not the source measure, Berman [11] was the first to obtain quantitative stability estimates without assuming regularity of the optimal transport map. He derived dimension-dependent stability exponents for $\rho$ bounded above and below on a compact, convex domain, using complex geometry. Delalande and Mérigot [42] then improved his stability exponent, making it dimension-free under the same assumptions on $\rho$. But more importantly, they introduced a robust proof technique based on the study of the Kantorovich functional, see Chapter 2.

John domains were named in honor of Fritz John who introduced them in his work on elasticity [60]; Martio and Sarvas [76] introduced this terminology. They appear also in the theory of quasi-conformal mappings and in geometric measure theory.

§1.8: The example in this section is due to [79].

# 2 The Kantorovich functional

In this chapter, we begin by carefully introducing the Kantorovich relaxation of the Monge problem, together with its dual formulation in terms of Kantorovich potentials. These are very classical results, presented with great precision in several textbooks (see the references in Section 2.7). Rather than stating them in their most general and technical form, we choose to present the results and the underlying ideas at a more intuitive level. This naturally leads us to the definition and study of the so-called *Kantorovich functional.* We then establish the Brascamp–Lieb inequality, which allows us to prove its strong convexity, and from this we deduce the stability of Kantorovich potentials when $\rho$ is log-concave.

## 2.1 The dual formulation of optimal transport

Monge formulated in 1781 the optimal transport problem as

$$\inf_{\substack{S:\mathcal{X}\to\mathcal{Y}\\ S_\#\rho=\mu}} \int_{\mathcal{X}} |x - S(x)| \mathrm{d}\rho(x) \tag{2.1}$$

where $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^d$, $\rho$ is a probability measure on $\mathcal{X}$, and $\mu$ a probability measure on $\mathcal{Y}$. For many reasons already explained, it is natural to put a square on the $|x - S(x)|$ term, thus yielding (1.4). It is only in 1942 that Kantorovich introduced what is now known as the Kantorovich relaxation already mentioned in (1.3) and which we recall:

$$\inf_{\gamma\in\Pi(\rho,\mu)} \int_{\mathcal{X}\times\mathcal{Y}} |x - y|^2 \mathrm{d}\gamma(x, y). \tag{2.2}$$

We show in this paragraph how to solve (2.2), and introduce along the way the primal and dual Kantorovich potentials, which play a prominent role in these notes.

First we notice that for $\gamma \in \Pi(\rho, \mu)$,

$$\int_{\mathcal{X}\times\mathcal{Y}} |x - y|^2 \mathrm{d}\gamma(x, y) = \int_{\mathcal{X}} |x|^2 \mathrm{d}\rho(x) + \int_{\mathcal{Y}} |y|^2 \mathrm{d}\mu(y) - 2 \int_{\mathcal{X}\times\mathcal{Y}} \langle x, y\rangle \ \mathrm{d}\gamma(x, y).$$

Since the first two terms in the right-hand side do not depend on $\gamma$, the quadratic optimal transport problem (1.3) is equivalent to

$$\sup_{\gamma\in\Pi(\rho,\mu)} \int_{\mathcal{X}\times\mathcal{Y}} \langle x, y\rangle \ \mathrm{d}\gamma(x, y). \tag{2.3}$$

We denote this supremum by $\mathcal{I}(\rho, \mu)$.

### 2.1.1 The dual problem

It is possible to remove the constraint $\gamma \in \Pi(\rho, \mu)$ in (2.3) by introducing appropriate Lagrange multipliers. This leads to what is called the *dual problem*, which turns out to be equivalent to the *primal problem* (2.3). To derive the dual problem, it is not a problem to work at a formal level; once derived formally however one needs to show its equivalence with the primal problem.

The formal derivation of the dual problem goes as follows. For any measure $\gamma$ on $\mathcal{X} \times \mathcal{Y}$,

$$\inf_{\substack{\phi\in C^0(\mathcal{X})\\ \psi\in C^0(\mathcal{Y})}} \int_{\mathcal{X}} \phi(x) \ \mathrm{d}\rho(x) + \int_{\mathcal{Y}} \psi(y) \ \mathrm{d}\mu(y) - \int_{\mathcal{X}\times\mathcal{Y}} \phi \oplus \psi \ \mathrm{d}\gamma(x, y) = \begin{cases} 0 & \text{if } \gamma \in \Pi(\rho, \mu) \\ -\infty & \text{otherwise} \end{cases}$$

where $\phi \oplus \psi : (x, y) \mapsto \phi(x) + \psi(y)$. This may be seen as follows: if for instance the first marginal of $\gamma$ is not equal to $\rho$, we consider a set $A \neq \emptyset$ such that $\rho(A) \neq \gamma(A \times \mathcal{Y})$, take for $\phi$ a continuous approximation of $\lambda \mathbf{1}_A$ and $\psi \equiv 0$, and let $\lambda$ go to $\pm\infty$. This proves that (2.3) is equivalent to

$$\sup_{\gamma} \inf_{\substack{\phi \in C^0(\mathcal{X}) \\ \psi \in C^0(\mathcal{Y})}} \int_{\mathcal{X} \times \mathcal{Y}} (\langle x, y \rangle - \phi \oplus \psi) \, \mathrm{d}\gamma(x, y) + \int_{\mathcal{X}} \phi(x) \, \mathrm{d}\rho(x) + \int_{\mathcal{Y}} \psi(y) \, \mathrm{d}\mu(y)$$

where the supremum is taken over all measures $\gamma$, not only those in $\Pi(\rho, \mu)$. The duality principle consists in exchanging the sup and the inf; we get

$$\inf_{\substack{\phi \in C^0(\mathcal{X}) \\ \psi \in C^0(\mathcal{Y})}} \left[ \int_{\mathcal{X}} \phi(x) \, \mathrm{d}\rho(x) + \int_{\mathcal{Y}} \psi(y) \, \mathrm{d}\mu(y) + \sup_{\gamma} \int_{\mathcal{X} \times \mathcal{Y}} (\langle x, y \rangle - \phi \oplus \psi) \, \mathrm{d}\gamma(x, y) \right] \tag{2.4}$$

(after rearranging the terms). As we will see, this new optimization problem is *equivalent* to (2.3), which may seem surprising at first sight. But let us first simplify (2.4) a bit. The supremum inside the brackets can itself be seen as a constraint: it is equal to 0 if $\phi \oplus \psi \geq \langle x, y \rangle$ for any $x \in \mathcal{X}, y \in \mathcal{Y}$, and equal to $+\infty$ otherwise. Therefore we end-up with the dual problem

$$\inf \left\{ \int_{\mathcal{X}} \phi(x) \, \mathrm{d}\rho(x) + \int_{\mathcal{Y}} \psi(y) \, \mathrm{d}\mu(y) \mid \phi \in C^0(\mathcal{X}), \psi \in C^0(\mathcal{Y}), \phi \oplus \psi \geq \langle x, y \rangle \right\}. \tag{2.5}$$

We denote this infimum by $\mathcal{J}(\rho, \mu)$.

It is immediate to check that $\mathcal{J}(\rho, \mu) \geq \mathcal{I}(\rho, \mu)$. Indeed, for any $\phi, \psi$ such that $\phi \oplus \psi \geq \langle x, y \rangle$, and any $\gamma \in \Pi(\rho, \mu)$ we have

$$\int_{\mathcal{X} \times \mathcal{Y}} \langle x, y \rangle \, \mathrm{d}\gamma(x, y) \leq \int_{\mathcal{X} \times \mathcal{Y}} \phi \oplus \psi \, \mathrm{d}\gamma(x, y) = \int_{\mathcal{X}} \phi(x) \, \mathrm{d}\rho(x) + \int_{\mathcal{Y}} \psi(y) \, \mathrm{d}\mu(y).$$

What is more surprising is that the reverse inequality $\mathcal{J}(\rho, \mu) \leq \mathcal{I}(\rho, \mu)$ is also true. The resulting equality

$$\mathcal{I}(\rho, \mu) = \mathcal{J}(\rho, \mu) \tag{2.6}$$

is often referred to as the *Kantorovich duality*, or *strong duality*. It holds in great generality, in Polish spaces and for any lower semi-continuous cost. This is a very classical result, but we shall not cover it here. We refer instead for instance to [105, Chapter 1].

The problem (2.3) admits a solution (i.e., a maximizer) for any $\rho, \mu \in \mathcal{P}_2(\mathbb{R}^d)$ (see [106, Theorem 5.10]). Similarly, the problem (2.5) admits a minimizer for any $\rho, \mu \in \mathcal{P}_2(\mathbb{R}^d)$, and a pair $(\phi, \psi)$ which minimizes (2.5) is called a pair of Kantorovich potentials. But for both (2.3) and (2.5), the optimizers are not necessarily unique.

### 2.1.2 Support of optimizers

Let us prove that if $\gamma$ is a maximizer in (2.3) and $(\phi, \psi)$ is a minimizer in (2.5), then

$$\operatorname{supp}(\gamma) \subset \{(x, y) \in \mathcal{X} \times \mathcal{Y} \mid \phi(x) + \psi(y) = \langle x, y \rangle\}. \tag{2.7}$$

For this we use (2.6) which yields

$$\int_{\mathcal{X} \times \mathcal{Y}} ((\phi \oplus \psi)(x, y) - \langle x, y \rangle) \, \mathrm{d}\gamma(x, y) = 0.$$

Since $\phi \oplus \psi \geq \langle x, y \rangle$, this gives (2.7). The converse is true: if (2.7) holds for some $\gamma \in \Pi(\rho, \mu)$ and $(\phi, \psi)$ such that $\phi \oplus \psi \geq \langle x, y \rangle$, then $\gamma$ and $(\phi, \psi)$ are solutions of their respective optimization problems, as a consequence of the definitions of $\mathcal{I}(\rho, \mu)$, $\mathcal{J}(\rho, \mu)$, and of the equality (2.6).

### 2.1.3 Semi-dual formulation

It is possible to give an equivalent unconstrained formulation of (2.5). Recall the definition of the *Legendre transform*:

$$\phi^*(y) = \sup_{x\in\mathbb{R}^d} \langle y, x\rangle - \phi(x). \tag{2.8}$$

If a function $\phi$ is defined only over a subset of $\mathbb{R}^d$, we first extend $\phi$ by $+\infty$ outside this subset to define its Legendre transform. As a consequence, $\phi^*(y) = \sup_{x\in\mathcal{X}} \langle y, x\rangle - \phi(x)$.

For a given $\phi$, the smallest possible $\psi$ that one may choose to satisfy the constraint $\phi \oplus \psi \geq \langle x, y\rangle$ in (2.5) is $\psi = \phi^*$. Similarly, for a given $\psi$, the smallest possible $\phi$ that one may choose is $\phi = \psi^*$. Therefore, one has

$$\mathcal{J}(\rho,\mu) = \inf\left\{\int_{\mathcal{X}} \phi \, \mathrm{d}\rho + \int_{\mathcal{Y}} \phi^* \, \mathrm{d}\mu \mid \phi \in C^0(\mathcal{X})\right\} = \inf\left\{\int_{\mathcal{X}} \psi^* \, \mathrm{d}\rho + \int_{\mathcal{Y}} \psi \, \mathrm{d}\mu \mid \psi \in C^0(\mathcal{Y})\right\}.$$

This leads us to the *semi-dual formulation* of (2.5):

$$\inf_{\psi\in C^0(\mathcal{Y})} \int_{\mathcal{X}} \psi^* \, \mathrm{d}\rho + \int_{\mathcal{Y}} \psi \, \mathrm{d}\mu. \tag{2.9}$$

If we want to solve (2.9), it seems natural to write the first and second-order optimality conditions with respect to $\psi$. This will be done in Section 2.3: the second integral is linear in $\psi$, hence easy to differentiate, but the first part is non-linear in $\psi$.

### 2.1.4 Convex functions and proof of Brenier's theorem

For any $\phi, \psi$ such that $\phi \oplus \psi \geq \langle x, y\rangle$, we have $\psi \geq \phi^*$ and $\phi \geq \phi^{**}$, hence

$$\int_{\mathcal{X}} \phi \, \mathrm{d}\rho + \int_{\mathcal{Y}} \psi \, \mathrm{d}\mu \geq \int_{\mathcal{X}} \phi^{**} \, \mathrm{d}\rho + \int_{\mathcal{Y}} \phi^* \, \mathrm{d}\mu.$$

Recall that a convex function is called proper if it has a non-empty domain, it never takes on the value $-\infty$ and also it is not identically equal to $+\infty$. The Fenchel-Moreau theorem ensures that if $\phi$ is a proper lower semi-continuous convex function, then $\phi^{**} = \phi$. Therefore, in (2.5) we may restrict the infimum to the set of pairs $(\phi, \phi^*)$ of proper lower semi-continuous conjugate functions on $\mathbb{R}^d$.

We recall that if $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is a convex function, then its subdifferential at $x \in \mathbb{R}^d$ is defined as

$$\partial\phi(x) = \{v \in \mathbb{R}^d \mid \forall z \in \mathbb{R}^d,\ \phi(z) \geq \phi(x) + \langle z - x, v\rangle\}.$$

The graph of the subdifferential is

$$\partial\phi = \bigcup_{x\in\mathbb{R}^d} \{x\} \times \partial\phi(x).$$

If $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is a proper lower semi-continuous convex function, then for all $x, y \in \mathbb{R}^d$,

$$\phi(x) + \phi^*(y) = \langle x, y\rangle \Leftrightarrow y \in \partial\phi(x) \Leftrightarrow x \in \partial\phi^*(y). \tag{2.10}$$

Indeed,

$$\begin{aligned}
\phi(x) + \phi^*(y) = \langle x, y\rangle &\Leftrightarrow \langle x, y\rangle \geq \phi(x) + \phi^*(y) \\
&\Leftrightarrow \forall z \in \mathbb{R}^d, \quad \langle x, y\rangle \geq \phi(x) + \langle y, z\rangle - \phi(z) \\
&\Leftrightarrow \forall z \in \mathbb{R}^d, \quad \phi(z) \geq \phi(x) + \langle y, z - x\rangle \\
&\Leftrightarrow y \in \partial\phi(x).
\end{aligned}$$

By symmetry, the other equivalence follows, since $\phi^{**} = \phi$ by the Fenchel-Moreau theorem (which uses the assumptions on $\phi$).

Combining (2.10) and (2.7), we have obtained:

**Proposition 2.1.** *Any optimal transport plan (i.e., any solution of* (2.3)*) has its support contained in the graph of the subdifferential of a convex function.*

The proof of Brenier's Theorem 1.2 is now straightforward.

*Proof of Brenier's theorem.* Let $\gamma$ be an optimal transport plan, and take $\psi$ a proper lower semi-continuous function on $\mathbb{R}^d$ that solves (2.9). For any $(x_0, y_0)$ in the support of $\gamma$ such that $\psi^*$ is differentiable at $x_0$, one has $y_0 = \nabla\psi^*(x_0)$ according to Proposition 2.1. Since $\psi^*$ is a proper lower semi-continuous convex function on $\mathbb{R}^d$, it is differentiable $\rho$-almost everywhere. This implies that $\gamma = (\mathrm{Id}, \nabla\psi^*)_\#\rho$. But $\psi$ and $\gamma$ have been chosen independently, so for any other optimal transport plan $\widetilde{\gamma}$ there also holds $\widetilde{\gamma} = (\mathrm{Id}, \nabla\psi^*)_\#\rho$. In other words, there is a unique optimal transport plan $\gamma$. If $f$ is another proper lower semi-continuous convex function such that $\mu = (\nabla f)_\#\rho$, then $\widetilde{\gamma} = (\mathrm{Id}, \nabla f)$ satisfies

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{R}^d} \langle x, y\rangle \, \mathrm{d}\gamma(x,y) \le \int_{\mathbb{R}^d} f(x)\, \mathrm{d}\rho(x) + \int_{\mathbb{R}^d} f^*(y)\, \mathrm{d}\mu(y) &= \int_{\mathbb{R}^d} (f(x) + f^*(\nabla f(x)))\, \mathrm{d}\rho(x) \\ &= \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle x, y\rangle \, \mathrm{d}\tilde{\gamma}(x,y)\end{aligned}$$

where the last equality comes from the Fenchel-Young equality case (2.10). Since $\gamma$ is the unique maximizer of (2.3), we get $\widetilde{\gamma} = \gamma$, and thus $\nabla f = \nabla\psi^*$ $\rho$-a.e. □

### 2.1.5 Kantorovich-Rubinstein formula

We conclude this section with the following important formula, which we will use several times.

**Theorem 2.2** (Kantorovich-Rubinstein duality formula)**.** *For any $\mu, \nu$ probability measures on a compact set $\mathcal{Y} \subset \mathbb{R}^d$,*

$$W_1(\mu,\nu) = \sup\left\{\int_{\mathcal{Y}} f \mathrm{d}\mu - \int_{\mathcal{Y}} f \mathrm{d}\nu \mid \mathrm{Lip}(f) \le 1\right\}. \tag{2.11}$$

*More generally, for any $\mu, \nu$ probability measures on a Polish space $\mathcal{Y}$ (i.e., a complete, separable, metric space, with distance denoted by* $\mathrm{dist}(\cdot,\cdot)$*),*

$$\inf_{\gamma\in\Pi(\mu,\nu)} \int_{\mathcal{Y}\times\mathcal{Y}} \mathrm{dist}(x,y)\mathrm{d}\gamma(x,y) = \sup\left\{\int_{\mathcal{Y}} f \mathrm{d}\mu - \int_{\mathcal{Y}} f \mathrm{d}\nu \mid \mathrm{Lip}(f) \le 1\right\}. \tag{2.12}$$

In these notes, we will only need the inequality $\ge$, which is easy to prove. For any $\gamma \in \Pi(\mu,\nu)$ and any 1-Lipschitz function $f$,

$$\int_{\mathcal{Y}} f \, \mathrm{d}\mu - \int_{\mathcal{Y}} f \, \mathrm{d}\nu = \int_{\mathcal{Y}\times\mathcal{Y}} (f(x) - f(y))\, \mathrm{d}\gamma(x,y) \le \int_{\mathcal{Y}\times\mathcal{Y}} \mathrm{dist}(x,y)\, \mathrm{d}\gamma(x,y).$$

Taking the infimum over $\gamma \in \Pi(\mu,\nu)$ in the right-hand side, and the supremum over 1-Lipschitz functions $f$ in the left-hand side, we get that in (2.11) the left-hand side is larger than the right-hand side. For the converse inequality, which is another instance of Kantorovich duality, we provide references at the end of this chapter. The version of the Kantorovich-Rubinstein duality formula in Polish spaces will be useful for us when dealing with Wasserstein barycenters, in Section 5.2.

## 2.2 Proof of Theorem 1.9

With the Kantorovich potentials at hand, we will be able to give in this section a short proof of Gigli's Theorem 1.9, which we recall here for convenience.

**Theorem** (Stability near regular optimal transport maps)**.** *Let $\rho$ be a probability measure on $\mathbb{R}^d$, absolutely continuous with respect to the Lebesgue measure, and with compact support. Let $\mathcal{Y} \subset \mathbb{R}^d$ be compact, and $K > 0$. Let $\mu, \nu \in \mathcal{P}(\mathcal{Y})$. If the optimal transport map $T_\mu$ from $\rho$ to $\mu$ is $K$-Lipschitz, then*

$$\|T_\mu - T_\nu\|_{L^2(\rho)} \leq C W_1(\mu, \nu)^{1/2}$$

*where $C = (2K \mathrm{diam}(\mathrm{supp}(\rho)))^{1/2}$.*

We start with a classical result which turns the Lipschitzness assumption on $T_\mu$ into strong convexity of $\psi_\mu$. For this we need the following definitions.

**Definition 2.3.** *For $K > 0$, a function $\phi : \mathbb{R}^d \to \mathbb{R}$ is $K$-smooth if $\frac{K}{2}|x|^2 - \phi(x)$ is convex, i.e.,*

$$\phi((1-t)x_0 + tx_1) + \frac{Kt(1-t)}{2}|x_0 - x_1|^2 \geq (1-t)\phi(x_0) + t\phi(x_1),$$

*for any $x_0, x_1 \in \mathbb{R}^d$ and $t \in [0,1]$. For $\lambda > 0$, a function $\psi : \mathbb{R}^d \to \mathbb{R}$ is $\lambda$-strongly convex if $\psi(y) - \frac{\lambda}{2}|y|^2$ is convex, i.e., for any $y_0, y_1 \in \mathbb{R}^d$ and $t \in [0,1]$,*

$$\psi((1-t)y_0 + ty_1) + \frac{\lambda t(1-t)}{2}|y_0 - y_1|^2 \leq (1-t)\psi(y_0) + t\psi(y_1). \tag{2.13}$$

When $\phi, \psi \in C^2$, the first condition is simply $D^2\phi \leq K$, and the second condition is $D^2\psi \geq \lambda$. Now, if $\phi$ and $\phi^*$ are $C^2$ convex functions, differentiating the identity $\nabla\phi(\nabla\phi^*) = \mathrm{Id}$ (which comes from (2.10)) we get $(\nabla^2\phi^*)^{-1} = \nabla^2\phi(\nabla\phi^*)$, and thus $\|\nabla^2\phi\|_{L^\infty} \leq K$ if and only if $\nabla^2\phi^* \geq \frac{1}{K}$, i.e., $\phi$ is $K$-smooth if and only if $\phi^*$ is $K^{-1}$-strongly convex. This result is actually true without assuming that $\phi, \phi^*$ are $C^2$ (see [73, Lemma 2.2] for a proof):

**Lemma 2.4.** *Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Then $\phi$ is $K$-smooth if and only if $\phi^*$ is $\lambda$-strongly convex for $\lambda = K^{-1}$.*

In the sequel, if $f \in C^0(\mathcal{Y})$ and $\mu$ is a real-valued Radon measure on $\mathcal{Y}$, then we set

$$\langle f \mid \mu \rangle = \int_{\mathcal{Y}} f \mathrm{d}\mu. \tag{2.14}$$

The (distinct) notation $\langle \cdot, \cdot \rangle$ denotes the Euclidean scalar product in $\mathbb{R}^d$.

The proof of Theorem 1.9 mainly relies on the following inequality:

**Lemma 2.5.** *Under the assumptions of Theorem 1.9, there holds*

$$\langle \psi_\mu - \psi_\nu \mid \nu - \mu \rangle \geq \frac{1}{2K}\|T_\nu - T_\mu\|^2_{L^2(\rho)} \tag{2.15}$$

*where $(\phi_\mu, \psi_\mu)$ (resp. $(\phi_\nu, \psi_\nu)$) is a pair of Kantorovich potentials associated to the transport from $\rho$ to $\mu$ (resp. $\rho$ to $\nu$).*

*Proof of Lemma 2.5.* Since $\mu = (T_\mu)_\#\rho$ and $\nu = (T_\nu)_\#\rho$ we have

$$\langle \psi_\mu \mid \nu - \mu \rangle = \int_{\mathbb{R}^d} (\psi_\mu(T_\nu(x)) - \psi_\mu(T_\mu(x)))\mathrm{d}\rho(x).$$

Since $T_\mu$ is $K$-Lipschitz, $\phi_\mu$ is $K$-smooth and therefore $\psi_\mu = \phi_\mu^*$ is $K^{-1}$-strongly convex by Lemma 2.4. Hence, letting $t \to 0$ in (2.13), we get that

$$\psi_\mu(y) - \psi_\mu(z) \geq \langle y - z, v\rangle + \frac{1}{2K}|y - z|^2$$

for any $v \in \partial\psi_\mu(z)$. Now we fix $x$ in the support of $\rho$ and choose $y = T_\nu(x)$ and $z = T_\mu(x)$. Therefore $z \in \partial\phi_\mu(x)$ (see Proposition 2.1) and (2.10) yields $x \in \partial\psi_\mu(z)$. We deduce

$$\langle \psi_\mu \mid \nu - \mu \rangle \geq \int_{\mathbb{R}^d} \Big(\langle T_\nu(x) - T_\mu(x), x\rangle + \frac{1}{2K}|T_\nu(x) - T_\mu(x)|^2\Big)\mathrm{d}\rho(x).$$

Since $\psi_\nu$ is also convex (but not necessarily strongly convex), choosing $y = T_\mu(x)$ and $z = T_\nu(x)$ we obtain similarly

$$\langle \psi_\nu \mid \mu - \nu \rangle \geq \int_{\mathbb{R}^d} \langle T_\mu(x) - T_\nu(x), x\rangle \; \mathrm{d}\rho(x).$$

Adding the two previous inequalities we get (2.15). □

*End of the proof of Theorem 1.9.* Using the Kantorovich-Rubinstein duality formula (2.11), we get that the left-hand side of (2.15) is bounded above by $\mathrm{Lip}(\psi_\mu - \psi_\nu)W_1(\mu, \nu)$. Finally, to conclude the proof, it remains to observe that $\psi_\mu - \psi_\nu$ is $\mathrm{diam}(\mathcal{X})$-Lipschitz. Essentially this is due to the fact that $\partial\psi_\mu$ and $\partial\psi_\nu$ are subsets of the support of $\rho$. Here is a formal proof: if $y, y' \in \mathcal{Y}$, let $x, x'$ in the support of $\rho$ such that $\psi_\mu(y) = \langle y, x\rangle - \phi_\mu(x)$ and $\psi_\nu(y') = \langle y', x'\rangle - \phi_\nu(x')$. These points exist due to (2.7) and (2.10). Then

$$\begin{aligned}(\psi_\mu - \psi_\nu)(y) - (\psi_\mu - \psi_\nu)(y') &= \psi_\mu(y) - \psi_\mu(y') + \psi_\nu(y') - \psi_\nu(y)\\ &\leq \langle y - y', x\rangle + \langle y' - y, x'\rangle \leq |y - y'|\mathrm{diam}(\mathcal{X}).\end{aligned}$$

Exchanging the roles of $y$ and $y'$ we get the Lipschitz bound and Theorem 1.9 follows. □

### 2.3 The Kantorovich functional: definition and derivatives

The *Kantorovich functional* is defined as

$$\mathcal{K}_\rho : \psi \mapsto \int_{\mathcal{X}} \psi^* \mathrm{d}\rho \tag{2.16}$$

for $\psi \in C^0(\mathcal{Y})$. This functional is one part of the quantity appearing in the semidual formulation (2.9). We prove here (some kind of) strong convexity of this functional under some assumptions on $\rho$, and explain how it implies stability properties for Kantorovich potentials. The other term $\int \psi \mathrm{d}\mu$ which appears in (2.9) is linear in $\psi$ and thus does not affect the convexity properties of $\mathcal{K}_\rho$. Also, it is immediate to see that $\mathcal{K}_\rho$ is *convex*, since it is a convex combination of the convex functions $\psi \mapsto \psi^*(x)$.

Let us explain on a basic example how one can deduce stability from strong convexity. Let $f : \mathbb{R}^d \to \mathbb{R}$ with $D^2 f \geq c\mathrm{Id}$ for some $c > 0$, where $D^2 f$ denotes the Hessian of $f$. Then for any $x_1, x_2 \in \mathbb{R}^d$,

$$c\|x_1 - x_2\|^2 \leq \langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2\rangle \leq |x_1 - x_2||\nabla f(x_1) - \nabla f(x_2)| \tag{2.17}$$

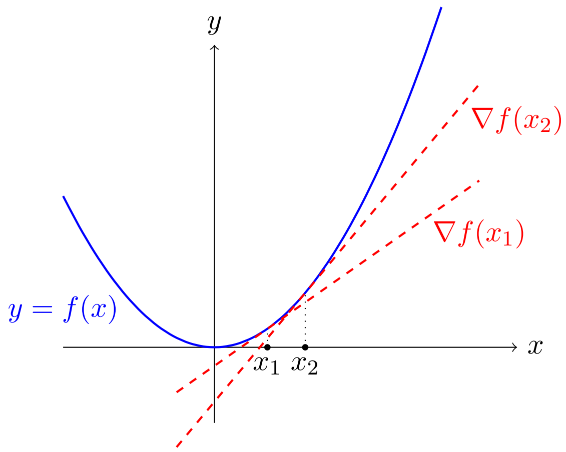



Figure 2: For a strongly convex function $f$, if $\nabla f(x_1)$ and $\nabla f(x_2)$ are close to each other, then $x_1$ and $x_2$ are close to each other.

hence $|x_1 - x_2| \leq c^{-1}|\nabla f(x_1) - \nabla f(x_2)|$. In particular, if $\nabla f(x_1)$ and $\nabla f(x_2)$ are close to each other, then $x_1$ and $x_2$ are close to each other. In other words, strong convexity of $f$ implies that the map $\nabla f(x) \mapsto x$ is well-defined and stable.

To prove our main results, we shall develop an analogous computation for $f = \mathcal{K}_\rho$ the Kantorovich functional, defined on $C^0(\mathcal{Y})$ instead of $\mathbb{R}^d$. We will see that if $\mathcal{K}_\rho$ is taken as (2.16) below, then for $\psi \in C^0(\mathcal{Y})$ the gradient $\nabla \mathcal{K}_\rho(\psi)$ is a measure, and

$$\nabla \mathcal{K}_\rho(\psi) = -(\nabla \psi^*)_\# \rho. \tag{2.18}$$

In particular, $\nabla \mathcal{K}_\rho(\psi_\mu) = -\mu$ for $\psi_\mu$ the dual Kantorovich potential from $\rho$ to $\mu$. The above computations show (heuristically) that if one is able to prove that $\mathcal{K}_\rho$ is strongly convex in some sense, then one gets a stability inequality of the form

$$\mu, \nu \text{ close to each other} \Rightarrow \psi_\mu, \psi_\nu \text{ close to each other.}$$

If we consider again Lemma 2.5 under this new light, we see that it plays the role of the left-hand side inequality in (2.17). And the Kantorovich-Rubinstein duality formula replaces the Cauchy-Schwarz inequality used in (2.17), and yields an upper bound on

$$\langle \psi_\mu - \psi_\nu \mid \nu - \mu \rangle$$

since $\psi_\mu$ and $\psi_\nu$ are Lipschitz by the following observation, which we will use many times:

**Lemma 2.6.** *If $\mathcal{X} \subset \mathbb{R}^d$ is a compact set and $f \in C^0(\mathcal{X})$, then the Legendre transform $f^*$ is Lipschitz, with Lipschitz constant at most $\sup_{x \in \mathcal{X}} |x|$.*

*Proof of Lemma 2.6.* Let $y, y' \in \mathbb{R}^d$. Let $x$ such that $f^*(y) = \langle x, y \rangle - f(x)$. Then

$$f^*(y) - f^*(y') = \langle x, y \rangle - f(x) - f^*(y') \leq \langle x, y - y' \rangle \leq |x||y - y'|.$$

Reversing the roles of $y$ and $y'$, we get the result. □

To study the (strong) convexity of $\mathcal{K}_\rho$, we compute its first two derivatives. The equality (2.19) below is a formal writing of (2.18).

**Lemma 2.7.** *Let $\phi_0, \phi_1 \in C^2(\mathbb{R}^d)$ be strongly convex functions. Define $\psi_0 = \phi_0^*$, $\psi_1 = \phi_1^*$, and $v = \psi_1 - \psi_0$. For $t \in [0,1]$, define $\psi_t = \psi_0 + tv$, and finally $\phi_t = \psi_t^*$. Then, $\phi_t$ is a strongly convex function, belongs to $C^2(\mathbb{R}^d)$, and*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}_\rho(\psi_t) = -\int_\mathcal{X} v(\nabla\phi_t(x))\mathrm{d}\rho(x) \tag{2.19}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{K}_\rho(\psi_t) = \int_\mathcal{X} \langle \nabla v(\nabla\phi_t(x)), D^2\phi_t(x)\cdot\nabla v(\nabla\phi_t(x))\rangle \mathrm{d}\rho(x). \tag{2.20}$$

*Proof.* A perfectly rigorous proof may be found in [42, Proposition 2.2]. Here, we shall remain at a formal level, assuming that all objects are well-defined. The maximum in

$$\max_{y\in\mathcal{Y}} \langle x, y\rangle - \psi_t(y)$$

is attained at $y_x \in \mathcal{Y}$ for which $x = \nabla\psi_t(y_x)$, which is equivalent to $y_x = \nabla\psi_t^*(x)$ according to (2.10). Therefore, by the envelope theorem,

$$\psi_{t+\varepsilon}^*(x) = \max_{y\in\mathcal{Y}} \langle x, y\rangle - \psi_t(y) - \varepsilon v(y) = \psi^*(x) - \varepsilon v(\nabla\psi_t^*(x)) + o(\varepsilon) \tag{2.21}$$

as $\varepsilon \to 0$. In other words $\frac{\mathrm{d}}{\mathrm{d}t}\psi_t^*(x) = -v(\nabla\psi_t^*(x))$, and integrating against $\rho$ we get (2.19).

For (4.24), applying (2.19) to $\psi_t$ we see that we need to evaluate

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\mathcal{X} v(\nabla\phi_t(x))\mathrm{d}\rho(x).$$

Using the chain rule and the fact that

$$\frac{\mathrm{d}}{\mathrm{d}t}\nabla\phi_t(x) = \nabla\frac{\mathrm{d}}{\mathrm{d}t}\phi_t(x) = \nabla\frac{\mathrm{d}}{\mathrm{d}t}\psi_t^*(x) = -D^2\phi_t(x)\cdot\nabla v(\nabla\phi_t(x))$$

due to (2.21), we get (4.24). □

## 2.4 Brascamp-Lieb and Prékopa-Leindler inequalities

In this section, we make a small detour in the world of *functional inequalities*: indeed, to show the strong convexity of the functional $\mathcal{K}_\rho$, the Brascamp-Lieb inequality (due to Herm Jan Brascamp and Eliott Lieb in 1976) will play a key role. Therefore, in this section, we state and prove this inequality, which is a kind of Poincaré inequality with respect to log-concave densities. It can be proved as a consequence of the following inequality due to Prékopa and Leindler.

**Theorem 2.8** (Prékopa-Leindler inequality)**.** *Let $f, g, h : \mathbb{R}^d \to [0,\infty)$ be integrable functions and $0 < \lambda < 1$. Assume that for any $x, y \in \mathbb{R}^d$,*

$$h(\lambda x + (1-\lambda)y) \geq f^\lambda(x) g^{1-\lambda}(y). \tag{2.22}$$

*Then,*

$$\int_{\mathbb{R}^d} h \geq \left(\int_{\mathbb{R}^d} f\right)^\lambda \left(\int_{\mathbb{R}^d} g\right)^{1-\lambda}.$$

Before showing Theorem 2.8, let us explain its relation to the classical *Brunn-Minkowski inequality*, which asserts that for any two compact subsets $A, B$ of $\mathbb{R}^d$,

$$|A+B|^{1/d} \geq |A|^{1/d} + |B|^{1/d} \tag{2.23}$$

where $|\cdot|$ denotes the $d$-dimensional volume. The Prékopa-Leindler inequality is actually a functional version of the Brunn-Minkowski inequality: for instance, Theorem 2.8 implies (2.23), as shown by the following elementary argument. Let $m = |A|^{1/d} + |B|^{1/d}$ and let $\lambda = \frac{|A|^{1/d}}{m}$, so that $1 - \lambda = \frac{|B|^{1/d}}{m}$. Applying Theorem 2.8 with $f = \mathbf{1}_{\widetilde{A}}$, $g = \mathbf{1}_{\widetilde{B}}$ and $h = \mathbf{1}_{\lambda\widetilde{A}+(1-\lambda)\widetilde{B}}$ where $\widetilde{A} = \frac{1}{|A|^{1/d}}A$ and $\widetilde{B} = \frac{1}{|B|^{1/d}}B$ (both $\widetilde{A}$ and $\widetilde{B}$ have volume 1) we get

$$1 \le |\lambda\widetilde{A} + (1-\lambda)\widetilde{B}| = \left|\frac{A+B}{m}\right|$$

from which (2.23) follows.

*Proof of Theorem 2.8.* We first show that it is sufficient to prove the result in dimension $d = 1$. Assume that Theorem 2.8 holds in dimensions $d_1$ and $d_2$, and let us show that it also holds in dimension $d = d_1 + d_2$. Let $f, g, h$ satisfy the assumptions in dimension $d = d_1 + d_2$, and notice that the functions $f(x_1, \cdot)$, $g(y_1, \cdot)$ and $h(\lambda x_1 + (1-\lambda)y_1, \cdot)$ satisfy (2.22) (in dimension $d_2$). Therefore,

$$\|h(\lambda x_1 + (1-\lambda)y_1, \cdot)\|_{L^1(\mathbb{R}^{d_2})} \ge \|f(x_1, \cdot)\|^{\lambda}_{L^1(\mathbb{R}^{d_2})}\|g(y_1, \cdot)\|^{1-\lambda}_{L^1(\mathbb{R}^{d_2})}$$

Defining $H(x_1) = \|h(x_1, \cdot)\|_{L^1(\mathbb{R}^{d_2})}$ and similarly $F, G$, we have

$$H(\lambda x_1 + (1-\lambda)y_1) \ge F(x_1)^{\lambda}G(y_1)^{1-\lambda}.$$

Applying Theorem 2.8 (in dimension $d_1$) we get

$$\|H\|_{L^1(\mathbb{R}^{d_1})} \ge \|F\|^{\lambda}_{L^1(\mathbb{R}^{d_2})}\|G\|^{1-\lambda}_{L^1(\mathbb{R}^{d_2})}.$$

But $\|H\|_{L^1(\mathbb{R}^{d_1})} = \|h\|_{L^1(\mathbb{R}^{d_1+d_2})}$ according to Tonelli's theorem, and similarly for $F, G$, which concludes the proof in dimension $d = d_1 + d_2$.

There remains to show that Theorem 2.8 holds for $d = 1$. By homogeneity we may assume $\int_{\mathbb{R}} f = \int_{\mathbb{R}} g = 1$. We may also assume by an approximation argument that $f, g$ are continuous and positive. We consider the probability measures $\mathrm{d}\mu(x) = f(x)\mathrm{d}x$ and $\mathrm{d}\nu(x) = g(x)\mathrm{d}x$, and denote by $F_\mu(t) = \int_{-\infty}^t f(x)\mathrm{d}x$ and $F_\nu(t) = \int_{-\infty}^t g(x)\mathrm{d}x$ their cumulant functions. Consider $T = F_\nu^{-1} \circ F_\mu$ the optimal transport map from $\mu$ to $\nu$, which is monotone according to Section 1.3 (even strictly monotone, since we assumed $f, g > 0$). The change of variables formula yields $f(x) = g(T(x))T'(x)$. Applying (2.22) with $y = T(x)$ and using the change of variables $z = \lambda x + (1-\lambda)T(x)$, we get

$$\begin{aligned}\int_{\mathbb{R}} h(z)\mathrm{d}z = \int_{\mathbb{R}} h(\lambda x + (1-\lambda)T(x))(\lambda + (1-\lambda)T'(x))\mathrm{d}x &\ge \int_{\mathbb{R}} f^{\lambda}(x)g^{1-\lambda}(T(x))T'(x)^{1-\lambda}\mathrm{d}x \\ &= \int_{\mathbb{R}} f^{\lambda}(x)f^{1-\lambda}(x)\mathrm{d}x = 1\end{aligned}$$

where we used the arithmetic-geometric inequality $\lambda a + (1-\lambda)b \ge a^{\lambda}b^{1-\lambda}$. □

It is also possible to prove the Prékopa-Leindler inequality directly in dimension $d$ using a generalization of the above one-dimensional argument, see [105, Section 6.1.4].

We now state the Brascamp-Lieb inequality and provide a proof extracted from Bobkov-Ledoux [12] (see also [65, Exercise 2.2.11]). Recall that a probability density $\sigma$ on a convex set $Q \subset \mathbb{R}^d$ is called *logarithmically concave*, or log-concave, if there exists a convex function

$V : Q \to \mathbb{R}$ such that $\sigma = e^{-V}$. Also, recall that the *variance* of a function $f$ with respect to a probability measure $\rho$ on a set $\mathcal{X}$ is defined as

$$\mathrm{Var}_\rho(f) = \int_{\mathcal{X}} \left( f - \int_{\mathcal{X}} f \mathrm{d}\rho \right)^2 \mathrm{d}\rho$$

and that it may be characterized as

$$\mathrm{Var}_\rho(f) = \min_{\lambda \in \mathbb{R}} \int_{\mathcal{X}} (f - \lambda)^2 \mathrm{d}\rho. \tag{2.24}$$

**Theorem 2.9** (Brascamp-Lieb variance inequality)**.** *Let $\rho_0 = e^{-V}\mathrm{d}x$ be a probability measure on $\mathbb{R}^d$, where $V \in C^2(\mathbb{R}^d)$ is assumed to be strictly convex. Then every smooth function $f$ on $\mathbb{R}^d$ verifies*

$$\mathrm{Var}_{\rho_0}(f) \leq \int_{\mathbb{R}^d} \langle \nabla f, (D^2 V)^{-1} \nabla f \rangle \mathrm{d}\rho_0.$$

As we said already, the Brascamp-Lieb inequality is a kind of Poincaré inequality with respect to log-concave densities. Its strength is that for an arbitrary strictly convex $V \in C^2(\mathbb{R}^d)$, the Poincaré constant is 1.

*Proof of Theorem 2.9.* For $u : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, let

$$\mathcal{I}(u) = \log \int_{\mathbb{R}^d} e^{-u^*}.$$

It follows from the definition of the Legendre transform that

$$(\lambda u + (1-\lambda)v)^*(\lambda x + (1-\lambda)y) \leq \lambda u^*(x) + (1-\lambda)v^*(y).$$

Hence applying Theorem 2.8 to $f = e^{-u^*}$, $g = e^{-v^*}$ and $h = e^{-(\lambda u + (1-\lambda)v)^*}$, we get that

$$e^{\mathcal{I}(\lambda u + (1-\lambda)v)} = \int_{\mathbb{R}^d} e^{-(\lambda u + (1-\lambda)v)^*} \geq \left( \int_{\mathbb{R}^d} e^{-u^*} \right)^\lambda \left( \int_{\mathbb{R}^d} e^{-v^*} \right)^{1-\lambda} = e^{\lambda \mathcal{I}(u) + (1-\lambda)\mathcal{I}(v)}$$

which means that $\mathcal{I}$ is concave. Let us compute the second derivative of $\mathcal{I}$. First, letting $u_t = u + tw$,

$$\frac{\mathrm{d}}{\mathrm{d}t} e^{\mathcal{I}(u_t)} = \frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} e^{-u_t^*} = \int_{\mathbb{R}^d} w(\nabla u_t^*) e^{-u_t^*}$$

thus

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{I}(u_t) = e^{-\mathcal{I}(u_t)} \int_{\mathbb{R}^d} w(\nabla u_t^*) e^{-u_t^*}.$$

Then

$$\begin{aligned}
\frac{\mathrm{d}^2}{\mathrm{d}t^2} \mathcal{I}(u_t) &= \left( -\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{I}(u_t) \right) e^{-\mathcal{I}(u_t)} \int_{\mathbb{R}^d} w(\nabla u_t^*) e^{-u_t^*} + e^{-\mathcal{I}(u_t)} \int_{\mathbb{R}^d} \frac{\mathrm{d}}{\mathrm{d}t} \left( w(\nabla u_t^*) e^{-u_t^*} \right) \\
&= -\left( e^{-\mathcal{I}(u_t)} \int_{\mathbb{R}^d} w(\nabla u_t^*) e^{-u_t^*} \right)^2 + e^{-\mathcal{I}(u_t)} \int_{\mathbb{R}^d} w(\nabla u_t^*)^2 e^{-u_t^*} \\
&\qquad - e^{-\mathcal{I}(u_t)} \int_{\mathbb{R}^d} \langle (D^2 u_t^*) \nabla w(\nabla u_t^*), \nabla w(\nabla u_t^*) \rangle e^{-u_t^*} \\
&= \mathrm{Var}_{\rho_t}(f) - \int_{\mathbb{R}^d} \langle (D^2 u_t^*)^{-1} \nabla f, \nabla f \rangle \mathrm{d}\rho_t.
\end{aligned}$$

where $\rho_t = e^{-\mathcal{I}(u_t)} e^{-u_t^*} = \frac{e^{-u_t^*}}{\int_{\mathbb{R}^d} e^{-u_t^*}}$ and $f = w \circ \nabla u_t^*$. Hence for any $f$ (since $\nabla u_t^*$ is a diffeomorphism)

$$\mathrm{Var}_{\rho_t}(f) \leq \int_{\mathbb{R}^d} \langle (D^2 u_t^*)^{-1} \nabla f, \nabla f \rangle \mathrm{d}\rho_t.$$

Taking $t = 0$ and $u = V^*$, i.e., $u_0^* = u^* = V$, this concludes the proof. □

For the application we have in mind, we need the following version of the Brascamp-Lieb inequality in compact convex sets (instead of $\mathbb{R}^d$), which can be deduced from Theorem 2.9 by an elementary approximation argument:

**Theorem 2.10** (Brascamp-Lieb inequality in compact convex sets)**.** *Let $\mathcal{X} \subset \mathbb{R}^d$ be a compact, convex set and let $\rho_0 = e^{-V}\mathrm{d}x$ be a probability measure on $\mathcal{X}$, where $V \in C^2(\mathcal{X})$ is assumed to be strictly convex. Then every smooth function $f$ on $\mathcal{X}$ verifies*

$$\mathrm{Var}_{\rho_0}(f) \leq \int_{\mathcal{X}} \langle \nabla f, (D^2 V)^{-1} \nabla f \rangle \mathrm{d}\rho_0$$

## 2.5 Variance inequality in compact convex sets

In this section, we prove a "*variance inequality*", i.e., an upper bound on the variance of the difference of two Kantorovich potentials corresponding to two different target measures. It reflects a form of strong convexity of $\mathcal{K}_\rho$. The estimate (2.25) below, proved thanks to the Brascamp-Lieb inequality, will be of fundamental importance in the sequel.

**Theorem 2.11** (Variance inequality for Kantorovich potentials)**.** *Let $Q \subset \mathbb{R}^d$ be a compact convex set with non-empty interior, let $\sigma$ be a log-concave probability density over $Q$ and let $\rho$ be another probability density over $Q$ satisfying $m_\rho \sigma \leq \rho \leq M_\rho \sigma$ for some constants $M_\rho \geq m_\rho > 0$. Let $\mathcal{Y} \subset \mathbb{R}^d$ be a compact set, and let $R_{\mathcal{Y}} = \max_{y \in \mathcal{Y}} \|y\|$. Then, for all $\psi_0, \psi_1 \in \mathcal{C}^0(\mathcal{Y})$,*

$$\frac{m_\rho}{M_\rho} \frac{1}{e R_{\mathcal{Y}} \mathrm{diam}(Q)} \mathrm{Var}_\rho(\psi_1^* - \psi_0^*) \leq \langle \psi_1 - \psi_0 \mid \nabla \psi^*_{0\#}\rho - \nabla \psi^*_{1\#}\rho \rangle. \tag{2.25}$$

An example to keep in mind is when $\sigma$ is the characteristic function of $Q$, normalized to be a probability density. Another important example is when $\rho$ itself is log-concave, in which case we may take $\sigma = \rho$ and $m_\rho / M_\rho = 1$.

The inequality (2.25) is not exactly a strong convexity estimate on $\mathcal{K}_\rho$ since primal (and not dual) Kantorovich potentials appear in the left-hand side. In the original proof of [42], a true strong convexity estimate with dual potentials was obtained. However it was not strong enough to imply Theorem 1.10, contrarily to (2.25). Moreover, the constant in (2.25) is *dimension-free*, whereas the estimate in [42] was not.

*Proof of Theorem 2.11.* Fix $\psi_0, \psi_1 \in \mathcal{C}^0(\mathcal{Y})$. Let $v = \psi_1 - \psi_0$ and $\psi_t = \psi_0 + tv = (1-t)\psi_0 + t\psi_1$ for $t \in [0,1]$. Set also $\phi_t = \psi_t^*$. In the sequel, we assume that $\phi_t$ has all the nice properties which make the involved objects well-defined, in particular we assume that Lemma 2.7 applies and that $\psi_0, \psi_1$ are $C^2$. The reduction to this case relies on approximation arguments written in detail in [72], which we do not reproduce here. Readers might wonder how it is possible to reduce to a case where so much regularity is assumed: at first sight, this puts us in the setting of Theorem 1.5, where the Lipschitzness of $T_\mu$ yields a $C^{1,1}$ bound on $\psi_\mu$. However, the key point here is that the final estimate (2.25) does not involve any bound on the regularity of primal and dual potentials, whereas the Lipschitz constant of $T_\mu$ appears explicitly in the constant $C$ in

Theorem 1.5. Therefore, (2.25) is amenable to regularization arguments, whereas Theorem 1.5 is not.

We have

$$\begin{aligned}\langle \psi_1 - \psi_0 \mid \nabla\psi^*_{0\#}\rho - \nabla\psi^*_{1\#}\rho\rangle &= \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}_\rho(\psi_t)\Big|_{t=1} - \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}_\rho(\psi_t)\Big|_{t=0} \\ &= \int_0^1 \frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{K}_\rho(\psi_t)\mathrm{d}t \\ &= \int_0^1 \int_Q \langle \nabla v(\nabla\phi_t), D^2\phi_t \cdot \nabla v(\nabla\phi_t)\rangle \mathrm{d}\rho\mathrm{d}t\end{aligned}$$

according to Lemma 2.7. We introduce $w_t(x) = v(\nabla\phi_t(x))$, then $\nabla w_t = D^2\phi_t \cdot \nabla v(\nabla\phi_t)$, and we get

$$\int_Q \langle \nabla v(\nabla\phi_t), D^2\phi_t \cdot \nabla v(\nabla\phi_t)\rangle \mathrm{d}\rho = \int_Q \langle \nabla w_t, (D^2\phi_t)^{-1} \cdot \nabla w_t\rangle \mathrm{d}\rho \tag{2.26}$$

($D^2\phi_t$ is invertible thanks to Lemma 2.4, because we assumed that $\psi_0, \psi_1$ are $C^2$).

There is a first approach to lower bound the last expression by directly applying the Brascamp-Lieb inequality (with $\rho_0 = e^{-\phi_t}$, properly normalized). This approach has the drawback that it yields a constant in (2.25) much worse than the one stated, in particular not good enough to prove Theorem 1.10.

To circumvent this, we write $\sigma = e^{-V}$ and set $\rho_t = Z_t^{-1}e^{-V-\phi_t}$ where $Z_t$ is a normalizing constant, so that $\rho_t$ is a probability measure. The idea is to apply the Brascamp-Lieb inequality with $\rho_t$. For this, we need to replace $\rho$ by $\rho_t$ in the right-hand side of (2.26). We denote by $m_t$ and $M_t$ the minimum and the maximum of $\phi_t(x)$ over $x \in Q$, and let $r = \sup_{t\in[0,1]} M_t - m_t$. We get

$$\begin{aligned}\rho_t(x) &\geq Z_t^{-1}e^{-M_t}\sigma \geq Z_t^{-1}e^{-M_t}M_\rho^{-1}\rho, \\ \rho_t(x) &\leq Z_t^{-1}e^{-m_t}\sigma \leq Z_t^{-1}e^{-m_t}m_\rho^{-1}\rho.\end{aligned}$$

In particular, using (2.24),

$$\mathrm{Var}_{\rho_t}(f) = \int_{\mathcal{X}} |f - \overline{f}|^2 \mathrm{d}\rho_t \geq \alpha \int_{\mathcal{X}} |f - \overline{f}|^2 \mathrm{d}\rho \geq \alpha \mathrm{Var}_\rho(f) \tag{2.27}$$

where $\alpha = Z_t^{-1}e^{-M_t}M_\rho^{-1}$ and $\overline{f} = \int_{\mathcal{X}} f\mathrm{d}\rho_t$. Then

$$\begin{aligned}\int_Q \langle \nabla w_t, (D^2\phi_t)^{-1} \cdot \nabla w_t\rangle \mathrm{d}\rho &\geq \int_Q \langle \nabla w_t, (D^2\phi_t + D^2V)^{-1} \cdot \nabla w_t\rangle \mathrm{d}\rho \\ &\geq Z_t e^{m_t} m_\rho \int_Q \langle \nabla w_t, (D^2\phi_t + D^2V)^{-1} \cdot \nabla w_t\rangle \mathrm{d}\rho_t \\ &\geq Z_t e^{m_t} m_\rho \mathrm{Var}_{\rho_t}(w_t) \qquad \text{(due to Theorem 2.10)} \\ &\geq e^{m_t - M_t}\frac{m_\rho}{M_\rho}\mathrm{Var}_\rho(w_t) \qquad \text{(due to (2.27))} \\ &\geq e^{-r}\frac{m_\rho}{M_\rho}\mathrm{Var}_\rho(w_t).\end{aligned}$$

Integrating this inequality over $t \in [0,1]$, there remains to lower bound $\int_0^1 \mathrm{Var}_\rho(w_t)\mathrm{d}t$. We notice that $\frac{\mathrm{d}}{\mathrm{d}t}\phi_t(x) = -v(\nabla\phi_t(x)) = -w_t(x)$ due to the same computation as in (2.21). Therefore we deduce from Minkowski's inequality

$$\int_0^1 \mathrm{Var}_\rho(w_t)\mathrm{d}t \geq \mathrm{Var}_\rho\left(\int_0^1 w_t \ \mathrm{d}t\right) = \mathrm{Var}_\rho\left(\int_0^1 \frac{\mathrm{d}\phi_t}{\mathrm{d}t}\mathrm{d}t\right) = \mathrm{Var}_\rho(\phi_1 - \phi_0).$$

All in all,

$$\langle \psi_1 - \psi_0 \mid \nabla\psi^*_{0\#}\rho - \nabla\psi^*_{1\#}\rho \rangle \geq e^{-r}\frac{m_\rho}{M_\rho}\mathrm{Var}_\rho(\phi_1 - \phi_0). \tag{2.28}$$

A scaling argument will allow us to replace $e^{-r}$ by $1/er$. Given $\lambda > 0$, we notice that $(\lambda\psi)^* = \lambda\psi^*(\cdot/\lambda)$. Applying the previous inequality to the functions $\psi_i^\lambda = \lambda\psi_i$ and to the dilated probability density $\rho_\lambda = (x \mapsto \lambda^{-1}x)_\#\rho$, and remarking that both $M_t$ and $m_t$ are multiplied by $\lambda$ under this scaling, we get

$$\lambda\langle \psi_1 - \psi_0 | \nabla\psi^*_{0\#}\rho - \nabla\psi^*_{1\#}\rho \rangle \geq \frac{m_\rho}{M_\rho}e^{-\lambda r}\lambda^2\mathrm{Var}_\rho(\psi_1^* - \psi_0^*).$$

Choosing $\lambda = r^{-1}$, we see that we can replace the constant $e^{-r}$ in (2.28) by $1/er$.

Finally, there remains to control $r$. Let $x, x' \in Q$, and $y \in \mathcal{Y}$ such that $\phi_t(x) = \langle x, y\rangle - \phi_t^*(y)$. Then

$$\phi_t(x') \geq \langle x', y\rangle - \phi_t^*(y) = \langle x' - x, y\rangle + \phi_t(x) \geq -\mathrm{diam}(Q)R_{\mathcal{Y}} + \phi_t(x). \tag{2.29}$$

Therefore $M_t - m_t \leq \mathrm{diam}(Q)R_{\mathcal{Y}}$, and taking the supremum over $t \in [0,1]$ we get that $r$ has the same upper bound, which concludes. □

**Remark 2.12.** *At this point, it is possible to conclude the proof of the stability of Kantorovich potentials when $\rho$ is supported on a compact,* convex *set, and bounded above and below on its support (i.e.,* (1.30) *when $\mathcal{X}$ is assumed to be convex). This recovers, with an improved constant, a result due to Delalande and Mérigot [42] (after anterior work by Berman, see the bibliographical notes in Section 2.7). For this, one just needs to take $\psi_0 = \psi_\mu$, $\psi_1 = \psi_\nu$, and upper bound the right-hand side in* (2.25) *thanks to the Kantorovich-Rubinstein duality formula. This is very similar to the argument at the end of the proof of Theorem 1.9 in Section 2.2.*

## 2.6 Stability of Kantorovich potentials for log-concave sources

In this section, we prove the first part of Theorem 1.10, namely (1.27). The key steps for this are to truncate the primal Kantorovich potentials in large balls, apply Theorem 2.11 and show that we do not lose too much by this truncation argument.

Let $\phi_\mu, \phi_\nu$ be the Kantorovich potentials in the quadratic optimal transport problems from $\rho$ to $\mu$ and $\rho$ to $\nu$ respectively. For $r > 0$ we set $\mathcal{B}_r = B(0, r)$ and we denote by $\phi_{\mu,r}, \phi_{\nu,r}$ the restriction of $\phi_\mu$ and $\phi_\nu$ to $\mathcal{B}_r$, extended by $+\infty$ outside $\mathcal{B}_r$. Then we set

$$\rho_r = \frac{\rho_{|\mathcal{B}_r}}{\rho(\mathcal{B}_r)}, \qquad \mu_r = (\nabla\phi_{\mu,r})_\#\rho_r, \qquad \nu_r = (\nabla\phi_{\nu,r})_\#\rho_r.$$

We also consider the dual Kantorovich potentials $\psi_{\mu,r} = \phi^*_{\mu,r}$, $\psi_{\nu,r} = \phi^*_{\nu,r}$ and recall that $\psi^*_{\mu,r} = \phi_{\mu,r}$ and $\psi^*_{\nu,r} = \phi_{\nu,r}$.

We apply Theorem 2.11 to $\rho_r$, taking for $\sigma$ the unique probability density over $\mathcal{B}_r$ whose density is proportional to $e^{-U}$ (recall that $\rho = e^{-U-F}$). This gives

$$\mathrm{Var}_{\rho_r}(\phi_{\mu,r} - \phi_{\nu,r}) \leq C_{\rho,\mathcal{Y}}\, r\langle \psi_{\nu,r} - \psi_{\mu,r} \mid \mu_r - \nu_r\rangle.$$

Since $\psi_{\mu,r} - \psi_{\nu,r}$ is $r$-Lipschitz (by a similar computation to (2.29)), we obtain by the Kantorovich-Rubinstein duality formula (2.11) that

$$\mathrm{Var}_{\rho_r}(\phi_{\mu,r} - \phi_{\nu,r}) \leq C_{\rho,\mathcal{Y}}\, r^2 W_1(\mu_r, \nu_r).$$

Using various truncation estimates which we do not detail here (but which are detailed in [72]), we get

$$\mathrm{Var}_\rho(\phi_\mu - \phi_\nu) \leq C_{\rho,\mathcal{Y}}\left(r^2 W_1(\mu,\nu) + r^2 m_0(r) + m_1(r)^2 + m_2(r)\right)$$

where

$$m_\ell(r) = \int_{\mathbb{R}^d \setminus \mathcal{B}_r} |x|^\ell d\rho(x).$$

Observe that $m_\ell(r) \leq C_{\rho,\ell} r^{d+\ell-2} e^{-\frac{1}{2}\kappa r^2}$ since $D^2 U \geq \kappa \mathrm{Id}$. Now, we notice that the quantity $\|\phi_\mu\|_{L^2(\rho)}$ is uniformly bounded over $\mu \in \mathcal{P}(\mathcal{Y})$, since $\phi_\mu$ is $R_{\mathcal{Y}}$-Lipschitz (due to Lemma 2.6), has vanishing mean, and $\rho$ has a finite second moment. Hence, $\|\phi_\mu - \phi_\nu\|_{L^2(\rho)}$ is uniformly bounded over $\mu, \nu \in \mathcal{P}(\mathcal{Y})$, and we may assume $W_1(\mu,\nu) < 1$ since otherwise the inequality (1.27) is trivial. Finally, we optimize over $r$, by taking

$$r = (4\kappa^{-1} |\log W_1(\mu,\nu)|)^{1/2}.$$

This yields (1.27).

## 2.7 Bibliographical notes

§2.1: Monge's original paper is [80]. The dual formulation of the Monge problem was introduced by Kantorovich, the founding father of linear programming, in [61]. His goal was to solve concrete problems for the Russian industry. The foreword to the English translation of his paper [61], written by an American scientist in the journal Management Science in 1958, is an historical gem: "[...] It is to be noted, however, that the problem of determining an effective method of actually acquiring the solution of a specific problem is not solved in this paper. In the category of development of such methods we seem to be currently, ahead of the Russians."

For a smooth introduction to duality in optimal transport, we refer to [105, Chapter 1]. Brenier's theorem was proved in [20], after several works by many authors, among which Knott and Smith, and Rachev and Rüschendorf. The strong duality theorem (2.6) can be formulated for general costs, see [106, Theorem 5.10]. The Kantorovich-Rubinstein formula is a particular case of this strong duality, when the cost is the distance function: $c(x,y) = |x-y|$.

§2.2: The first result similar to Theorem 1.9 is due to Gigli [53], and in the form stated in Section 2.2 it is due to Mérigot-Delalande-Chazal [79]. Similar ideas have been developed later for numerical purposes (see for instance [73]) and in statistical optimal transport (see Section 5.1).

§2.3: For a rigorous proof of Lemma 2.7, see [42, Proposition 2.2].

§2.4: The Prékopa-Leindler inequality dates back to [92], [69], and the Brascamp-Lieb inequality to [18]. We refer to [105, Chapter 6] for a nice presentation of these inequalities, and related ones. The Brascamp-Lieb inequality in compact convex sets (Theorem 2.10) is a particular case of [71], [66], but in the simple Euclidean setting considered here, it follows directly from Theorem 2.9.

§2.5: Berman [11] was the first to prove an inequality of the form (2.25), using complex geometry. In his result corresponding to the variance inequality (2.25), the right-hand side is raised to the power $1/2^{d-1}$, which makes it non-optimal. Inspired by his paper, Mérigot-Delalande-Chazal proved in [79] the inequality $\|T_\mu - T_\nu\|_{L^2(\rho)} \leq C W_p(\mu,\nu)^{2/15}$, using a very instructive proof technique. Their arguments are a sort of "discrete version" of the arguments later developed in [42]. They first rely on an approximation argument allowing them to assume that $\mu, \nu$ are discrete, and then they leverage specific features of semi-discrete optimal transport, notably the geometry of Laguerre cells. In their paper, the Brunn-Minkowski inequality (2.23) plays the same role as the Brascamp-Lieb inequality in [42] and in the proof of Theorem 2.11 presented above. In the paper [42], Delalande and Mérigot prove the same inequality as in Theorem 2.11, except with a dimension-dependent constant which is worse constant than ours (notably because they do not compare $\rho$ to a log-concave measure $\sigma$). Inspired by the paper [81] by Mischler and Trevisan, which proves a variance inequality with a good constant for log-concave source measures, we found with Mérigot in [72] the simple proof of Theorem 2.11 presented in Section 2.5, which shortcuts several arguments of [42]. The approximation arguments which are not presented in the proof of Theorem 2.11 above are written in detail in [72].

Another approach to variance inequalities is possible, using entropic optimal transport and the Prékopa-Leindler inequality. This approach is presented in Chapter 4.

# 3 Gluing methods

In this chapter, we pursue a broad generalization of the results from the previous chapter. To this end, we introduce what we call gluing methods (not to be confused with the 'gluing lemma' in optimal transport). These methods combine two key ingredients: on the one hand, decompositions of domains and measures into local pieces, to which the techniques of Chapter 2 apply; on the other hand, arguments that assemble these local estimates into global ones. The approach is inspired by proofs of Sobolev–Poincaré inequalities developed in the 1980s. We then present in Section 3.7 a series of examples and counterexamples, providing explicit source and target measures for which optimal transport maps can be computed. These cases serve both to test the optimality of the main results and to illustrate the sharpness of their assumptions. They also demonstrate how optimal transport maps can be computed explicitly in concrete situations, thereby making the stability theory more tangible. We finally explain in Section 3.8 how quantitative stability estimates on optimal transport maps can be derived from the quantitative stability of Kantorovich potentials. These estimates are likely not optimal, and this part of the theory remains therefore not entirely satisfactory. However, at the time of writing, the results presented here represent the current state of the art.

## 3.1 Gluing arguments in a nutshell

This section remains at a panoramic level, while complete arguments are provided in the next sections. We start with a toy example. Let $Q_1, Q_2$ be two open sets in $\mathbb{R}^d$ such that $Q_1 \cap Q_2 \neq \emptyset$, and let $\rho$ be a probability density over $Q_1 \cup Q_2$, bounded above and below on $Q_1 \cup Q_2$. For $i = 1, 2$, let

$$\rho_{Q_i} = \frac{\rho_{|Q_i}}{\rho(Q_i)} \tag{3.1}$$

be the restriction of $\rho$ to $Q_i$, normalized to be a probability measure. We show that

$$\mathrm{Var}_\rho(f) \leq C(\mathrm{Var}_{\rho_{Q_1}}(f) + \mathrm{Var}_{\rho_{Q_2}}(f)) \tag{3.2}$$

for some explicit constant $C$, roughly proportional to the quotient $\max(\rho(Q_1), \rho(Q_2))/\rho(Q_1 \cap Q_2)$. This is a quantitative version of the fact that if $f$ is constant on $Q_1$ and constant on $Q_2$, then it is constant on $Q_1 \cup Q_2$, since $Q_1 \cap Q_2 \neq \emptyset$. The less $Q_1$ and $Q_2$ overlap, the larger $C$ has to be; and if $Q_1 \cap Q_2 = \emptyset$, then (3.2) becomes false.

We will prove a general version of (3.2) for a *finite or infinite collection* $\mathcal{F}$ of well-chosen sets $Q_i$ (often chosen to be cubes) whose union is equal to the whole domain $\mathcal{X}$:

$$\mathrm{Var}_\rho(f) \leq C \sum_{Q_i \in \mathcal{F}} \rho(Q_i) \mathrm{Var}_{\rho_{Q_i}}(f) \tag{3.3}$$

for some $C < +\infty$ (and any $f$). For this inequality to be true, one has to make some assumptions on $\rho$; and to carefully design the family $\mathcal{F}$.

### 3.1.1 From (3.3) to the stability of Kantorovich potentials.

Once (3.3) is shown, it takes only a few lines to complete the proof of the stability of Kantorovich potentials. For instance, if the $Q_i$ are cubes covering $\mathcal{X} \subset \mathbb{R}^d$, what we need to assume is:

(1) the variation of $\rho$ on each cube is uniformly bounded over $\mathcal{F}$:

$$\sup_{Q_i \in \mathcal{F}} \frac{\sup_{Q_i} \rho}{\inf_{Q_i} \rho} \leq E < +\infty \tag{3.4}$$

(2) there exists $A > 0$ such that any $Q_i \in \mathcal{F}$ intersects at most $A$ other cubes $Q_j \in \mathcal{F}$ (including itself).

So let us show how to deduce stability of Kantorovich potentials from (3.3).

*Claim.* There exists $C' > 0$ such that $\psi_0, \psi_1 \in C^0(\mathcal{Y})$, there holds

$$\mathrm{Var}_\rho(\psi_1^* - \psi_0^*) \leq C' \langle \psi_0 - \psi_1 \mid (\nabla \psi_1^*)_\# \rho - (\nabla \psi_0^*)_\# \rho \rangle. \tag{3.5}$$

*Proof of the claim.* For this we apply Theorem 2.11 in each cube $Q_i$, to $\rho_{Q_i}$ defined by (3.1) (with $\sigma$ the normalized Lebesgue measure on $Q_i$). We get

$$\mathrm{Var}_{\rho_{Q_i}}(\psi_1^* - \psi_0^*) \leq eER_{\mathcal{Y}} \mathrm{diam}(\mathcal{X}) \langle \psi_0 - \psi_1 \mid (\nabla \psi_1^*)_\# \rho_{Q_i} - (\nabla \psi_0^*)_\# \rho_{Q_i} \rangle \tag{3.6}$$

Combining with (3.3) we get

$$\mathrm{Var}_\rho(f) \leq CeER_{\mathcal{Y}} \mathrm{diam}(\mathcal{X}) \sum_{Q_i \in \mathcal{F}} \langle \psi_0 - \psi_1 \mid (\nabla \psi_1^*)_\# \rho_{|Q_i} - (\nabla \psi_0^*)_\# \rho_{|Q_i} \rangle \tag{3.7}$$

(recall that $\rho_{|Q_i}$ is simply the restriction of $\rho$ to $Q_i$, not renormalized). The right-hand side in (3.7) is a sum of positive terms (due to (3.6)) and $\sum \rho_{|Q_i} \leq A\rho$ (when absolutely continuous measures and densities are identified), so (3.5) is almost proved.

Here is a formal proof. We define a partition $\mathcal{F}'$ of $\mathcal{X}$ into convex sets as follows: $x, x' \in \mathcal{X}$ belong to the same element $P \in \mathcal{F}'$ if and only if they belong exactly to the same elements in $\mathcal{F}$. Each $P \in \mathcal{F}'$ is an intersection of $n_P \leq A$ cubes according to (2), thus it is convex. The sum in (3.7) may be written equivalently as

$$\sum_{P \in \mathcal{F}'} n_P \langle \psi_0 - \psi_1 \mid (\nabla \psi_1^*)_\# \rho_{|P} - (\nabla \psi_0^*)_\# \rho_{|P} \rangle$$

Moreover, each term in this sum is non-negative due to Theorem 2.11 (or more directly due to the convexity of $\mathcal{K}_{\rho_P}$ - we do not need strong convexity here). Recalling that the elements of $\mathcal{F}'$ form a partition of $\mathcal{X}$, we obtai

$$\begin{aligned}\sum_{P \in \mathcal{F}'} n_P \langle \psi_0 - \psi_1 \mid (\nabla \psi_1^*)_\# \rho_{|P} - (\nabla \psi_0^*)_\# \rho_{|P} \rangle &\leq A \sum_{P \in \mathcal{F}'} \langle \psi_0 - \psi_1 \mid (\nabla \psi_1^*)_\# \rho_{|P} - (\nabla \psi_0^*)_\# \rho_{|P} \rangle \\ &= A \langle \psi_0 - \psi_1 \mid (\nabla \psi_1^*)_\# \rho - (\nabla \psi_0^*)_\# \rho \rangle.\end{aligned}$$

Therefore (3.5) holds with $C' = ACeER_{\mathcal{Y}} \mathrm{diam}(\mathcal{X})$. □

Finally, we apply the claim (3.5) to $\psi_0 = \phi_\mu^*$ and $\psi_1 = \phi_\nu^*$, where the Legendre transform is computed as a supremum over $\mathcal{X}$, and $\phi_\mu, \phi_\nu$ are the Kantorovich potentials from $\rho$ to $\mu$ and $\rho$ to $\nu$. We get

$$\mathrm{Var}_\rho(\psi_1^* - \psi_0^*) \leq C' \langle \phi_\mu - \phi_\nu \mid \nu - \mu \rangle \leq C' \mathrm{diam}(\mathcal{Y}) W_1(\mu, \nu)$$

where the last inequality comes from the Kantorovich-Rubinstein duality formula and the fact that $\phi_\mu$ and $\phi_\nu$ are $\mathrm{diam}(\mathcal{Y})$-Lipschitz due to Proposition 2.1. This concludes the proof.

### 3.1.2 Two strategies to prove (3.3).

All in all, to establish the stability of Kantorovich potentials (1.30) (or stability of Kantorovich potentials in other situations, like those described in Section 1.7) we only need to show that (3.3) holds for some well-chosen family $\mathcal{F}$. To prove (3.3) we designed two strategies, which are complementary in terms of the families of probability measures $\rho$ that they allow to handle:

- Strategy 1: an approach inspired by the proofs of Sobolev-Poincaré inequalities in the 1980's, where the elements of $\mathcal{F}$ are cubes, and we consider chains of cubes, called Boman chains, in which the variances are controlled. When applicable, this approach yields the results in their sharpest forms, and is exactly tailored to handle delicate cases like John domains in Theorem 1.12, and degenerate densities in bounded domains as in Section 1.7.
- Strategy 2: an approach through spectral graph theory. In this case, the constant $C$ in (3.3) is related to the spectral gap of the Laplacian on a natural graph constructed from the family $\mathcal{F}$. This strategy is sufficient to prove most of our results, but sometimes in slightly weaker forms - for example with this approach we are able to prove Theorem 1.12 only for bounded, connected Lipschitz domains. In some specific cases, this strategy works whereas the first one fails (explaining why we care about it!), for instance for measures $\rho$ on $\mathbb{R}^d$ which decay polynomially at infinity.

We describe these two strategies in more details in Sections 3.3 and 3.5. In the first strategy, the construction of the family $\mathcal{F}$ is based on a classical decomposition of any open set as a union of cubes, called the Whitney decomposition. In Section 3.2 we recall this decomposition and construct the family $\mathcal{F}$ when $\mathcal{X}$ is assumed to be a John domain.

## 3.2 Whitney decomposition and Boman chain condition

In this section, we first discuss the Whitney decomposition of open sets. Recall that a dyadic cube in $\mathbb{R}^d$ is a cube of the form

$$\Big\{(x_1, \ldots, x_d) \in \mathbb{R}^d \mid m_j 2^{-\ell} \leq x_j \leq (m_j+1)2^{-\ell} \text{ for any } j \in [d]\Big\} \tag{3.8}$$

for some $\ell \in \mathbb{Z}$ and $m = (m_1, \ldots, m_d) \in \mathbb{Z}^d$. The Whitney decomposition is a decomposition of any open set $\mathcal{X}$ into dyadic cubes in a way that each of these cubes has a sidelength which is comparable to its distance from the boundary of $\mathcal{X}$. Discovered in 1934 by Hassler Whitney to prove its extension theorem, it has been used since then in many areas to localize arguments: in harmonic analysis to localize singular integrals, in PDEs to localize estimates near boundaries of irregular domains, etc. An illustration is provided in Figure 3.

**Proposition 3.1** (Whitney decomposition)**.** *Let $\mathcal{X}$ be an open, proper and non-empty subset of $\mathbb{R}^d$. Then there exists a family of closed dyadic cubes $(P_j)_{j\in\mathbb{N}}$ such that*

*(a) The $P_j$'s have disjoint interiors and*

$$\bigcup_{j\in\mathbb{N}} P_j = \mathcal{X}.$$

*(b) If $\ell(P)$ denotes the sidelength of a cube $P$ and $\mathcal{X}^c = \mathbb{R}^d \setminus \mathcal{X}$, then for any $j \in \mathbb{N}$,*

$$\sqrt{d}\ell(P_j) \leq \operatorname{dist}(P_j, \mathcal{X}^c) \leq 4\sqrt{d}\ell(P_j). \tag{3.9}$$

*(c) If the boundaries of two cubes $P_j$ and $P_k$ touch then*

$$\frac{1}{4} \leq \frac{\ell(P_j)}{\ell(P_k)} \leq 4. \tag{3.10}$$

*(d) For a given $P_j$ there exist at most $12^d - 4^d$ cubes $P_k$'s that touch it.*

(e) *Let $1 < \sigma < 5/4$ and, for $j \in \mathbb{N}$, denote by $Q_j$ the cube with same center as $P_j$ and $\ell(Q_j) = \sigma\ell(P_j)$. Then*

$$\sum_{j\in\mathbb{N}} \mathbf{1}_{Q_j} \le 12^d$$

*where $\mathbf{1}_{Q_j}$ is the characteristic function of $Q_j$.*

*Proof of Proposition 3.1.* Observe that if $A, B$ are dyadic cubes, then either $A \cap B = \emptyset$, or $A \subset B$, or $B \subset A$. Thus for any $x \in \mathcal{X}$, it makes sense to consider the largest dyadic cube $P$ that contains $x$ and such that $\sqrt{d}\ell(P) \le \operatorname{dist}(P, \mathcal{X}^c)$. Denoting this cube by $P^{(x)}$, we see that the family $\{P^{(x)}\}$, once removed the redundancies, is a family of disjoint dyadic cubes. Let us check that this family has all desired properties.

First, each $x \in \mathcal{X}$ belongs to one of these cubes, and $P^{(x)} \subset \mathcal{X}$ for any $x \in \mathcal{X}$, therefore (a) holds. The left-hand side inequality in (3.9) is satisfied by definition. For the other inequality, we assume for the sake of a contradiction that $\operatorname{dist}(P_j, \mathcal{X}^c) > 4\sqrt{d}\ell(P_j)$ for some $j \in \mathbb{N}$, and we consider $P'_j$ any dyadic cube with same center as $P_j$, and doubled sidelength. Then, since the diameter of $P_j$ is equal to $\sqrt{d}\ell(P_j)$, we have

$$\operatorname{dist}(P'_j, \mathcal{X}^c) \ge \operatorname{dist}(P_j, \mathcal{X}^c) - \operatorname{diam}(P'_j) > 4\sqrt{d}\ell(P_j) - \operatorname{diam}(P'_j) = \operatorname{diam}(P'_j)$$

which contradicts the definition of $P_j$ and concludes the proof of (b). Assume now that $P_j$ and $P_k$ touch. Then

$$\sqrt{d}\ell(P_j) \le \operatorname{dist}(P_j, \mathcal{X}^c) \le \operatorname{dist}(P_j, P_k) + \operatorname{dist}(P_k, \mathcal{X}^c) \le 0 + 4\sqrt{d}\ell(P_k)$$

which proves (c). To prove (d), note that any cube $P_j$ is touched by exactly $3^d - 1$ dyadic cubes of the same sidelength. But each of them can contain at most $4^d$ cubes $P_k$ of length at least one-quarter of the length of $P_j$. This fact combined with (c) yields (d). Finally, to prove (e), we first observe that each $Q_j$ is contained in $\mathcal{X}$ by (b). If $x \in \mathcal{X}$, then $x \in P_{j_0}$ for some $j_0 \in \mathbb{N}$. If $P_j$ does not touch $P_{j_0}$, then $Q_j$ does not touch $P_{j_0}$ by the definition (3.8) of dyadic cubes. Therefore, $x$ may belong only to the cubes $Q_j$ for which $P_j$ touches $P_{j_0}$, and there are at most $12^d - 4^d + 1$ such cubes (including $P_{j_0}$) according to (d). □

Jan Boman discovered in 1982 that under some regularity assumptions on an open set $\mathcal{X} \subset \mathbb{R}^d$, its Whitney decomposition could be used to construct chains of cubes with certain useful properties, going from one "central cube" to cubes near the boundary. He used it to show $L^p$ estimates for some overdetermined elliptic systems of PDEs, and it was discovered a few years later by Bogdan Bojarski that these same "Boman chains" could be leveraged to prove Sobolev-Poincaré inequalities in Euclidean open sets (in a rather sharp form). We will extract from Bojarski's arguments an inequality of the form (3.3), which is exactly what is needed to prove the stability of Kantorovich potentials as we explained in Section 3.1. But let us first introduce the Boman chain condition, and show that it is satisfied in John domains.

**Definition 3.2** (Boman chain condition)**.** *A probability measure $\rho$ on an open set $\mathcal{X} \subset \mathbb{R}^d$ is said to satisfy the Boman chain condition with parameters $A, B, C > 0$ if there exists a covering $\mathcal{F}$ of $\mathcal{X}$ by open cubes $Q \in \mathcal{F}$ such that*

- *Any point cannot belong to more than $A$ cubes $Q \in \mathcal{F}$: for any $x \in \mathbb{R}^d$,*

$$\sum_{Q\in\mathcal{F}} \mathbf{1}_Q(x) \le A\mathbf{1}_{\mathcal{X}}(x). \tag{3.11}$$

- *For some fixed cube $Q_0$ in $\mathcal{F}$, called the central cube, and for every $Q \in \mathcal{F}$, there exists a sequence $Q_0, Q_1, \ldots, Q_N = Q$ of distinct cubes from $\mathcal{F}$ (a "Boman chain") such that for any $j \in \{0, \ldots, N\}$,*

$$Q \subset BQ_j \tag{3.12}$$

*where $BQ_j$ is the cube with same center as $Q_j$ and sidelength multiplied by $B$.*

- *Consecutive cubes of the above chain overlap quantitatively: for any $j \in \{0, \ldots, N-1\}$,*

$$\rho(Q_j \cap Q_{j+1}) \geq C^{-1} \max(\rho(Q_j), \rho(Q_{j+1})). \tag{3.13}$$

We notice that consecutive cubes of a Boman chain are comparable in size: for any $j \in \{0, \ldots, N-1\}$,

$$C^{-1} \leq \frac{\rho(Q_j)}{\rho(Q_{j+1})} \leq C. \tag{3.14}$$

Indeed,

$$\frac{\rho(Q_j)}{\rho(Q_{j+1})} \geq \frac{\rho(Q_j \cap Q_{j+1})}{\rho(Q_{j+1})} \geq C^{-1}$$

as a consequence of (3.13), and the reverse bound in (3.14) follows by the same argument.

The main case where Bomain chains exist is the following:

**Proposition 3.3** (John implies Boman)**.** *If $\rho$ is a probability measure on a John domain $\mathcal{X} \subset \mathbb{R}^d$, with a density bounded above and below on $\mathcal{X}$, then $\rho$ satisfies the Boman chain condition (for some $A, B, C$).*

The cubes $Q$ of the Boman chain condition are obtained by dilating the cubes of the Whitney decomposition of $\mathcal{X}$ by a factor slightly greater than 1. We refer to Figure 4 for an illustration of the cubes of the Boman chain condition. This picture should be compared with Figure 3 in which the Whitney decomposition of the same domain is drawn.

*Proof of Proposition 3.3.* Fix an arbitrary $\sigma \in (1, 5/4)$ and denote by $\mathcal{F}$ the family of cubes $Q_j = \sigma P_j$ obtained from the Whitney decomposition (Proposition 3.1). Without loss of generality, we may assume that $Q_0$ contains the central point $x_0$ highlighted in the definition of John domains. Consider an arbitrary $Q \in \mathcal{F}$, and denote by $x$ be the center of $Q$. Let $\gamma : [0, T] \to \mathcal{X}$ be a curve from $x$ to $x_0$ satisfying (1.29), and enumerate the cubes $Q = Q_N, \ldots, Q_0$ intersecting $\gamma$, ordered in such a way that $Q_j \cap Q_{j+1}$ is non-empty for all $j$. We may also remove redundancies and assume that in this chain of cubes, any two cubes are distinct.

First, (3.11) is a consequence of property (e) of Proposition 3.1. Moreover, since $Q_j \cap Q_{j+1} \neq \emptyset$, as we already saw in the proof of property (e) in Proposition 3.1, necessarily the boundaries of the Whitney cubes $P_j$ and $P_{j+1}$ touch, and thus (3.10) holds. From this, (3.13) follows immediately.

There only remains to check (3.12). Let $y_j$ denote the center of $Q_j$ (in particular $y_N = x$, and $y_j$ is also the center of $P_j$). By the triangle inequality, the distance from $y_j$ to any point in $Q_N$ is at most $|x - y_j| + \sqrt{d}\ell(Q_N)$, hence (3.12) boils down to showing the existence of $B$ depending only on $\mathcal{X}$ such that

$$\frac{|x - y_j| + \sqrt{d}\ell(Q_N)}{\frac{1}{2}\ell(Q_j)} \leq B. \tag{3.15}$$

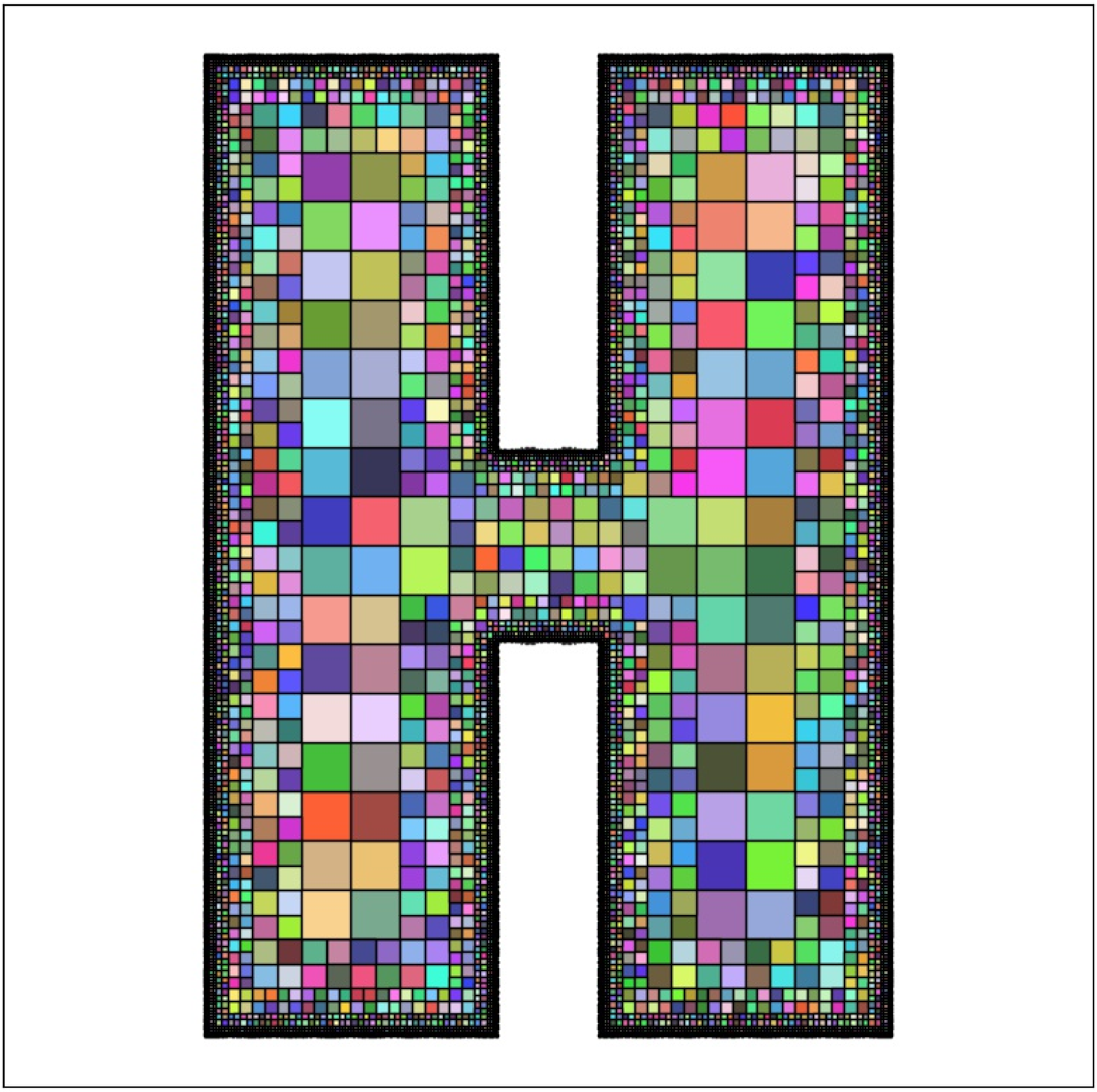


Figure 3: The Whitney decomposition of an H-shape in 2 dimensions. Each cube has a sidelength comparable to its distance to the boundary of the H. Courtesy of Quentin Mérigot.

Let us show that there exists $\kappa > 0$ depending only on $\mathcal{X}$ such that

$$|x - y_j| \leq \kappa \mathrm{dist}(y_j, \mathcal{X}^c), \qquad \ell(Q_N) \leq \kappa \mathrm{dist}(y_j, \mathcal{X}^c). \tag{3.16}$$

Since $\mathrm{d}(y_j, \mathcal{X}^c) \leq \sqrt{d}\ell(P_j) \leq \sqrt{d}\ell(Q_j)$, (3.15) follows immediately from (3.16). We turn to the proof of (3.16). Pick $t \in [0, T]$ such that $\gamma(t) \in Q_j$. Then

$$|x - y_j| \leq |x - \gamma(t)| + |\gamma(t) - y_j| \leq \frac{1}{\eta}\mathrm{dist}(\gamma(t), \mathcal{X}^c) + \mathrm{diam}(Q_j) \leq \frac{1}{\eta}\mathrm{dist}(y_j, \mathcal{X}^c) + \frac{\eta+1}{\eta}\mathrm{diam}(Q_j)$$

thanks to the triangle inequality and (1.29). Since the diameter of $Q_j$ is bounded above by $\sigma \mathrm{dist}(y_j, \mathcal{X}^c)$ according to (3.9), we obtain the first inequality in (3.16). Without loss of gener-

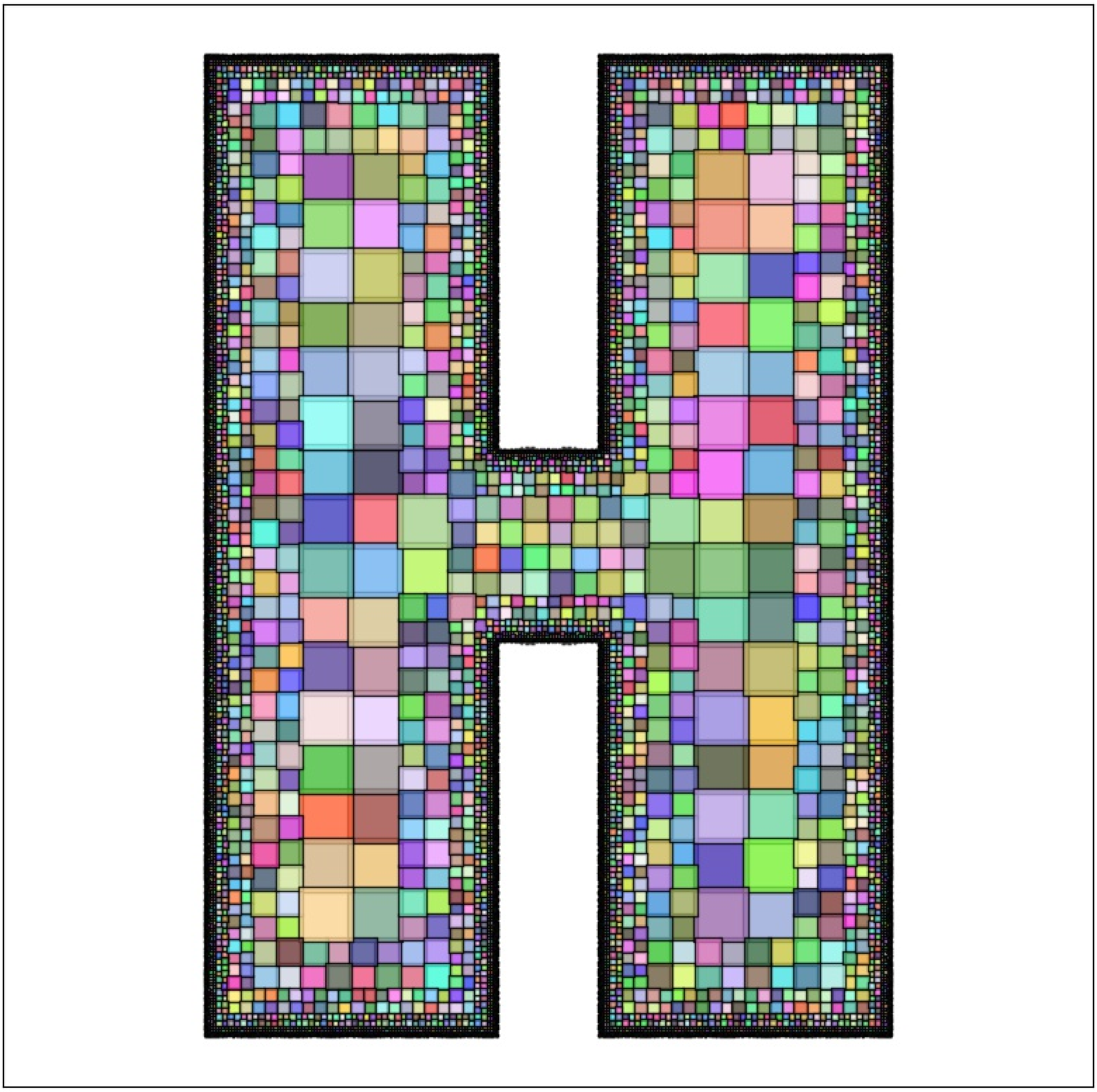


Figure 4: A Boman family for the uniform probability density on the H-shape, obtained by enlarging the sidelength of each cube of the Whitney decomposition by the same factor. This induces some overlap between the cubes. Courtesy of Quentin Mérigot.

ality, we assume from now on that $\eta \leq 1$, and we prove that for any $t \in [0, T]$,

$$\frac{\operatorname{dist}(\gamma(t), \mathcal{X}^c)}{\operatorname{dist}(x, \mathcal{X}^c)} \geq \frac{\eta}{2}. \tag{3.17}$$

By contradiction, if this does not hold, then

$$\operatorname{dist}(x, \mathcal{X}^c) \leq |x - \gamma(t)| + \operatorname{dist}(\gamma(t), \mathcal{X}^c) \leq |x - \gamma(t)| + \frac{\eta}{2}\operatorname{dist}(x, \mathcal{X}^c).$$

Since $\eta \leq 1$, we deduce $\operatorname{dist}(x, \mathcal{X}^c) \leq 2|x - \gamma(t)| \leq \frac{2}{\eta}\operatorname{dist}(\gamma(t), \mathcal{X}^c)$ where the last inequality comes from (1.29). Therefore (3.17) holds. To show the second inequality in (3.16), we finally write

$$\ell(Q_N) = \sigma\ell(P_N) \leq \frac{\sigma}{\sqrt{d}}\mathrm{dist}(x, \mathcal{X}^c) \leq \frac{2\sigma}{\eta\sqrt{d}}\mathrm{dist}(\gamma(t), \mathcal{X}^c)$$

where $t \in [0, T]$ is chosen to satisfy $\gamma(t) \in Q_j$. Since

$$\mathrm{dist}(\gamma(t), \mathcal{X}^c) \leq \mathrm{dist}(y_j, \mathcal{X}^c) + \mathrm{diam}(Q_j) \leq (1+\sigma)\mathrm{dist}(y_j, \mathcal{X}^c),$$

this concludes the proof of (3.16). □

Let us mention that some partial converse to Proposition 3.3 holds: if the characteristic function $\rho$ of some bounded open set $\mathcal{X}$ satisfies the Boman chain condition, then $\mathcal{X}$ is a John domain.

### 3.3 Strategy 1: Gluing variances with Boman chains

In this section we explain the first gluing technique. This technique applies only to measures with bounded support in $\mathbb{R}^d$. Roughly, for a measure $\rho$ satisfying the Boman chain condition (for some $A, B, C$), it consists in estimating the variance in any cube $Q \in \mathcal{F}$ to the variance in the central cube $Q_0$, via the construction of a Boman chain of overlapping cubes going from $Q_0$ to $Q$. The proof of the first part of Theorem 1.12, namely the stability of Kantorovich potentials (1.30), follows almost directly from the following proposition:

**Proposition 3.4** (Gluing variances with Boman chains). *Let $\rho$ be a probability measure satisfying the Boman chain condition, with covering family $\mathcal{F}$. Assume also that $\rho$ is doubling on the family $\mathcal{F}$: there exists $D > 0$ such that*

$$\forall Q \in \mathcal{F}, \qquad \rho(2Q) \leq D\rho(Q) \tag{3.18}$$

*where $2Q$ denotes the cube with same center as $Q$, and doubled sidelength. Then there exists $C > 0$ such that for any $f$,*

$$\mathrm{Var}_\rho(f) \leq C \sum_{Q_i \in \mathcal{F}} \rho(Q_i)\mathrm{Var}_{\rho_{Q_i}}(f). \tag{3.19}$$

*Proof of the stability of Kantorovich potentials* (1.30) *in John domains.* Let $\rho$ be a probability density on a John domain $\mathcal{X} \subset \mathbb{R}^d$, and assume that $\rho$ is bounded above and below on $\mathcal{X}$ by positive constants. According to Proposition 3.3, it satisfies the Boman chain condition for some $A, B, C > 0$. Therefore, the condition (3.4) holds, and moreover each cube of the family $\mathcal{F}$ appearing in the Definition 3.2 of the Boman chain condition intersects at most $A + 1$ other cubes. Therefore, we can apply Section 3.1.1, and it follows that (1.30) holds. □

*Proof of Proposition 3.4.* We set $f_Q = \frac{1}{\rho(Q)}\int_Q f \mathrm{d}\rho$ and $a_Q = (\mathrm{Var}_{\rho_Q}(f))^{1/2}$. Then

$$\begin{aligned}
\mathrm{Var}_\rho(f) \leq \int_{\mathcal{X}} |f(x) - f_{Q_0}|^2 \mathrm{d}\rho(x) &\leq \sum_{Q\in\mathcal{F}} \int_Q |f(x) - f_{Q_0}|^2 \mathrm{d}\rho(x) \\
&\leq 2\sum_{Q\in\mathcal{F}} \int_Q |f(x) - f_Q|^2 \mathrm{d}\rho(x) + \int_Q |f_Q - f_{Q_0}|^2 \mathrm{d}\rho(x) \qquad (3.20) \\
&= 2\sum_{Q\in\mathcal{F}} \rho(Q)\mathrm{Var}_{\rho_Q}(f) + 2\sum_{Q\in\mathcal{F}} \rho(Q)|f_Q - f_{Q_0}|^2.
\end{aligned}$$

The first sum is bounded above by the right-hand side in (3.19), therefore we only need to upper bound the second sum. The triangle inequality yields

$$|f_Q - f_{Q_0}| \leq \sum_{j=0}^{N-1} |f_{Q_j} - f_{Q_{j+1}}| \tag{3.21}$$

We estimate each term in the sum separately:

$$\begin{aligned}
|f_{Q_j} - f_{Q_{j+1}}|^2 &= \frac{1}{\rho(Q_j \cap Q_{j+1})} \int_{Q_j \cap Q_{j+1}} |f_{Q_j} - f_{Q_{j+1}}|^2 \\
&\leq \frac{2}{\rho(Q_j \cap Q_{j+1})} \left( \int_{Q_j \cap Q_{j+1}} |f_{Q_j} - f(x)|^2 \mathrm{d}\rho(x) + \int_{Q_j \cap Q_{j+1}} |f_{Q_{j+1}} - f(x)|^2 \mathrm{d}\rho(x) \right) \\
&\leq \frac{2}{\rho(Q_j \cap Q_{j+1})} \left( \int_{Q_j} |f_{Q_j} - f(x)|^2 \mathrm{d}\rho(x) + \int_{Q_{j+1}} |f_{Q_{j+1}} - f(x)|^2 \mathrm{d}\rho(x) \right) \\
&\leq 2C(a_{Q_j}^2 + a_{Q_{j+1}}^2).
\end{aligned}$$

Taking the square root and plugging into (3.21), we obtain

$$|f_Q - f_{Q_0}| \leq (2C)^{1/2} \sum_{j=0}^{N-1} a_{Q_j} + a_{Q_{j+1}} \leq (8C)^{1/2} \sum_{Q \subset B\widetilde{Q}} a_{\widetilde{Q}}$$

where the sum $\sum_{Q \subset B\widetilde{Q}}$ means that we sum over all cubes $\widetilde{Q} \in \mathcal{F}$ such that $Q \subset B\widetilde{Q}$. By the Boman chain condition, $Q_j$ and $Q_{j+1}$ have this property, and notice that we use here the fact that the elements of the Boman chain in Definition 3.2 are distinct.

Therefore,

$$\begin{aligned}
\rho(Q)|f_Q - f_{Q_0}|^2 \leq 8C\rho(Q)\Big( \sum_{Q \subset B\widetilde{Q}} a_{\widetilde{Q}} \Big)^2 &= 8C \int_Q \Big( \sum_{Q \subset B\widetilde{Q}} a_{\widetilde{Q}} \Big)^2 \mathrm{d}\rho(x) \\
&= 8C \int_Q \Big( \sum_{\widetilde{Q} \in \mathcal{F}} a_{\widetilde{Q}} \mathbf{1}_{B\widetilde{Q}}(x) \Big)^2 \mathrm{d}\rho(x)
\end{aligned}$$

since for any $x \in Q$, the sum in the first line is equal to the sum in the second line.

Then we use an important lemma which says that

$$\Big\| \sum_{\widetilde{Q} \in \mathcal{F}} a_{\widetilde{Q}} \mathbf{1}_{B\widetilde{Q}} \Big\|_{L^2(\rho)} \lesssim \Big\| \sum_{\widetilde{Q} \in \mathcal{F}} a_{\widetilde{Q}} \mathbf{1}_{\widetilde{Q}} \Big\|_{L^2(\rho)}. \tag{3.22}$$

Its proof is postponed slightly below.

All in all,

$$\begin{aligned}
\sum_{Q \in \mathcal{F}} \rho(Q)|f_Q - f_{Q_0}|^2 &\lesssim \int_{\mathcal{X}} \Big( \sum_{\widetilde{Q} \in \mathcal{F}} a_{\widetilde{Q}} \mathbf{1}_{B\widetilde{Q}}(x) \Big)^2 \mathrm{d}\rho(x) \lesssim \int_{\mathcal{X}} \Big( \sum_{\widetilde{Q} \in \mathcal{F}} a_{\widetilde{Q}} \mathbf{1}_{\widetilde{Q}}(x) \Big)^2 \mathrm{d}\rho(x) \\
&\lesssim \int_{\mathcal{X}} \sum_{\widetilde{Q} \in \mathcal{F}} a_{\widetilde{Q}}^2 \mathbf{1}_{\widetilde{Q}}(x) \mathrm{d}\rho(x) = \int_{\mathcal{X}} \sum_{\widetilde{Q} \in \mathcal{F}} \rho(\widetilde{Q}) a_{\widetilde{Q}}^2
\end{aligned}$$

where in the third inequality we used the Cauchy-Schwarz inequality and the first condition in Definition 3.2. Plugging into (3.20) we get the result. □

*Proof of* (3.22). Without loss of generality, we assume in the sequel that $B \geq 1$ (where $B$ appears in (3.12)). For $x \in \mathcal{X}$, denote by $\mathcal{F}_x$ the set of $Q \in \mathcal{F}$ for which $x \in Q$. For any $g \in L^1(\rho)$, set

$$Mg(x) = \sup_{Q \in \mathcal{F}_x} \frac{1}{\rho(BQ)} \int_{BQ} |g(y)| \mathrm{d}\rho(y).$$

We first prove that

$$\|Mg\|_{L^2(\rho)} \leq D' \|g\|_{L^2(\rho)} \tag{3.23}$$

for some $D' > 0$. To this end it is sufficient to prove that there exists $D > 0$ such that

$$\forall \alpha > 0, \quad \rho(S_\alpha) \leq D \alpha^{-1} \|g\|_{L^1(\rho)} \tag{3.24}$$

where $S_\alpha = \{x \in \mathcal{X} \mid Mg(x) > \alpha\}$. Indeed, (3.23) follows directly from (3.24), together with the $L^\infty(\rho)$-boundedness of $M$ and the Marcinkiewicz interpolation theorem [45, Theorem 9.1 in Chapter VIII.9].

Let us prove (3.24). For any $x \in S_\alpha$, let $Q_x \in \mathcal{F}_x$ such that

$$\int_{BQ_x} |g(y)| d\rho(y) \geq \alpha \rho(BQ_x).$$

From the standard Vitali covering argument, we get a countable subset $S'_\alpha \subset S_\alpha$ such that

$$S_\alpha \subset \bigcup_{x \in S'_\alpha} 5B\sqrt{d} Q_x,$$

and $BQ_x \cap BQ_{x'} = \emptyset$ for any distinct $x, x' \in S'_\alpha$. Using that $\rho$ is doubling (3.18), that $B \geq 1$ and the disjointness of the sets $BQ_x$, we obtain (with $D$ changing from one inequality to the other)

$$\begin{aligned}
\rho(S_\alpha) \leq D \sum_{x \in S'_\alpha} \rho(Q_x) \leq D \sum_{x \in S'_\alpha} \rho(BQ_x) &\leq \alpha^{-1} D \sum_{x \in S'_\alpha} \int_{BQ_x} |g(y)| \mathrm{d}\rho(y) \\
&\leq \alpha^{-1} D \|g\|_{L^1(\rho)}
\end{aligned}$$

which concludes the proof of (3.24) and (3.23).

We turn to the proof of (3.22). Recall that $a_Q \geq 0$ for any $Q \in \mathcal{F}$ (this is the only property of the sequence $(a_Q)_{Q \in \mathcal{F}}$ that we use below). Also, there holds $Mg(y) \geq \frac{1}{\rho(BQ)} \int_{BQ} |g(x)| \mathrm{d}\rho(x)$ for any $y \in Q$. Hence,

$$\begin{aligned}
\Big| \int_{\mathbb{R}^d} \sum_{Q \in \mathcal{F}} a_Q \chi_{BQ}(x) g(x) \mathrm{d}\rho(x) \Big| &\leq \sum_{Q \in \mathcal{F}} a_Q \rho(BQ) \frac{1}{\rho(BQ)} \int_{BQ} |g(x)| \mathrm{d}\rho(x) \\
&\leq \sum_{Q \in \mathcal{F}} a_Q \frac{\rho(BQ)}{\rho(Q)} \int_Q Mg(y) \mathrm{d}\rho(y) \\
&\leq C \sum_{Q \in \mathcal{F}} a_Q \int_Q Mg(y) \mathrm{d}\rho(y) \\
&= C \int_{\mathbb{R}^d} \sum_{Q \in \mathcal{F}} a_Q \chi_Q(y) Mg(y) \mathrm{d}\rho(y),
\end{aligned}$$

where we used again (3.18) to get the last inequality. Combining with (3.23), we get

$$\begin{aligned}\Big|\int_{\mathbb{R}^d}\sum_{Q\in\mathcal{F}} a_Q\chi_{BQ}(x)g(x)\mathrm{d}\rho(x)\Big| &\le C\Big\|\sum_{Q\in\mathcal{F}} a_Q\chi_Q\Big\|_{L^2(\rho)}\|Mg\|_{L^2(\rho)}\\ &\le CD'\Big\|\sum_{Q\in\mathcal{F}} a_Q\chi_Q\Big\|_{L^2(\rho)}\|g\|_{L^2(\rho)}\end{aligned}$$

which concludes the proof, by duality. □

## 3.4 The Cheeger inequality in spectral graph theory

Coming back to our initial goal of proving inequalities of the form

$$\mathrm{Var}_\rho(f) \le C\sum_{Q_i\in\mathcal{F}}\rho(Q_i)\mathrm{Var}_{\rho_{Q_i}}(f)$$

we develop in this section and the next one a second strategy. It is based on spectral graph theory, and notably on the Cheeger inequality which is an inequality between the spectral gap of a graph, i.e., the lowest eigenvalue of the associated Laplacian, and a geometric constant called the isoperimetric constant. We gather in this section general facts regarding Laplacians in infinite weighted graphs, and establish the appropriate version of the Cheeger inequality.

### 3.4.1 Weighted graphs

Let $V$ be a countable set equipped with the discrete topology, and endowed with a function $\delta : V \to (0,\infty)$. We denote by $\delta_i$ the value of $\delta$ at $i \in V$. The function $\delta$ can be turned into a Radon measure on $V$ of full support by the formula $\delta(U) = \sum_{i\in U}\delta_i$ for $U \subseteq V$. Next, let $w : V\times V \to [0,\infty)$, and denote by $w_{ij} = w(i,j)$ its values. We assume that it is symmetric ($w_{ij} = w_{ji}$), vanishing on the diagonal ($w_{ii} = 0$), and that it satisfies

$$\forall i\in V,\quad \sum_{j\in V} w_{ij} < +\infty.$$

We denote by $E$ its set of edges, i.e., the set of all $(i,j)\in V\times V$ such that $w_{ij} > 0$. If $(i,j)\in E$, we say that $i$ and $j$ are neighbors. When each vertex has only finitely many neighbors, we say that the weighted graph $(V,E,\delta,w)$ is locally finite.

### 3.4.2 Graph Laplacians

We denote by $C_c(V)$ the space of real valued functions on $V$ with finite support and consider the weighted $\ell^2$-space on vertices

$$\ell^2(V,\delta) = \Big\{u : V\to\mathbb{R} \mid \sum_{i\in V}\delta_i u(i)^2 < +\infty\Big\}.$$

We endow $\ell^2(V,\delta)$ with the scalar product $\langle u,v\rangle_\delta = \sum_{i\in V}\delta_i u(i)v(i)$ and denote by $\|u\|_\delta = \sqrt{\langle u,u\rangle_\delta}$ the corresponding norm. Let $\mathcal{Q} = \mathcal{Q}_w$ be the quadratic form with domain $\mathcal{D}$ given by

$$\mathcal{Q}(u) = \frac{1}{2}\sum_{i,j\in V} w_{ij}(u(i)-u(j))^2,\quad \mathcal{D} = \left\{u\in\ell^2(V,\delta) \mid \mathcal{Q}(u) < \infty\right\}.$$

This is a Dirichlet form, playing the role of an energy. The corresponding Laplacian is the positive selfadjoint operator $L$ acting as

$$Lu(i) = \frac{1}{\delta_i} \sum_{j \in V} w_{ij}(u(i) - u(j)). \tag{3.25}$$

Notice that $\mathcal{Q}(u) = \frac{1}{2}\langle Lu, u\rangle_\delta$. In the case that will be considered in the next section, there exists $C < +\infty$ such that

$$\forall i \in V, \quad C\delta_i \geq \sum_{j \in V} w_{ij}. \tag{3.26}$$

In particular, $L$ is a bounded operator. The function $\mathbf{1} \in \mathcal{D}$ equal to 1 on all $V$ is in the kernel of $L$, therefore we define the spectral gap of $L$ by

$$\lambda_2(L) = \inf\{\mathcal{Q}(u) \mid \|u\|_\delta = 1, \langle u, \mathbf{1}\rangle_\delta = 0\}.$$

It is non-negative, but not necessarily strictly positive, even if the graph is connected. The Cheeger inequality is a lower bound on $\lambda_2(L)$ in terms of a constant measuring how well the graph is connected, called the isoperimetric constant of $G$.

### 3.4.3 Isoperimetric constant and Cheeger inequality

For $U \subset V$ we denote its volume by

$$\mathrm{vol}(U) = \sum_{i \in U} \delta_i$$

and size of its boundary by

$$|\partial U| = \sum_{i \in U,\ j \notin U} w_{ij}. \tag{3.27}$$

The isoperimetric constant of $G$ is defined as

$$h = \inf_{\substack{U \subset V \\ 0 < \mathrm{vol}(U) \leq \frac{1}{2}}} \frac{|\partial U|}{\mathrm{vol}(U)}. \tag{3.28}$$

Equivalently, we may take the infimum over all sets $U$ (without the restriction on the volume), but in this case we need to replace the denominator $\mathrm{vol}(U)$ by $\min(\mathrm{vol}(U), \mathrm{vol}(V \setminus U))$. In other words, the isoperimetric constant is large if for any set $U \subset V$, the size of its boundary is non-negligible compared either to $\mathrm{vol}(U)$ or to $\mathrm{vol}(V \setminus U)$ (the latter case arises typically if $\mathrm{vol}(U)$ is too large).

In this context, the Cheeger inequality reads:

**Proposition 3.5** (Cheeger inequality in weighted graphs)**.** *If* (3.26) *holds, then*

$$\lambda_2(L) \geq \frac{h^2}{2C}$$

*where $C$ is the constant in* (3.26)*.*

The Cheeger inequality originates from the work of Cheeger [28], who proved an analogous inequality on manifolds.

*Proof of Proposition 3.5.* We provide an elementary proof assuming that $V$ is finite. A proof of Proposition 3.5 for infinite graphs may be found in [82, Theorem 3.5]. Let $n = |V|$. For $u \in \ell^2(V, \delta)$, we consider the Rayleigh quotient

$$R(u) = \frac{\mathcal{Q}(u)}{\|u\|_\delta^2}. \tag{3.29}$$

Let $\varepsilon > 0$ and $u : V \to \mathbb{R}$ such that $R(u) \le \lambda_2(L) + \varepsilon$ and $\langle u, \mathbf{1}\rangle_\delta = 0$. Up to relabelling the vertices, we may assume that

$$u(1) \ge \ldots \ge u(n).$$

Let $S_k = \{1, \ldots, k\}$ for $k \in [n]$, and

$$\alpha_G = \min_{k \in [n]} \frac{|\partial S_k|}{\min(\mathrm{vol}(S_k), \mathrm{vol}(V \setminus S_k))}.$$

Let $r \in [n]$ denote the largest integer such that $\mathrm{vol}(S_r) \le \frac{1}{2}\mathrm{vol}(G)$. Since $\sum_{i \in [n]} \delta_i u(i) = 0$,

$$\sum_{i \in [n]} \delta_i u(i)^2 = \min_{c \in \mathbb{R}} \sum_{i \in [n]} \delta_i (u(i) - c)^2 \le \sum_{i \in [n]} \delta_i (u(i) - u(r))^2$$

The positive part and the negative part of $u(i) - u(r)$, denoted by $u_+(i)$ and $u_-(i)$ respectively, are defined as follows:

$$u_+(i) = \begin{cases} u(i) - u(r) \text{ if } u(i) \ge u(r) \\ 0 \text{ otherwise} \end{cases} \qquad u_-(i) = \begin{cases} |u(i) - u(r)| \text{ if } u(i) \le u(r) \\ 0 \text{ otherwise.} \end{cases}$$

Then

$$\begin{aligned} R(u) &= \frac{\sum_{i,j} w_{ij}(u(i) - u(j))^2}{\sum_i \delta_i u(i)^2} \\ &\ge \frac{\sum_{i,j} w_{ij}(u(i) - u(j))^2}{\sum_i \delta_i (u(i) - u(r))^2} \\ &\ge \frac{\sum_{i,j} w_{ij}\left((u_+(i) - u_+(j))^2 + (u_-(i) - u_-(j))^2\right)}{\sum_i \delta_i (u_+(i)^2 + u_-(i)^2)}. \end{aligned}$$

Without loss of generality we have $R(u_+) \le R(u_-)$ and therefore $\lambda_2(L) + \varepsilon \ge R(u_+)$ since $\frac{a+c}{b+d} \ge \min(\frac{a}{c}, \frac{b}{d})$. If we assume $\lambda_2(L) + \varepsilon \ge R(u_-)$ instead, the subsequent computations can be carried out in the same way. Then we have

$$\begin{aligned} \lambda_2(L) + \varepsilon \ge R(u_+) &= \frac{\sum_{i,j} w_{ij}(u_+(i) - u_+(j))^2}{\sum_i \delta_i u_+(i)^2} \\ &= \frac{\sum_{i,j} w_{ij}(u_+(i) - u_+(j))^2}{\sum_i \delta_i u_+(i)^2} \cdot \frac{\sum_{i,j} w_{ij}(u_+(i) + u_+(j))^2}{\sum_{i,j} w_{ij}(u_+(i) + u_+(j))^2} && (3.30) \\ &\ge \frac{\left(\sum_{i,j} w_{ij}|u_+(i)^2 - u_+(j)^2|\right)^2}{2C\left(\sum_i \delta_i u_+(i)^2\right)^2} \text{ (see explanations below)} && (3.31) \\ &= \frac{\left(\sum_i (u_+(i)^2 - u_+(i+1)^2)|\partial S_i|\right)^2}{2C\left(\sum_i \delta_i u_+(i)^2\right)^2} && (3.32) \end{aligned}$$

where in the last line $|\partial S| = \sum_{k\in S,\ell\notin S} w_{k\ell}$. To go from (3.30) to (3.31) we apply the Cauchy-Schwarz inequality for the numerator; for the denominator we use (3.26). To go from (3.31) to (3.32) we observe that in the numerator of (3.31) each edge $i \sim j$ contributes $w_{ij}|u_+(i) - u_+(j)|^2$ to the sum, while in the numerator of (3.32), each edge $i \sim j$ (with for instance $j > i$) is in the boundary of $\partial S_{i'}$ for $i' = i, \dots, j-1$, and thus contributes $w_{ij}\sum_{i'=i}^{j-1} u_+(i)^2 - u_+(i+1)^2$ which is exactly equal to $w_{ij}|u_+(i) - u_+(j)|^2$.

We finish our computations: we set $\mathrm{vol}'(S) = \min(\mathrm{vol}(S), \mathrm{vol}(G) - \mathrm{vol}(S))$, we have

$$\lambda_2(L) + \varepsilon \geq \frac{\left(\sum_i (u_+(i)^2 - u_+(i+1)^2)\alpha_G \mathrm{vol}'(S_i)\right)^2}{2C\left(\sum_i \delta_i u_+(i)^2\right)^2} \tag{3.33}$$
$$= \frac{\alpha_G^2}{2C}\frac{\left(\sum_i u_+(i)^2(\mathrm{vol}'(S_i) - \mathrm{vol}'(S_{i-1}))\right)^2}{\left(\sum_i \delta_i u_+(i)^2\right)^2}$$
$$= \frac{\alpha_G^2}{2C}$$

where (3.33) follows from (3.32) and the definition of $\alpha_G$, and in the last line we used the fact that $\mathrm{vol}(S_r) \leq \frac{1}{2}\mathrm{vol}(G)$. This being true for any $\varepsilon > 0$, we obtain $\lambda_2(L) \geq \frac{\alpha_G^2}{2C}$, which concludes the proof since $\alpha_G \geq h$. □

## 3.5 Strategy 2: Gluing variances with spectral graph theory

Armed with these elements of spectral theory, we will explain in this section the second strategy for proving an inequality of the type

$$\mathrm{Var}_\rho(f) \leq C \sum_{Q_i\in\mathcal{F}} \rho(Q_i)\mathrm{Var}_{\rho_{Q_i}}(f). \tag{3.34}$$

Assume that $\rho$ is a probability measure on a metric space $\mathcal{X}$, and that $\mathcal{F} = \{Q_i\}_{i\in V}$ is a countable family of subsets of $\mathcal{X}$ such that

$$\bigcup_{i\in V} Q_i = \mathrm{supp}(\rho).$$

We do not assume that the $Q_i$'s are cubes (and actually, neither that $\mathcal{X} \subset \mathbb{R}^d$). However, as in Section 3.1, we assume:

(2) there exists $A > 0$ such that any $Q_i \in \mathcal{F}$ intersects at most $A$ other subsets $Q_j \in \mathcal{F}$ (including itself).

We construct the following graph: its vertices are given by the set $V$, and there is an edge between $i, j \in V$ if and only if $\rho(Q_i \cap Q_j) > 0$, in which case we write $i \sim j$. Each vertex $i \in V$ is endowed with a weight $\delta_i = \rho(Q_i)$, and each edge $(i, j)$ with a weigth $w_{ij} = \rho(Q_i \cap Q_j)$. We do not repeat the definitions of Section 3.4: in the sequel we consider the weighted $\ell^2$-space $\ell^2(V, \delta)$, the scalar product $\langle u, v\rangle_\delta$, the corresponding norm $\|\cdot\|_\delta$, the quadratic form $\mathcal{Q}$ with domain $\mathcal{D}$, and the Laplacian

$$Lu(i) = \frac{1}{\delta_i}\sum_{j\sim i} w_{ij}(u(i) - u(j)).$$

Due to (2), we know that for any $i \in V$,

$$\sum_{i\sim j} w_{ij} \leq A\delta_i. \tag{3.35}$$

We also have $\sum_{i\in V}\delta_i < +\infty$ since $\rho(\mathcal{X}) = 1$; hence the constant function $\mathbf{1} : V \to \mathbb{R}$ belongs to $\ell^2(V, \delta)$.

The next lemma, which does not assume anything on $\rho$, is the sought inequality of the form (3.34). Notice that it is useless if $\lambda_2(L) = 0$, e.g., if the graph is not connected.

**Lemma 3.6** (Gluing variances with spectral graph theory)**.** *Let $\rho_{Q_i} = \frac{1}{\rho(Q_i)}\rho_{|Q_i}$ for any $i \in V$. Then, for any continuous function $f : \mathbb{R}^d \to \mathbb{R}$, there holds*

$$\mathrm{Var}_\rho(f) \le A\left(1 + \frac{2A}{\lambda_2(L)}\right)\sum_{i\in V}\rho(Q_i)\mathrm{Var}_{\rho_{Q_i}}(f).$$

*Proof of Lemma 3.6.* Let $m_i = \int_{Q_i} f\mathrm{d}\rho_{Q_i}$ be the mean of $f$ over $Q_i$, and let $S = \sum_{i\in V}\delta_i \le A$. Developing the variance $\mathrm{Var}_\rho(f)$ thanks to the identity $f(x) - f(y) = f(x) - m_i + m_i - m_j + m_j - f(y)$, we get

$$\mathrm{Var}_\rho(f) \le S\sum_{i\in V}\delta_i\mathrm{Var}_{\rho_{Q_i}}(f) + \frac{1}{2}\sum_{i,j\in V}(m_i - m_j)^2\delta_i\delta_j. \tag{3.36}$$

We only need to find an upper bound on the second term in the right-hand side. We observe that $\langle m - \widetilde{m}, \mathbf{1}\rangle_\delta = 0$ where $\widetilde{m} = S^{-1}\sum_{i\in V}\delta_i m_i$ is "the mean of the means". Hence by definition of the spectral gap,

$$\begin{aligned}\frac{1}{2}\sum_{i,j\in V}(m_i - m_j)^2\delta_i\delta_j = S\|m - \widetilde{m}\|_\delta^2 &\le \frac{S}{\lambda_2(L)}\langle m - \widetilde{m}, L(m - \widetilde{m})\rangle_\delta \\ &= \frac{S}{2\lambda_2(L)}\sum_{i,j\in V}w_{ij}(m_i - m_j)^2. \end{aligned}\tag{3.37}$$

For any $i \neq j$ such that $w_{ij} > 0$, consider the mean $m_{i\cap j} = \frac{1}{w_{ij}}\int_{Q_i\cap Q_j} f\mathrm{d}\rho$ of $f$ over $Q_i \cap Q_j$. There holds

$$\frac{1}{2}(m_i - m_j)^2 \le (m_{i\cap j} - m_i)^2 + (m_{i\cap j} - m_j)^2.$$

For any such $i, j \in V$,

$$\begin{aligned}(m_{i\cap j} - m_i)^2 &= \left(\frac{1}{\rho(Q_i\cap Q_j)}\int_{Q_i\cap Q_j}(f - m_i)\mathrm{d}\rho\right)^2 \le \frac{1}{\rho(Q_i\cap Q_j)}\int_{Q_i\cap Q_j}(f - m_i)^2\mathrm{d}\rho \\ &\le \frac{1}{\rho(Q_i\cap Q_j)}\int_{Q_i}(f - m_i)^2\mathrm{d}\rho = \frac{\rho(Q_i)}{w_{ij}}\mathrm{Var}_{\rho_{Q_i}}(f),\end{aligned}$$

and similarly for $(m_{i\cap j} - m_j)^2$. We deduce

$$\frac{1}{2}(m_i - m_j)^2 \le \frac{\delta_i}{w_{ij}}\mathrm{Var}_{\rho_{Q_i}}(f) + \frac{\delta_j}{w_{ij}}\mathrm{Var}_{\rho_{Q_j}}(f).$$

Plugging into (3.37) we get

$$\begin{aligned}\frac{1}{2}\sum_{i,j\in V}(m_i - m_j)^2\delta_i\delta_j &\le \frac{S}{\lambda_2(L)}\sum_{i\in V}\sum_{j|Q_i\cap Q_j\neq\emptyset}(\delta_i\mathrm{Var}_{\rho_{Q_i}}(f) + \delta_j\mathrm{Var}_{\rho_{Q_j}}(f)) \\ &\le \frac{2AS}{\lambda_2(L)}\sum_{i\in V}\delta_i\mathrm{Var}_{\rho_{Q_i}}(f).\end{aligned}$$

Together with (3.36) and recalling $S \le A$, this concludes the proof of Lemma 3.6. □

In applications of Lemma 3.6 to concrete cases, with explicit $\rho$ and $\mathcal{F}$, one needs to prove $\lambda_2(L) > 0$ (otherwise the bound is trivial!). In general, proving a spectral gap for an operator can be a hard task. Here, it seems natural to rely on the Cheeger inequality (Proposition 3.5):

$$\lambda_2(L) \geq \frac{h^2}{2A}.$$

Depending on $\rho$, on the construction of the family $\mathcal{F}$, and thus on the corresponding graph, we need to check in concrete applications whether $h > 0$ or not. If $h > 0$, then $\lambda_2(L) > 0$ and Lemma 3.6 together with Section 3.1.1 can be used to prove the stability of Kantorovich potentials. If $h = 0$, we cannot draw any conclusion from Cheeger's inequality.

One example where we have been able to construct $\mathcal{F}$, to check that $h > 0$, and to conclude that Kantorovich potentials are stable, is the family of polynomially decaying probability densities on $\mathbb{R}^d$:

**Theorem 3.7** (Stability of Kantorovich potentials for polynomially decaying densities)**.** *Assume that $\rho(x) = c(1 + |x|)^{-\beta}$ on $\mathbb{R}^d$, with $\beta > d + 2$ and $c > 0$ a normalizing constant. Let $\mathcal{Y} \subset \mathbb{R}^d$ be a compact set. Then, there exists $C > 0$ such that for any probability measures $\mu, \nu$ supported in $\mathcal{Y}$,*

$$\|\phi_\mu - \phi_\nu\|_{L^2(\rho)} \leq CW_1(\mu, \nu)^\theta$$
$$\|T_\mu - T_\nu\|_{L^2(\rho)} \leq CW_1(\mu, \nu)^{\theta'}$$

*where $\theta = \frac{1}{2}(1 - \frac{2}{\beta - d}) > 0$ and $\theta' = \frac{\beta - d - 2}{8\beta - 2d - 4} > 0$. Moreover, the exponent $\theta$ in the first bound is sharp.*

Due to the radial symmetry of these distributions, we have constructed the family $\mathcal{F}$ in a way that each set $Q$, except a central one, is the intersection of an annulus and an angular sector (see Figure 5). The associated graph is very simple, it is essentially the union of $2^d$ line graphs (see Figure 5). The ratios $|\partial U|/\mathrm{vol}(U)$ can be lower bounded "by hand", but we shall not detail the computations here, we refer to [72] for details.

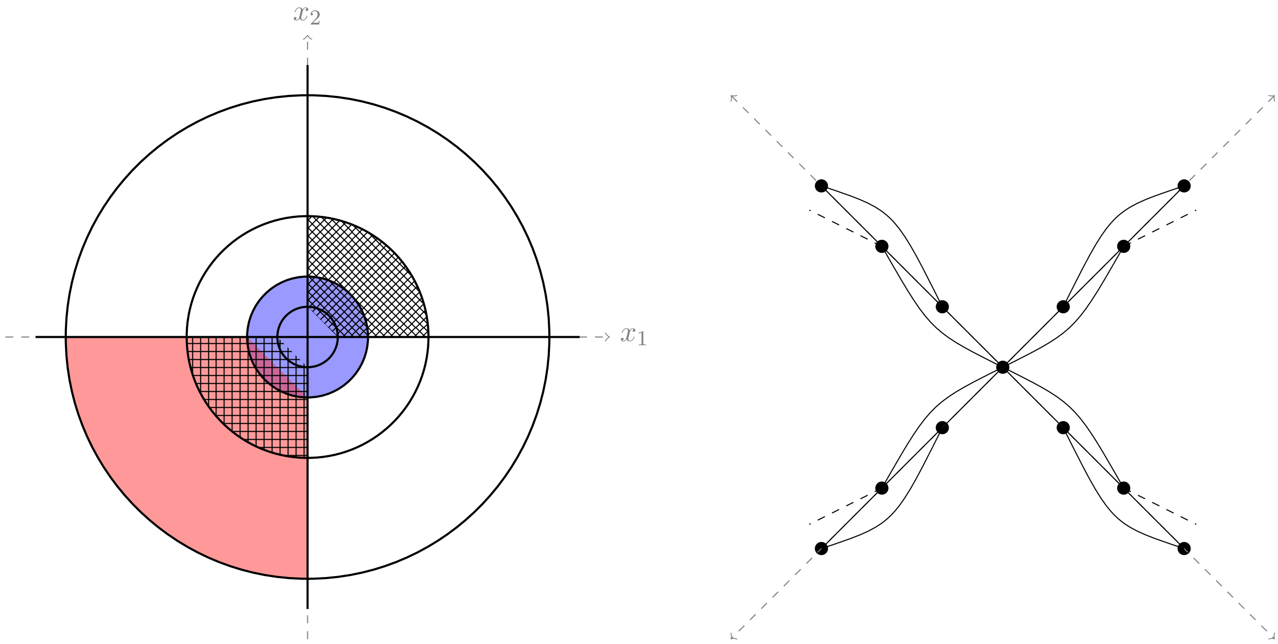



Figure 5: Here, $\rho$ is radially symmetric, for instance $\rho(x) = c(1 + |x|)^{-\beta}$. On the left, a few examples of sets $Q$ (in red, in blue with a grid and with crosshatches). The central set $Q$ in blue is different from the other ones, it covers the full unit disk, to make the graph on the right connected. On the right, the associated graph in a neighborhood of its central point.

### 3.6 Comments related to Poincaré inequalities

**Stability of Kantorovich potentials and Sobolev-Poincaré inequalities.** Let us make a few comments on the relation between the stability of Kantorovich potentials, and Sobolev-Poincaré inequalities. Maybe the easiest way to see a link between them is to write that if the stability inequality for Kantorovich potentials (1.16) (with $p = 2$) holds, then

$$\|\phi_\mu - \phi_\nu\|_{L^2(\rho)} \leq C W_2(\mu, \nu)^\alpha \leq C \|\nabla\phi_\mu - \nabla\phi_\nu\|_{L^2(\rho)}^\alpha$$

according to (1.15). Hence some kind of Poincaré inequality with an exponent holds, but only for differences of convex functions (with uniformly bounded gradient). We are not aware of any literature on this kind of inequalities.

Also, some part of our proof of Theorem 1.12 is shared with the proof of Sobolev-Poincaré inequalities in John domains. Let us recall the result:

**Theorem 3.8** (Sobolev-Poincaré inequality in John domains). *Let $f_\mathcal{X} = |\mathcal{X}|^{-1} \int_\mathcal{X} f \mathrm{d}x$. When $\mathcal{X} \subset \mathbb{R}^d$ is a John domain, the Sobolev-Poincaré inequality*

$$\left( \int_\mathcal{X} |f - f_\mathcal{X}|^{pd/(d-p)} \mathrm{d}x \right)^{(d-p)/pd} \leq C \left( \int_\mathcal{X} |\nabla f|^p \mathrm{d}x \right)^{1/p} \tag{3.38}$$

*holds for $1 \leq p < d$.*

Theorem 3.8 has been shown around 1985 by Bojarski, following ideas that he attributes to Boman, and our gluing techniques are adapted from this literature. We shall not demonstrate this theorem, but only the Poincaré inequality

$$\exists C_P > 0,\ \forall f \in C^1(\mathcal{X}), \qquad \mathrm{Var}(f) \leq C_P \int_\mathcal{X} |\nabla f|^2 \mathrm{d}x \tag{3.39}$$

when $\mathcal{X}$ is a John domain. Here the variance is taken with respect to the normalized Lebesgue measure $\rho = dx$ on $\mathcal{X}$. The Poincaré inequality (3.39) is weaker than (3.38), since it can be deduced from (3.38) (for $p = 2$, and $d \geq 3$) using Hölder's inequality.

To prove (3.39), we observe that there exists $C'_P > 0$ such that for any $Q \in \mathcal{F}$ there holds

$$\mathrm{Var}_{\rho_Q}(f) \leq C'_P \int_Q |\nabla f|^2 d\rho_Q. \tag{3.40}$$

This is because all $Q$ are cubes, with uniformly bounded diameter, and $\rho_Q$ is the normalized Lebesgue measure on $Q$. Summing (3.40) over $Q \in \mathcal{F}$ with weights $\rho(Q)$, and using (3.11) we get (3.39).

**Towards an equivalence in Theorem 1.12?** Theorem 3.8 can be strengthened into an equivalence, under an additional assumption. If $\mathcal{X} \subset \mathbb{R}^d$ is a domain of finite volume that satisfies a separation property, and $1 \leq p < d$, then

$$\mathcal{X} \text{ satisfies (3.38)} \Leftrightarrow \mathcal{X} \text{ is a John domain.} \tag{3.41}$$

The separation property, which we do not discuss here, is automatically valid for simply connected planar domains. And without an additional assumption on $\mathcal{X}$ such as the separation property, the equivalence (3.41) is not true. Let us illustrate this on an example: take $\mathcal{X} = \mathbb{D} \backslash E$ where $\mathbb{D}$ is the unit disk and $E = \bigcup_{k=1}^{\infty} E_k$ where $E_k$ consists of $k!$ equally spaced points on the circle $\{|x| = 1 - 2^{-k}\}$. Then $\mathcal{X}$ is not a John domain, but since $E$ is of dimension 0, the Sobolev-Poincaré inequality (3.38) holds in $\mathcal{X}$ (it can be deduced by integration by parts from the Sobolev-Poincaré inequality in $\mathbb{D}$).

This example may be transposed to the optimal transport setting with source measure $\rho$ equal to the uniform density on $\mathcal{X}$. Then optimal transport maps and potentials coincide with those obtained when the source measure is equal to the uniform probability density on $\mathbb{D}$. And for the latter, stability follows from Theorem 1.12. Therefore we have exhibited a non-John domain for which optimal transport stability inequalities hold. When trying to prove a converse statement to Theorem 1.12, one should keep this example in mind.

Several families of examples of bounded connected domains which are not John domains have been considered in the literature. For instance, domains with an outward cusp, and the so-called room-and-passage domains. In Section 3.7.3 we show that in these examples, stability of Kantorovich potentials fails, even in a very weak sense. This shows the relevance of the John domain condition in Theorem 1.12, at least regarding stability of Kantorovich potentials.

## 3.7 Examples and counterexamples

In the previous sections, we did not discuss the sharpness of our results. The only example we provided, in Section 1.8, showed that for optimal transport *maps*, without further assumptions on the target measures, the inequality

$$\|T_\mu - T_\nu\|_{L^2(\rho)} \leq C W_1(\mu, \nu)^\alpha \tag{3.42}$$

fails for $\alpha > 1/2$: there exists no $C > 0$ such that for any $\mu, \nu$ supported in $\mathcal{Y} = \mathbb{D}^2$, (3.42) holds.

In this section, we discuss the sharpness of our results regarding stability of *Kantorovich potentials*, by providing explicit computations on carefully chosen examples. This allows us to show some kind of sharpness in two respects:

- we show that the stability exponents $\alpha = 1/2$ for Kantorovich potentials in (1.27) is sharp in the Gaussian case (but the additional log-factor is probably not sharp), and that the exponent 1/2 in (1.30) is asymptotically sharp as the dimension tends to $+\infty$.

- We show that in typical examples of domains $\mathcal{X}$ which are not John domains, no bound of the form (1.14) can hold when $\rho$ has a density bounded above and below on $\mathcal{X}$. This shows that our "John domain" assumption in Theorem 1.12 is truly meaningful.

The main idea which guides the design of our examples is the following: to test our quantitative stability estimates, it seems much easier to choose convex potentials $\phi_1, \phi_2$ and to compute the Wasserstein distance between $(\nabla\phi_1)_{\#}\rho$ and $(\nabla\phi_2)_{\#}\rho$, than to choose two measures $\mu, \nu$ and to try to compute the associated potentials $\phi_\mu, \phi_\nu$. Indeed, solving a given optimal transport problem is difficult, whereas pushing forward through gradients of convex functions provides us directly with optimal transport maps and Kantorovich potentials between pairs of probability measures.

However, we should warn the reader that for the stability of optimal transport *maps*, we unfortunately have no good example to test the sharpness of stability exponents beside that of Section 1.8. In particular, we do not know the optimal exponents in (1.28) and (1.31).

### 3.7.1 Asymptotic sharpness of exponent in the ball

When $\rho$ is the uniform density on the unit ball $B_d(0,1)$ of $\mathbb{R}^d$, Theorem 1.12 provides us with an inequality of the form

$$\|\phi_\mu - \phi_\nu\|_{L^2(\rho)} \le C_d W_1(\mu,\nu)^\alpha \tag{3.43}$$

with $\alpha = 1/2$. We show that the exponent $\alpha = 1/2$ is asymptotically sharp as $d \to +\infty$:

**Proposition 3.9.** *Let $\rho_d$ be the uniform density on the unit ball $B_d(0,1)$ of $\mathbb{R}^d$. If (3.43) holds for any $\mu, \nu$ supported in the unit ball, then*

$$\alpha \le \frac{d+2}{2d}$$

*(a quantity which tends to 1/2 as $d \to +\infty$).*

*Proof.* Denote by $\omega_d$ the Euclidean volume of the unit ball $B_d(0,1)$ of $\mathbb{R}^d$ and by $\sigma_{d-1}$ the Euclidean area of the unit sphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$. In particular,

$$\rho_d(x) = \frac{1}{\omega_d}\mathbf{1}_{B_d(0,1)}.$$

Consider for any $\varepsilon \in (0,1)$ the radial and convex functions

$$\phi_\varepsilon^{(1)}(x) = |x|, \qquad \phi_\varepsilon^{(2)}(x) = \max(|x|, \varepsilon).$$

Then

$$\int_{B_d(0,1)} \left(\phi_\varepsilon^{(2)} - \phi_\varepsilon^{(1)}\right) \mathrm{d}\rho = \frac{\sigma_{d-1}}{\omega_d}\int_0^\varepsilon r^{d-1}(\varepsilon - r)\mathrm{d}r = \frac{\varepsilon^{d+1}}{d+1}$$

and

$$\int_{B_d(0,1)} \left(\phi_\varepsilon^{(2)} - \phi_\varepsilon^{(1)}\right)^2 \mathrm{d}\rho = \frac{\sigma_{d-1}}{\omega_d}\int_0^\varepsilon r^{d-1}(\varepsilon - r)^2\mathrm{d}r = \frac{2\varepsilon^{d+2}}{(d+1)(d+2)}.$$

Hence,

$$\mathrm{Var}(\phi_\varepsilon^{(2)} - \phi_\varepsilon^{(1)})^{1/2} \sim C_d \varepsilon^{(d+2)/2} \tag{3.44}$$

as $\varepsilon \to 0$, with $C_d = (2/(d+1)(d+2))^{1/2}$. Finally, denoting by $\delta_{\mathbb{S}^{d-1}}$ the uniform probability measure on $\mathbb{S}^{d-1}$,

$$(\nabla\phi_\varepsilon^{(1)})_\# \rho_d = \delta_{\mathbb{S}^{d-1}}, \qquad (\nabla\phi_\varepsilon^{(2)})_\# \rho_d = (1-\varepsilon^d)\delta_{\mathbb{S}^{d-1}} + \varepsilon^d \delta_0$$

hence

$$W_1((\nabla\phi_\varepsilon^{(1)})_\# \rho_d, (\nabla\phi_\varepsilon^{(2)})_\# \rho_d) = \varepsilon^d.$$

We conclude that for any $d$,

$$\mathrm{Var}(\phi_\varepsilon^{(2)} - \phi_\varepsilon^{(1)})^{1/2} \underset{\varepsilon\to 0}{\sim} C_d W_1((\nabla\phi_\varepsilon^{(1)})_\# \rho_d, (\nabla\phi_\varepsilon^{(2)})_\# \rho_d)^{(d+2)/2d}$$

i.e., it is necessary that $\alpha \le (d+2)/2d$ in order for (3.43) to be true. □

### 3.7.2 (Almost) sharpness of exponent for Gaussians

It seems natural to test the sharpness of our exponents in the Gaussian case too. Let

$$\rho(x) = (2\pi)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2}} \tag{3.45}$$

be the standard Gaussian. In this case, recall that the stability inequality for Kantorovich potentials is given by (1.27):

$$\|\phi_\mu - \phi_\nu\|_{L^2(\rho)} \le C W_1(\mu,\nu)^{1/2}(1 + |\log W_1(\mu,\nu)|^{1/2}).$$

The following proposition shows that this bound is sharp, up to the log factor.

**Proposition 3.10.** *Let $\rho$ be the standard Gaussian given by* (3.45). *If an inequality of the form*

$$\|\phi_\mu - \phi_\nu\|_{L^2(\rho)} \le C_d W_1(\mu,\nu)^\alpha$$

*holds for any $\mu, \nu$ supported in the unit ball, then $\alpha \le 1/2$.*

Actually, we prove something slightly stronger: if

$$\|\phi_\mu - \phi_\nu\|_{L^2(\rho)} \le C_d W_1(\mu,\nu)^{1/2}(1 + |\log W_1(\mu,\nu)|^\beta)$$

holds, then $\beta \ge -1$.

*Proof of Proposition 3.10.* Consider for any $r \in (0, +\infty)$ the radial and convex function

$$\phi_r(x) = (|x| - r)_+ - c_r \tag{3.46}$$

with $c_r$ chosen in a way that $\int_{\mathbb{R}^d} \phi_r(x) d\rho(x) = 0$. Brenier's theorem guarantees that $\nabla\phi_r$ is the optimal transport map from $\rho$ to $(\nabla\phi_r)_\#\rho$. For $r$ and $r'$ close enough (and $r$ large enough), we compare $\|\phi_r - \phi_{r'}\|_{L^2(\rho)}$ to $W_1(\mu,\nu)$ where $\mu = (\nabla\phi_r)_\#\rho$ and $\nu = (\nabla\phi_{r'})_\#\rho$.

For $r$ large, we set $r' = r + \frac{1}{r}$ and compute

$$\begin{aligned}
(2\pi)^{d/2}(c_r - c_{r'}) &= \int_r^{+\infty} (s-r)s^{d-1}e^{-s^2/2}ds - \int_{r'}^{+\infty} (s-r')s^{d-1}e^{-s^2/2}ds \\
&= (r'-r)\int_{r'}^{+\infty} s^{d-1}e^{-s^2/2}ds + \int_r^{r'} (s-r)s^{d-1}e^{-s^2/2}ds \\
&= O(r^{d-3}e^{-r^2/2})
\end{aligned}$$

and

$$\|(|\cdot|-r)_+-(|\cdot|-r')_+\|^2_{L^2(\rho)}=(r'-r)^2\int_{r'}^{+\infty}s^{d-1}e^{-s^2/2}ds+\int_r^{r'}(s-r)^2s^{d-1}e^{-s^2/2}ds$$
$$=\Theta(r^{d-4}e^{-r^2/2})$$

where we write $f(r)=\Theta(g(r))$ if the quotient $f(r)/g(r)$ remains bounded above and below by positive constants as $r\to+\infty$. We deduce that

$$\|\phi_{r'}-\phi_r\|^2_{L^2(\rho)}=\|(|\cdot|-r)_+-(|\cdot|-r')_+\|^2_{L^2(\rho)}-|c_r-c_{r'}|^2=\Theta(r^{d-4}e^{-r^2/2}). \tag{3.47}$$

We then turn to the computation of $W_1((\nabla\phi_{r'})_\#\rho,(\nabla\phi_r)_\#\rho)$. We observe that

$$(\nabla\phi_r)_\#\rho=\rho(B(0,r))\delta_0+(1-\rho(B(0,r))\sigma_{\mathbb{S}^{d-1}}$$

where $\sigma_{\mathbb{S}^{d-1}}$ is the uniform probability measure on $\mathbb{S}^{d-1}$. We have an analogous expression for $(\nabla\phi_{r'})_\#\rho$, and we deduce

$$W_1((\nabla\phi_{r'})_\#\rho,(\nabla\phi_r)_\#\rho)=\rho(B(0,r'))-\rho(B(0,r))=\Theta(r^{d-2}e^{-r^2/2}). \tag{3.48}$$

It follows from (3.47) and (3.48) that

$$\|\phi_{r'}-\phi_r\|_{L^2(\rho)}=\Theta(W_1^{1/2}|\log W_1|^{-1})$$

where $W_1$ is a short notation for $W_1((\nabla\phi_{r'})_\#\rho,(\nabla\phi_r)_\#\rho)$. □

We also observe that the above example does not prove the sharpness of the exponent of stability of optimal transport maps (1.28) (and indeed, we conjecture that the correct exponent is 1/2 and not 1/6).

The above proof can be adapted to other contexts. For instance, when $\rho(x)=c_{\beta,d}(1+|x|)^{-\beta}$ ($\beta>d+2$), it is possible to derive sharp stability exponents for Kantorovich potentials using the same family of radial Kantorovich potentials (3.46). Also, when $\rho$ blows up at the boundary of a ball or is the spherical uniform distribution (see "Degenerate densities $\rho$ in bounded domains" in Section 1.7), this same family may be used to find upper bounds on the stability exponents for Kantorovich potentials.

### 3.7.3 Strong instability for room-and-passage domains

We turn to another explicit computation, this time aimed at showing the relevance of the John domain condition in Theorem 1.12. For this, we consider domains that are considered in the literature as typical instances of non-John domains, and show that if $\rho$ is bounded above and below on such domain, then stability of Kantorovich potentials cannot hold, even in a very weak sense.

We could seek for even stronger, and hope that the John domain condition is necessary and sufficient for Theorem 1.12 to hold. However, this cannot be true, as explained in Section 3.6. In analogy, John domains support Sobolev-Poincaré inequalities, but there exist non-John domains which also support Sobolev-Poincaré inequalities. To remedy this issue, it has been shown that a domain satisfying a certain separation property supports Sobolev-Poincaré inequalities if and only if it is a John domain. The proof of this fact is delicate, and it would be interesting to look for an analogous converse result to Theorem 1.12.

In this section we prove:

**Theorem 3.11** (Instability of Kantorovich potentials in room-and-passage domains)**.** *There exists a non-empty, bounded, path-connected domain $\mathcal{X} \subset \mathbb{R}^d$ such that for any probability density $\rho$ bounded above and below on $\mathcal{X}$ the inequality*

$$\forall \mu, \nu \in \mathcal{P}(B(0,1)), \qquad \|\phi_\mu - \phi_\nu\|_{L^2(\rho)} \leq C W_p(\mu,\nu)^\alpha \tag{3.49}$$

*fails for any $C, \alpha > 0$ and $p \in [1, +\infty)$ (where $\phi_\mu, \phi_\nu$ denote the Kantorovich potentials between $\rho$ and $\mu$ and $\rho$ and $\nu$ respectively).*

The counterexample $\mathcal{X}$ is a so-called "room-and-passage" domain, a typical example of a non-John domain. It is endowed with a probability density which is bounded above and below on the support, for instance the uniform density. We consider the case $d = 2$ for simplicity, but the computations may be modified to cover any dimension $d$.

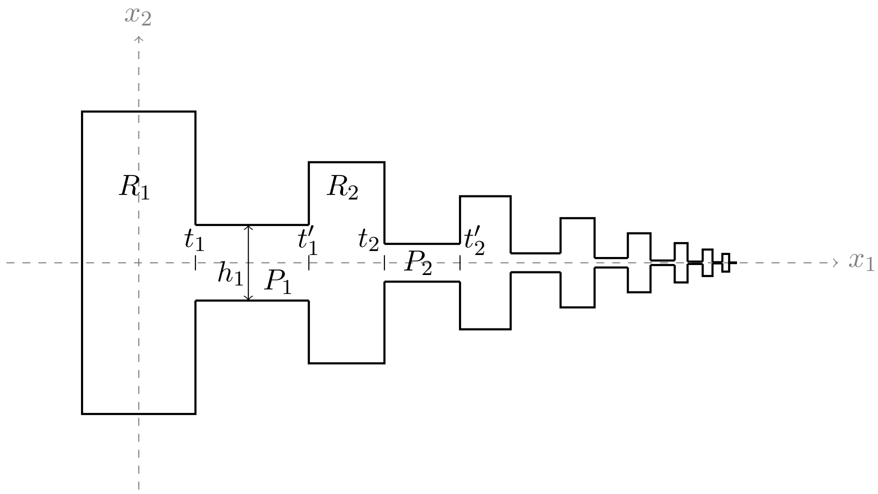



Figure 6: A room-and-passage domain

*Proof of Theorem 3.11.* As depicted on Figure 6, a room-and-passage domain in $\mathbb{R}^2$ is a connected and bounded set made of an infinite union of rectangles with variable lengths and widths. For simplicity, we assume that the axes of these rectangles are parallel to the coordinate axes. We call length of a rectangle the length of its side parallel to the $x_1$ axis, and width that of its side parallel to the $x_2$ axis. The rectangles are of two types, which alternate along the $x_1$-axis: the rooms $R_n$, $n \in \mathbb{N}$; and the passages $P_n$, $n \in \mathbb{N}$. The key assumption we make is that the passages have a width $h_n$ which decreases very fast as $n$ tends to $+\infty$, much faster than the other typical lengths of $R_n$ and $P_n$. To start, we keep $h_n$ free, as well as the other parameters of the rectangles, but we shall fix them later.

We write $P_n = [t_n, t'_n] \times [-h_n/2, h_n/2]$, and set

$$\phi_n(x) = |x_1 - t_n|, \qquad \phi'_n(x) = |x_1 - t'_n| \tag{3.50}$$

for $x = (x_1, x_2) \in \mathbb{R}^2$. Since $\phi_n$ (resp. $\phi'_n$) is convex, it differs from the Kantorovich potential from $\rho$ to $(\nabla\phi_n)_\#\rho$ (resp. $(\nabla\phi'_n)_\#\rho$) only by a constant. Now, the idea is that $\nabla\phi_n$ and $\nabla\phi'_n$ coincide on $\mathcal{X} \setminus P_n$, and this set has $\rho$-volume almost 1, which makes $(\nabla\phi_n)_\#\rho$ and $(\nabla\phi'_n)_\#\rho$ extremely close in Wasserstein distance: their Wasserstein distance is proportional to $\rho(P_n)$ which is of order $h_n(t'_n - t_n)$. The quantity $\mathrm{Var}(\phi_n - \phi'_n)$ is much larger (but very small too!) since $|\phi_n - \phi'_n|$ is equal to $|t'_n - t_n|$ in the largest part of $\mathcal{X}$.

More precisely, both $\mu_n = (\nabla\phi_n)_\#\rho$ and $\mu_n' = (\nabla\phi_n')_\#\rho$ are supported on $\{A, B\}$ where $A = (-1, 0)$ and $B = (1, 0)$, and the subset of points of $\mathcal{X}$ such that $\nabla\phi_n \neq \nabla\phi_n'$ is $P_n$. Since $\mathrm{dist}(A, B) = 2$, we get that for any $p \geq 1$,

$$W_p(\mu_n, \mu_n') = 2\rho(P_n)^{1/p}. \tag{3.51}$$

We turn to the computation of $\mathrm{Var}_\rho(\phi_n - \phi_n')$. For this, we observe that

$$\phi_n'(x) - \phi_n(x) = \begin{cases} t_n' - t_n & \text{if } x_1 \leq t_n \\ t_n - t_n' & \text{if } x_1 \geq t_n' \end{cases} \tag{3.52}$$

and

$$|\phi_n(x) - \phi_n'(x)| \leq |t_n - t_n'| \text{ if } x \in P_n. \tag{3.53}$$

Therefore

$$\|\phi_n - \phi_n'\|_{L^2(\rho)}^2 \geq |t_n - t_n'|^2(1 - \rho(P_n)). \tag{3.54}$$

Then, we evaluate the mean of $\phi_n - \phi_n'$. We set

$$v_n = \rho(\{x \in \mathcal{X} \mid x_1 \leq t_n\}) \qquad \text{and} \qquad w_n = \rho(\{x \in \mathcal{X} \mid x_1 \geq t_n'\}).$$

Then $v_n \to 1$ and $w_n \to 0$ as $n \to +\infty$, and for any $n \in \mathbb{N}^*$,

$$v_n + w_n + \rho(P_n) = 1.$$

Therefore, for $n$ large enough, using (3.52) and (3.53),

$$\begin{aligned} 0 \leq \int_{\mathcal{X}} (\phi_n' - \phi_n)\mathrm{d}\rho &\leq (t_n' - t_n)v_n + (t_n - t_n')w_n + \rho(P_n)|t_n - t_n'| \\ &= (1 - 2w_n)(t_n' - t_n). \end{aligned} \tag{3.55}$$

We deduce that

$$\begin{aligned} \mathrm{Var}_\rho(\phi_n - \phi_n') &= \|\phi_n - \phi_n'\|_{L^2(\rho)}^2 - \left(\int_{\mathcal{X}} (\phi_n - \phi_n')\mathrm{d}\rho\right)^2 \\ &\geq |t_n - t_n'|^2(1 - \rho(P_n) - (1 - 2w_n)^2) \\ &= |t_n - t_n'|^2(4w_n - \rho(P_n) - 4w_n^2) \\ &\geq |t_n - t_n'|^2(\rho(R_{n+1}) - \rho(P_n)) \end{aligned} \tag{3.56}$$

since $w_n \to 0$ and $w_n \geq \rho(R_{n+1})$.

There remains to choose the parameters of the rooms $R_n$ and the passages $P_n$. We choose $h_n$ small enough compared to all other lengths, in particular $\rho(R_{n+1}) - \rho(P_n) \geq \frac{1}{2}\rho(R_{n+1}) \gtrsim \lambda(R_{n+1})$ where recall that $\rho$ has density bounded below on $\mathcal{X}$. Then for any $\alpha > 0$ and $p \in [1, +\infty)$,

$$\frac{W_p(\mu_n, \mu_n')^\alpha}{\mathrm{Var}_\rho(\phi_n - \phi_n')} \lesssim \frac{\lambda(P_n)^{\alpha/p}}{|t_n - t_n'|^2\lambda(R_{n+1})} \lesssim \frac{|t_n - t_n'|^{(\alpha-2p)/p}h_n^{\alpha/p}}{\lambda(R_{n+1})}.$$

and we see that choosing $h_n$ small enough compared to all other parameters, this quantity tends to 0 as $n \to +\infty$, for any $\alpha > 0$, $p \in [1, +\infty)$, which concludes the proof. □

**Remark 3.12.** *The above computations being essentially* 1*-dimensional, one may easily turn them into an example of a source measure* $\rho$ *whose support is a segment of* $\mathbb{R}$*, and for which stability does not hold even in a very weak sense.*

To understand the general philosophy of the room-and-passage example, let us recall that if the support of a probability density $\rho$ is disconnected, Kantorovich potentials are not unique in general. Room-and-passage domains, while connected, are "nearly disconnected": at finer and finer scales, the passages that keep them connected become extremely narrow compared to the size of the pieces they connect. We can make a thought experiment: choose one of the passages, and send its width $h$ to 0. When $h = 0$, the domain is disconnected and potentials are not unique; when $h$ goes to 0, the Kantorovich potentials, while unique, become more and more unstable. This is to be expected: just before losing uniqueness, stability is lost.

Beside room-and-passage domains, domains with an outward cusp are another well-known category of non-John domains. By outward cusp, we mean that in some local coordinates, the equation defining the domain is $|y| \leq f(x)$ for some $f : [0, +\infty) \to \mathbb{R}$ with $f'(0) = 0$. For instance $f(x) = x^s$ with $s > 1$, or $f(x) = e^{-1/x^2}$. The former are called (polynomial) $s$-cusps, the latter exponential cusps. It is not difficult to modify the above computations to show that domains with an exponential cusp could also be used to prove Theorem 3.11.

Regarding polynomial cusps, we need another definition. For $s \geq 1$, an $s$-John domain is a domain for which the condition (1.29) is replaced by

$$\mathrm{dist}(\gamma(t), \mathcal{X}^c) \geq \eta t^s$$

i.e., the same condition as John domains except that $t$ is replaced by $t^s$ in the right-hand side. It follows that $s$-John domains may have polynomial $s$-cusps.

**Open question 3.13.** *Does a stability inequality $\|\phi_\mu - \phi_\nu\|_{L^2(\rho)} \leq C W_p(\mu, \nu)^\alpha$ hold in $s$-John domains, for some $C, p, \alpha$ which may depend on the domain (notably on $s$)?*

In $s$-John domains with large $s$, the exponent $\alpha$ necessarily has to be less good (i.e., larger) than 1/2. One can compute as an exercise an upper bound on the stability exponent in a domain containing an $s$-cusp, using similar sequences as in (3.50).

Let us hazard a final speculative comment: it seems to us that there is at least a formal resemblance between the example detailed in Section 3.7.3 and the Kannan-Lovasz-Simonovitz conjecture. This conjecture asserts that the Poincaré constant of log-concave measures can be checked on linear test functions. The analogy we see with our example is that we only need simple test functions (namely, distance functions to hyperplanes) to prove that optimal transport potentials are unstable. Therefore, it is tempting to formulate the following vague question:

**Open question 3.14.** *Is it true that for more general $\rho$'s, some simple family of test functions is sufficient to guarantee stability/instability of optimal transport potentials?*

## 3.8 Stability for quadratic optimal transport maps in $\mathbb{R}^d$

Until now, we have only proved quantitative stability estimates for Kantorovich potentials. To conclude this chapter, we establish the quantitative stability of optimal transport maps, under some assumptions. The arguments used here are very different from those used so far. Actually, the stability of optimal transport maps is deduced from that of Kantorovich potentials, using an additional final step: the main tool in this section will be an inequality of the form

$$\|\nabla f - \nabla g\|_{L^2(\rho)} \leq C L^{2/3} \|f - g\|_{L^2(\rho)}^{1/3} \tag{3.57}$$

for $f, g$ convex and $L$-Lipschitz on the support of $\rho$ (of course (3.57) cannot hold without some assumption on $f, g$). Applying it to $f = \phi_\mu$, $g = \phi_\nu$ and using the stability of Kantorovich potentials will immediately yield the stability of optimal transport maps in John domains (1.31).

To get stability of optimal transport maps when the source measure $\rho$ is log-concave, i.e., (1.28), one needs additional truncation arguments that we will not detail here.

The surprising inequality (3.57) crucially relies on the assumption that both $\phi_\mu$ and $\phi_\nu$ are convex. Its proof relies on two main ingredients:

- first, a one-dimensional version of the inequality (3.57);
- second, integral geometric techniques which allow to extend the one-dimensional inequality to higher dimensions.

Until now, this strategy has not led to optimal results, in the sense that the exponent 1/6 in Theorem 1.12 does not match the upper bound 1/2 on the exponent provided by the explicit example written in Section 1.8. Perhaps one would need a direct approach to stability of optimal transport maps, not going through stability of Kantorovich potentials, to get sharper exponents. This is of the main open questions in the field.

### 3.8.1 The 1d inequality

We start with a one-dimensional version of (3.57).

**Proposition 3.15.** *Let $I \subset \mathbb{R}$ be a compact segment. Let $u, v : I \to \mathbb{R}$ be two convex functions whose derivatives (defined a.e. on $I$) are uniformly bounded over $I$. Then*

$$\|u' - v'\|_{L^2(I)}^2 \leq 8(\|u'\|_{L^\infty(I)} + \|v'\|_{L^\infty(I)})^{4/3} \|u - v\|_{L^2(I)}^{2/3}. \tag{3.58}$$

This inequality looks like a Poincaré inequality, but in the wrong sense! It holds only because we are applying it to a difference of convex functions. Indeed, taking $I = [0, 1]$, $u = 0$ and $v = \sin(nx)$ shows that (3.58) cannot hold without assuming something on $u, v$. One can get some intuition about (3.58) by drawing the graphs of $u'$ and $v'$, which are non-decreasing functions. Then $u, v$ are obtained as areas under the curves and it may be seen that $|u' - v'|$ cannot be large on some quantitative fraction of $I$ without having $|u - v|$ large at some point. Another remark is that (3.58) is invariant under affine transformations, hence it is sufficient to prove the result on $I = [0, 1]$. Finally, the exponents in (3.58) are optimal, as may be seen by taking $u(x) = L|x - \frac{1}{2}|$ and $v = \max(u, \varepsilon)$.

*Proof of Proposition 3.15.* As explained above, we may assume $I = [0, 1]$ thanks to a scaling argument. First integrating by parts,

$$\begin{aligned}\int_0^1 |u' - v'|^2 &= \left[(u - v)(u' - v')\right]_0^1 - \int_0^1 (u - v)(u'' - v'') \\ &\leq 2\|u - v\|_{L^\infty}(\|u'\|_{L^\infty} + \|v'\|_{L^\infty}) + \|u - v\|_{L^\infty}\left(\int_0^1 |u''| + \int_0^1 |v''|\right)\end{aligned}$$

But since $u$ is convex,

$$\int_0^1 |u''| = \int_0^1 u'' = u'(1) - u'(0) \leq 2\|u'\|_{L^\infty},$$

and similarly for $v$, thus we conclude that

$$\int_0^1 |u' - v'|^2 \leq 4\|u - v\|_{L^\infty}(\|u'\|_{L^\infty} + \|v'\|_{L^\infty}). \tag{3.59}$$

The second step is to bound the $L^\infty$ norm of $f = u - v$ with its $L^2$-norm using that the Lipschitz constant of $f = u - v$ is less than $L = \|u'\|_{L^\infty} + \|v'\|_{L^\infty}$. This second step does not use the fact that $f$ is the difference of two convex functions: considering the worst case scenario where $f$ is piecewise affine, equal to 0 except around the maximum of $\|f\|_{L^\infty}$ where it looks like a "tent" with slope $L$, we get

$$\|u - v\|_{L^2(I)}^2 \geq \frac{1}{4} \min\left(\frac{\varepsilon}{2L}, 1\right) \varepsilon^2 \tag{3.60}$$

where $\varepsilon = \|u - v\|_{L^\infty}$.

We separate two cases. If $\varepsilon \geq 2L$, then we deduce from (3.60) that $\|u - v\|_{L^2(I)} \geq \frac{\varepsilon}{2}$. Hence

$$8(\|u'\|_{L^\infty(I)} + \|v'\|_{L^\infty(I)})^{4/3}\|u - v\|_{L^2(I)}^{2/3} \geq 8L^{4/3}(\varepsilon/2)^{2/3} \geq 8L^2 \geq \|u' - v'\|_{L^2}^2,$$

which concludes in this case. If $\varepsilon \leq 2L$, then (3.60) yields $\varepsilon^3 \leq 8L\|u - v\|_{L^2(I)}^2$, hence

$$\|u - v\|_{L^\infty(I)} \leq 2L^{1/3}\|u - v\|_{L^2(I)}^{2/3}$$

and plugging into (3.59) we also get (3.58). □

### 3.8.2 Higher dimension: an integral-geometric argument

The second step is to generalize Proposition 3.15 to higher dimensions:

**Proposition 3.16.** *Let $L > 0$ and let $K$ be a compact subset of $\mathbb{R}^d$ whose boundary has finite $(d-1)$-dimensional measure. Then there exists $C > 0$ such that for any $u, v : K \to \mathbb{R}$ convex on any segment included in $K$ and $L$-Lipschitz,*

$$\|\nabla u - \nabla v\|_{L^2(K)} \leq C\|u - v\|_{L^2(K)}^{1/3}.$$

*Proof.* To prove Proposition 3.16, one possibility is to rely on integral-geometric techniques, i.e., expressing a multidimensional integral in terms of integrals over lines (or geodesics, in Riemannian geometry). We start from the formula

$$\int_{\mathbb{R}^d} f(x)^2 \mathrm{d}x = \int_{e^\perp} \int_{\mathbb{R}} f(y + te)^2 \mathrm{d}t\mathrm{d}y.$$

valid for any $e \in \mathbb{S}^{d-1}$, where $e^\perp$ denotes the hyperplane (through the origin) perpendicular to the unit vector $e$. Applying this to $f(x) = \langle F(x), e\rangle$, and then integrating over $e \in \mathbb{S}^{d-1}$, we get

$$\int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}^d} \langle F(x), e\rangle^2 \mathrm{d}x\mathrm{d}\sigma(e) = \int_{\mathbb{S}^{d-1}} \int_{e^\perp} \int_{\mathbb{R}} \langle F(y + te), e\rangle^2 \mathrm{d}t\mathrm{d}y\mathrm{d}\sigma(e)$$

where $\sigma$ is the uniform probability measure on $\mathbb{S}^{d-1}$. We observe that the left-hand side is equal to $C_d\|F\|_{L^2(\mathbb{R}^d)}^2$ for some $C_d > 0$ depending only on $d$. We apply this to $F$ given by $\nabla u - \nabla v$ inside $K$, and extended by 0 outside $K$. We get

$$\|\nabla u - \nabla v\|_{L^2(K)}^2 = C_d^{-1} \int_{\mathbb{S}^{d-1}} \int_{e^\perp} \|u'_{\ell_e^y} - v'_{\ell_e^y}\|_{L^2(\ell_e^y \cap K)}^2 \mathrm{d}y\mathrm{d}\sigma(e) \tag{3.61}$$

where $\ell_e^y$ denotes the oriented line $y+e\mathbb{R}$ and $u_{\ell_e^y} = u|_{\ell_e^y\cap K}$, $v_{\ell_e^y} = v|_{\ell_e^y\cap K}$. The set $\ell_e^y \cap K$ may be decomposed as a finite union of intervals $I^i_{\ell_e^y}$, $i = 1, \dots, n_{\ell_e^y}$ in which we can apply Proposition 3.15 (the 1d inequality, using that $\|u'\|_{L^\infty}, \|v'\|_{L^\infty} \leq L$). We get

$$\|u'_{\ell_e^y} - v'_{\ell_e^y}\|^2_{L^2(\ell_e^y\cap K)} = \sum_{i=1}^{n_{\ell_e^y}} \|u'_{\ell_e^y} - v'_{\ell_e^y}\|^2_{L^2(I^i_{\ell_e^y})} \leq 8(2L)^{4/3} \sum_{i=1}^{n_{\ell_e^y}} \|u_{\ell_e^y} - v_{\ell_e^y}\|^{2/3}_{L^2(I^i_{\ell_e^y})}$$
$$\leq 8(2L)^{4/3} n_{\ell_e^y}^{2/3} \|u_{\ell_e^y} - v_{\ell_e^y}\|^{2/3}_{L^2(\ell_e^y\cap K)}$$

where the last inequality comes from Jensen's inequality. Combining this with (3.61) and then using Hölder's inequality, we get

$$\|\nabla u - \nabla v\|^2_{L^2(K)} \lesssim \int_{\mathbb{S}^{d-1}} \int_{e^\perp} n_{\ell_e^y}^{2/3} \|u_{\ell_e^y} - v_{\ell_e^y}\|^{2/3}_{L^2(\ell_e^y\cap K)} \mathrm{d}y \mathrm{d}\sigma(e)$$
$$\lesssim \left(\int_{\mathbb{S}^{d-1}} \int_{e^\perp} n_{\ell_e^y} \mathrm{d}y \mathrm{d}\sigma(e)\right)^{2/3} \left(\int_{\mathbb{S}^{d-1}} \int_{e^\perp} \|u_{\ell_e^y} - v_{\ell_e^y}\|^2_{L^2(\ell_e^y\cap K)} \mathrm{d}y \mathrm{d}\sigma(e)\right)^{1/3}.$$

The second parenthesis is equal to $C_d\|u - v\|^2_{L^2(K)}$ due to the same argument which led to (3.61). Regarding the first parenthesis, we observe that $n_{\ell_e^y} \leq \#(\ell_e^y \cap \partial K)$ and then we use the Cauchy-Crofton formula, which asserts that

$$\int_{\mathbb{S}^{d-1}} \int_{e^\perp} \#(\ell_e^y \cap \partial K) \mathrm{d}y \mathrm{d}\sigma(e) = \mathscr{H}^{d-1}(\partial K) < +\infty \tag{3.62}$$

where $\mathscr{H}^{d-1}$ denotes the $(d-1)$-dimensional Hausdorff measure. This concludes the proof of Proposition 3.16.

Here, we should warn the reader that for the Cauchy-Crofton formula (3.62) to hold, one actually needs to assume that the boundary $\partial K$ is rectifiable. If this is not assumed, then it is still true that the left-hand side in (3.62) is finite when $\mathscr{H}^{d-1}(\partial K) < +\infty$ (but the equality in (3.62) does not necessarily hold). To prove this and also to extend Theorem 1.12 to Riemannian manifolds, we devised a more robust argument based on the definition of the integral-geometric measure, with the help of Antoine Julia and Federer's book [49]. This outer measure, defined following Caratheodory's construction, compares easily (almost by definition) with the $(d-1)$-dimensional Hausdorff measure and may be shown to count the number of intersections of short geodesic curves with $\partial\mathcal{X}$. □

### 3.8.3 Conclusion of the proof of Theorem 1.12

To prove the stability of optimal transport maps in the John domain case, i.e., (1.31), we simply combine the stability of Kantorovich potentials (1.30) (proved in Section 3.3) with Proposition 3.16 for $K = \overline{\mathcal{X}}$.

We shall not detail the proof of the stability of optimal transport maps in the log-concave case (i.e., (1.28)). It relies not only on Proposition 3.16, but also on truncation arguments closely related to those of Section 2.6.

To conclude this chapter, we would like to ask the following question: where did we lose sharpness of the exponents? Indeed, as we already mentioned, there is a gap between the stability inequality for optimal transport maps (1.31), which displays an exponent 1/6, and the concrete examples, in which the optimal transport maps are always more stable than this and the stability exponent is 1/2 (see for instance Section 1.8). If we summarize our proofs, we have proved stability of optimal transport maps thanks to a chain of inequalities of the form

$$\|T_\mu - T_\nu\|^6_{L^2(\rho)} \lesssim \|\phi_\mu - \phi_\nu\|^2_{L^2(\rho)} \lesssim \langle \psi_\nu - \psi_\mu, \mu - \nu \rangle \lesssim W_1(\mu, \nu). \tag{3.63}$$

Indeed, the right-most inequality is the "Kantorovich-Rubinstein" argument done several times, the middle inequality has been proven using the strong convexity of the Kantorovich functional (together with gluing arguments), and the left-most inequality comes from Proposition 3.16.

As seen in Section 3.7.1 the two inequalities on the right become asymptotically sharp as $d \to +\infty$. The inequality on the left is sharp in any dimension: take $\rho = \mathbf{1}_{[0,1]^d}$ and

$$\phi_\varepsilon^{(1)}(x) = |x_1|, \qquad \phi_\varepsilon^{(2)}(x) = \max(\phi_\varepsilon^{(1)}, \varepsilon)$$

for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$. Then $\|\nabla\phi_\varepsilon^{(1)} - \nabla\phi_\varepsilon^{(2)}\|^2_{L^2(\rho)} = \varepsilon$, whereas $\mathrm{Var}(\phi_\varepsilon^{(1)} - \phi_\varepsilon^{(2)}) \approx \varepsilon^3$, showing the sharpness of the exponents in the left hand side inequality.

In other words, each of the three inequalities in (3.63) is sharp! But their equality cases do not match, and we have no example where this chain of inequalities is indeed saturated. Thus, one of the main open problems that remains to be solved is the following:

**Open question 3.17.** *Obtain sharp exponents for the stability of optimal transport maps (in* (1.28) *and* (1.31) *for instance).*

## 3.9 Bibliographical notes

§3.1: The starting point of our collaboration with Quentin Mérigot [72], in which we worked out the general gluing methods presented here, was the work [26] in which a basic gluing argument for finite families $\mathcal{F}$ was performed in the slightly different context of stability of Wasserstein barycenters. Our methods turned out to be useful to address stability of optimal transport in Riemannian manifolds too [64] (see Section 4.4). Other methods for decomposing measures and proving Poincaré inequalities (or other functional inequalities) exist in the literature, for instance Eldan's stochastic localization [48].

§3.2: The Whitney decomposition was discovered in 1934 by Hassler Whitney [108]. Boman introduced in [14] the chains now known as Boman chains. His goal was to prove $L^p$ estimates for solutions to some over-determined elliptic systems of PDEs in regions with irregular boundary. Bojarski discovered in [13] how to use these chains to prove Sobolev-Poincaré inequalities in John domains. The proof of Proposition 3.3 is due to him. The converse fact that any bounded open subset of $\mathbb{R}^d$ supporting Boman chains is a John domain was proved in [22].

§3.3: The proof of Proposition 3.4 is inspired by Bojarski's computations in [13].

§3.4: Spectral graph theory is a classical topic, see for instance the books [34], [99] and the beautiful expository notes by Luca Trevisan [103]. The Cheeger inequality in finite graphs is of course covered in these references, and our proof is adapted from [34, Theorem 2.2]. For infinite graphs, the book chapter [62, Chapter 13.1] is particularly clear. A more general Cheeger inequality, applying as well to infinite graphs which do not verify (3.26) has been proved in [62, Theorem 13.4]. Our presentation follows closely [10].

§3.5: I did not find any reference for the elementary Lemma 3.6, but it seems difficult to believe that no one ever used such arguments.

§3.6: The converse to Bojarski's result, namely the left-to-right implication in (3.41), was proved in the paper [21]. We borrowed our discussion about the separation property from this paper. Since the 1980's, many authors have been using variants of the Boman chain condition. It turns out that for many results, a good framework is that of metric spaces endowed with a doubling measure, see the memoir [55] which develops the theory of Sobolev spaces and proves Poincaré-type inequalities in this setting.

§3.7: The example in Section 3.7.1 comes from from [64], and those of Sections 3.7.2 and 3.7.3 come from [72]. Room-and-passage domains date back at least to the 1937 monograph by Courant and Hilbert [37, pp. 521-523], who used them to show that the embedding of $H^1(\mathcal{X})$ in $L^2(\mathcal{X})$ is not necessarily compact (see also [4]). Indeed, consider $\phi_n$ a function which is equal to a constant in the $n$-th room $R_n$ and which drops linearly to 0 in the adjacent passages $P_{n-1}$ and $P_n$, reaching the value 0 at the midpoint of each of these passages. Choosing the constant in a way that $\|\phi_n\|_{L^2(\mathcal{X})} = 1$, we obtain an orthonormal family of functions. If the passages are narrow enough (i.e., if $h_n$ is small enough), then $\|\nabla\phi_n\|_{L^2(\mathcal{X})} \to 0$ as $n \to +\infty$. Since no subsequence of $(\phi_n)$ converges in $L^2(\mathcal{X})$, $H^1(\mathcal{X})$ is not compactly embedded in $L^2(\mathcal{X})$. This example is fundamentally related to the fact that the Poincaré inequality fails in $\mathcal{X}$, and that 0 is in the essential spectrum of the Neumann Laplacian on $\mathcal{X}$. In this direction, Hempel-Seco-Simon [56] used room-and-passage domains to provide examples of domains with prescribed essential spectrum. Finally, many papers have considered $s$-John domains, see for instance [54].

§3.8: The one-dimensional inequality given in Proposition 3.15 was proved in [42]. This is a refinement of Theorem 3.5 in [27], in which the upper bound involved the uniform distance $\|u - v\|_{L^\infty}$. Federer's book [49] provides a general perspective on integral-geometric techniques, notably those of Section 3.8.2. As already mentioned, there is an alternative path to prove Proposition 3.16, which relies on the Caratheodory construction, see for instance [49, Chapter 2.10]. This alternative strategy was used in [64] to extend Theorem 1.12 to Riemannian manifolds.

# 4 Stability of optimal transport for more general costs

While the previous chapters were focused on *quadratic* optimal transport in *Euclidean spaces*, we turn in this chapter to more general costs. Our goal is to prove quantitative stability estimates for optimal transport maps and Kantorovich potentials obtained as solutions to optimal transport problems with more general costs.

We focus on the generalization of Theorem 1.12. To this end, we make a small detour in the world of entropic optimal transport (EOT): this penalized version of the optimal transport problem is both an important theoretical topic and a major computational tool. Section 4.1 is a brief self-contained introduction to this subject. In Section 4.2, inspired by EOT, we define the regularized Kantorovich functional, a variant of the Kantorovich functional. Finally, in Sections 4.3 and 4.4 we state the main results of this chapter, namely the quantitative stability bounds for the $p$-cost in $\mathbb{R}^d$ and for the squared distance on Riemannian manifolds. We discuss their proofs which rely on strong concavity estimates for the regularized Kantorovich functional.

Another possible direction would be to extend the validity of Gigli's Theorem 1.9, which shows stability "around regular transport maps". The generalization of Gigli's theorem has been achieved in the paper [3] for the squared distance on Riemannian manifolds, and in [52] for more general costs. For instance, the results of [52] apply to the so-called reflector cost

$$c(x, y) = -\log(1 - x \cdot y) \quad \text{on} \quad \mathbb{S}^{d-1}.$$

We do not discuss these developments here.

**Notation.** Recall that the duality pairing between continuous functions and real-valued Radon measures is denoted by $\langle \cdot \mid \cdot \rangle$. For instance, $\langle c \mid \gamma \rangle$ stands for the integral $\int_{\mathcal{X}\times\mathcal{Y}} c(x, y)\mathrm{d}\gamma(x, y)$.

## 4.1 Entropic optimal transport

Let $\mathcal{X}$ and $\mathcal{Y}$ be compact metric spaces. Given $\varepsilon > 0$, a cost function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ bounded from below, and two probability measures $\rho$ and $\mu$ on $\mathcal{X}$ and $\mathcal{Y}$ respectively, the *entropic optimal transport* problem reads

$$\inf_{\gamma\in\Pi(\rho,\mu)} \langle c \mid \gamma \rangle + \varepsilon H(\gamma \mid \rho \otimes \mu) \tag{4.1}$$

where $H$ denotes the *entropy*

$$H(\alpha \mid \beta) = \begin{cases} \int f(\log(f) - 1)\, \mathrm{d}\beta & \text{if } \alpha \text{ has a density w.r.t. } \beta\text{, denoted by } f \\ +\infty & \text{otherwise.} \end{cases} \tag{4.2}$$

Usually, the entropy is simply $f \log f$, but subtracting $f$ makes later computations slightly nicer (and when both $\alpha$ and $\beta$ are probability measures, subtracting $f$ in the integral simply amounts to subtracting a constant to the entropy). In (4.2), for short, we denote by $x(\log x - 1)$ the function equal to $x(\log x - 1)$ where $x > 0$, equal to 0 where $x = 0$, and equal to $+\infty$ where $x < 0$. For $\varepsilon = 0$, one recovers the usual quadratic optimal transport problem in its Kantorovich formulation (1.3). For $\varepsilon > 0$, the penalization term $\varepsilon H(\gamma \mid \rho \otimes \mu)$ forces the optimal transport plan to have some density with respect to the transport plan $\rho \otimes \mu$, which is the most naive transport plan between $\rho$ and $\mu$ since it distributes equally over $\mu$ any piece of mass of $\rho$.

Entropic optimal transport has the big advantage that (4.1) always has a *unique solution*, whereas recall that multiple solutions of the Kantorovich problem may exist.

**Proposition 4.1.** *Given $\varepsilon > 0$ and two probability measures $\rho$ and $\mu$, the entropic optimal transport problem* (4.1) *admits a unique solution.*

*Proof.* The existence follows from the lower semi-continuity of the functional

$$\gamma \mapsto \langle c \mid \gamma \rangle + \varepsilon H(\gamma \mid \rho \otimes \mu)$$

over the compact set $\Pi(\rho, \mu)$. The uniqueness is a consequence of the strict convexity of $H$, which comes from the strict convexity of $x(\log x - 1)$. □

The computation of the unique solution to (4.1) is more tractable from a computational point of view than that of the unregularized optimal transport problem (1.3), thanks to the *Sinkhorn algorithm* presented in Section 5.3, which has an exponential convergence rate under some assumptions on the marginals.

**Remark 4.2.** *Entropic optimal transport plans, i.e., minimizers of* (4.1)*, have full support with respect to* $\rho \otimes \mu$*. The proof of this fact is left to the reader: it follows from the infinite negative slope of* $f \log f$ *which allows to create a transport plan with strictly lower cost from any transport plan whose support is strictly contained in the support of* $\rho \otimes \mu$*. As a consequence, the unique solution to* (4.1) *is not a map in general (but it is a transport plan between* $\rho$ *and* $\mu$*).*

As for (1.3), there is an equivalent dual formulation of (4.1). It is based on the following formula:

**Proposition 4.3.** *Let* $\beta$ *be a finite (positive) Radon measure on a compact set* $Z$*, and let* $\alpha$ *be a real-valued Radon measure on* $Z$*. Then,*

$$H(\alpha \mid \beta) = \sup_{f \in C^0(Z)} \langle f \mid \alpha \rangle - \langle e^f \mid \beta \rangle. \tag{4.3}$$

*In particular,* $\alpha \mapsto H(\alpha \mid \beta)$ *is convex and weakly-* lower semi-continuous. In addition, the supremum in* (4.3) *is attained at* $f \in C^0(Z)$ *if and only if* $e^f$ *is the density of* $\alpha$ *with respect to* $\beta$*.*

**Remark 4.4.** *The above formula is quite close to the Donsker-Varadhan variational formula, which is stated in* (5.26)*. But to obtain the standard formulation of duality for entropic optimal transport,* (4.3) *is the correct identity to use.*

*Proof of Proposition 4.3.* Let $h : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ defined by $h(x) = x(\log x - 1)$ if $x > 0$, $h(0) = 0$ and $h(x) = +\infty$ for $x < 0$. Its Legendre transform is

$$h^*(s) = \sup_{x>0} xs - h(x) = e^s.$$

Assume first that $\alpha$ has a density $g$ with respect to $\beta$. Since $h$ is convex and lower semi-continuous, there holds $h = h^{**}$ by the Fenchel-Moreau theorem, hence

$$H(\alpha \mid \beta) = \int h(g(x))\mathrm{d}\beta(x) = \int h^{**}(g(x))\mathrm{d}\beta(x) = \int \sup_{s \in \mathbb{R}} (sg(x) - h^*(s))\mathrm{d}\beta(x).$$

In particular, for any bounded measurable function $f$, we have

$$H(\alpha \mid \beta) \geq \langle f \mid g\beta \rangle - \langle e^f \mid \beta \rangle = \langle f \mid \alpha \rangle - \langle e^f \mid \beta \rangle$$

with equality if $g = e^f$ almost everywhere. This proves (4.3) in this case.

Let us now assume that $\alpha$ does not have a density with respect to $\beta$, and let us prove that the supremum in the right-hand side of (4.3) is also equal to $+\infty$. Pick $A \subset Z$ such that $\beta(A) = 0$ and $\alpha(A) \neq 0$. Let $f = \lambda \mathbf{1}_A$ for some $\lambda \in \mathbb{R}$. Recall Lusin's theorem (see [96, Memo 1.6]), adapted for signed measures: if $f : Z \to \mathbb{R}$ is measurable, then for every $\varepsilon > 0$, there exists a compact set $K \subset Z$ and a continuous function $\widetilde{f} : Z \to \mathbb{R}$ such that $|\alpha(Z \setminus K)| < \varepsilon$, and $f = \widetilde{f}$ on $K$, and $\|\widetilde{f}\|_{L^\infty(Z)}$ is bounded above independently of $\varepsilon$. Taking for $f_n$ the function $\widetilde{f}$ associated to a sequence $\varepsilon_n \to 0$, we get

$$\langle f_n \mid \alpha \rangle - \langle e^{f_n} \mid \beta \rangle \underset{n\to+\infty}{\longrightarrow} \langle f \mid \alpha \rangle - \langle e^f \mid \beta \rangle. \tag{4.4}$$

Finally, letting $\lambda \to \pm\infty$, we get that (4.3) holds also when $\alpha$ is not absolutely continuous with respect to $\beta$.

The strict convexity of $\alpha \mapsto H(\alpha \mid \beta)$ is an immediate consequence of the strict convexity of $x \mapsto x(\log(x) - 1)$. □

**Dual formulation.** To establish the dual formulation of (4.1) we first remove the positivity constraint on $\gamma$ in (4.1), which is guaranteed to be satisfied for minimizers due to the definition of the entropy: in other words, we denote by $\Pi'(\rho, \mu)$ the set of *signed* measures $\gamma$ on $\mathcal{X} \times \mathcal{Y}$ such that for all measurable sets $A \subset \mathcal{X}$, $B \subset \mathcal{Y}$,

$$\gamma(A \times \mathcal{Y}) = \rho(A) \qquad \text{and} \qquad \gamma(\mathcal{X} \times B) = \mu(B).$$

and we consider

$$\inf_{\gamma\in\Pi'(\rho,\mu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y)\mathrm{d}\gamma(x, y) + \varepsilon H(\gamma \mid \rho \otimes \mu). \tag{4.5}$$

Although the infimum is taken over a larger set in (4.5) than in (4.1), these two infima are equal. Indeed, if $\gamma$ is *not* non-negative, then we can check that $H(\gamma \mid \rho \otimes \mu) = +\infty$: taking $g$ such that $\langle g \mid \gamma \rangle < 0$ for some continuous function $g \geq 0$, and $f = -\lambda g$ in Proposition 4.3, there holds

$$H(\gamma \mid \rho \otimes \mu) \geq \lambda \langle \gamma \mid -g \rangle - \langle e^{-\lambda g} \mid \rho \otimes \mu \rangle \underset{\lambda\to+\infty}{\longrightarrow} +\infty.$$

Hence, we focus on (4.5) in the sequel, and we proceed as in Section 2.1.1 for the dual formulation of the Kantorovich problem. Using (4.3), the minimization problem (4.1) may be equivalently written as

$$\inf_{\gamma\in\Pi'(\rho,\mu)} \sup_{\phi,\psi,f} \langle c - \phi \oplus \psi \mid \gamma \rangle + \langle \phi \mid \rho \rangle + \langle \psi \mid \mu \rangle + \varepsilon \langle f \mid \gamma \rangle - \varepsilon \langle e^f \mid \rho \otimes \mu \rangle$$

where the supremum is taken over continuous functions $\phi, \psi, f$. The dual problem is obtained by inverting the inf and the sup:

$$\sup_{\phi,\psi,f} \inf_{\gamma\in\Pi'(\rho,\mu)} \langle c - \phi \oplus \psi + \varepsilon f \mid \gamma \rangle + \langle \phi \mid \rho \rangle + \langle \psi \mid \mu \rangle - \varepsilon \langle e^f \mid \rho \otimes \mu \rangle. \tag{4.6}$$

In (4.6), the infimum over $\gamma$ is equal to $-\infty$, unless $c - \phi \oplus \psi + \varepsilon f = 0$ in which case $f = (\phi \oplus \psi - c)/\varepsilon$. Therefore (4.6) reads

$$\sup_{(\phi,\psi)\in C^0(\mathcal{X})\times C^0(\mathcal{Y})} \mathcal{J}^\varepsilon(\phi, \psi) \tag{4.7}$$

where

$$\mathcal{J}^\varepsilon(\phi, \psi) = \langle \phi \mid \rho \rangle + \langle \psi \mid \mu \rangle - \varepsilon \langle e^{\frac{\phi\oplus\psi-c}{\varepsilon}} \mid \rho \otimes \mu \rangle.$$

The problem (4.7) is called the *dual problem*, as opposed to (4.1), which is the *primal problem*. It is a *concave* maximization problem, due to the convexity of the exponential. To find the solution $(\phi, \psi)$ to the dual problem (4.7), we write the first-order optimality conditions. For this we introduce

$$\psi^{c,\varepsilon}(x) = -\varepsilon \log\left(\int_{\mathcal{Y}} e^{\frac{\psi(y)-c(x,y)}{\varepsilon}} \mathrm{d}\mu(y)\right). \tag{4.8}$$

for $\psi \in C^0(\mathcal{Y})$, and

$$\phi^{\bar{c},\varepsilon}(y) = -\varepsilon \log\left(\int_{\mathcal{X}} e^{\frac{\phi(x)-c(x,y)}{\varepsilon}} \mathrm{d}\rho(x)\right) \tag{4.9}$$

for $\phi \in C^0(\mathcal{X})$. These so-called *regularized c-transforms* play the same role as the Legendre transform (2.8) in the unregularized optimal transport problem. And the optimality conditions below are analogues of the relations $\phi = \psi^*$ and $\psi = \phi^*$ for Kantorovich potentials in the unregularized case (see Section 2.1.3).

**Proposition 4.5** (First-order optimality conditions)**.** *The optimality conditions $\nabla_\phi \mathcal{J}^\varepsilon = 0$ and $\nabla_\psi \mathcal{J}^\varepsilon = 0$ are respectively equivalent to $\phi(x) = \psi^{c,\varepsilon}(x)$ for $\rho$-a.e. $x \in \mathcal{X}$, and $\psi(y) = \phi^{\bar{c},\varepsilon}(y)$ for $\mu$-a.e. $y \in \mathcal{Y}$.*

*Proof.* We have for $v \in C^0(\mathcal{X})$

$$\frac{1}{t}\left(\mathcal{J}^\varepsilon(\phi + tv, \psi) - \mathcal{J}^\varepsilon(\phi, \psi)\right) = \langle v \mid \rho \rangle - \frac{\varepsilon}{t}\left(\langle e^{\frac{(\phi+tv)\oplus\psi - c}{\varepsilon}} - e^{\frac{\phi\oplus\psi-c}{\varepsilon}} \mid \rho \otimes \mu\rangle\right)$$

hence

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{J}^\varepsilon(\phi + tv, \psi)\bigg|_{t=0} = \langle v \mid \rho \rangle - \langle v e^{\frac{\phi\oplus\psi-c}{\varepsilon}} \mid \rho \otimes \mu\rangle. \tag{4.10}$$

The first-order optimality condition is thus for $\rho$-a.e. $x \in \mathcal{X}$,

$$1 = \int_{\mathcal{Y}} e^{\frac{\phi(x)+\psi(y)-c(x,y)}{\varepsilon}} \mathrm{d}\mu(y) = e^{\frac{\phi(x)}{\varepsilon}} \int_{\mathcal{Y}} e^{\frac{\psi(y)-c(x,y)}{\varepsilon}} \mathrm{d}\mu(y),$$

i.e., $\phi(x) = \psi^{c,\varepsilon}(x)$. The relation $\psi = \phi^{\bar{c},\varepsilon}$ is proved similarly. □

There is another way to formulate the first-order optimality conditions, which derives directly from (4.10). Fix $(\phi, \psi)$ and let

$$\gamma = e^{\frac{\phi\oplus\psi-c}{\varepsilon}} \rho \otimes \mu.$$

Let also $\Pi_{\mathcal{X}}$ and $\Pi_{\mathcal{Y}}$ denote the projections onto the first and second variables respectively. Then

$$\nabla_\phi \mathcal{J}^\varepsilon(\phi, \psi) = 0 \Leftrightarrow \Pi_{\mathcal{X}\#}\gamma = \rho \tag{4.11}$$

and

$$\nabla_\psi \mathcal{J}^\varepsilon(\phi, \psi) = 0 \Leftrightarrow \Pi_{\mathcal{Y}\#}\gamma = \mu.$$

**Theorem 4.6** (Strong duality)**.** *There exists a solution $(\phi, \psi) \in C^0(\mathcal{X}) \times C^0(\mathcal{Y})$ to the dual problem* (4.7)*. Moreover, the strong duality relation Primal = Dual holds, i.e., the supremum* (4.7) *is equal to the infimum* (4.1)*. Finally, the unique minimizer $\gamma$ in* (4.1) *is given by*

$$\gamma = e^{\frac{\phi\oplus\psi-c}{\varepsilon}} \rho \otimes \mu. \tag{4.12}$$

We give a complete proof of Theorem 4.6. We first show the following weak duality result, which is an easy step:

**Lemma 4.7** (Weak duality)**.** *For any potentials $(\phi, \psi) \in C^0(\mathcal{X}) \times C^0(\mathcal{Y})$ and any transport plan $\gamma \in \Pi(\rho, \mu)$, one has*

$$\mathcal{J}^\varepsilon(\phi, \psi) \leq \int_{\mathcal{X}\times\mathcal{Y}} c \, \mathrm{d}\gamma + \varepsilon H(\gamma \mid \rho \otimes \mu) \tag{4.13}$$

*with equality if $\gamma = e^{\frac{\phi \oplus \psi - c}{\varepsilon}} \rho \otimes \mu$. In particular the weak duality inequality Dual $\leq$ Primal holds, i.e., the supremum (4.7) is upper bounded by the infimum (4.1).*

*Proof of Lemma 4.7.* Let $f = \frac{1}{\varepsilon}(\phi \oplus \psi - c)$. Then

$$\begin{aligned}\mathcal{J}^\varepsilon(\phi, \psi) &= \langle c \mid \gamma\rangle + \varepsilon\langle f \mid \gamma\rangle - \varepsilon\langle e^f \mid \rho \otimes \mu\rangle \\ &\leq \langle c \mid \gamma\rangle + \varepsilon H(\gamma \mid \rho \otimes \mu),\end{aligned}$$

with equality according to Proposition 4.3 if and only if the density of $\gamma$ with respect to $\rho \otimes \mu$ is $e^f$. □

Then, we prove an important fact regarding the modulus of continuity of regularized $c$-transforms. A modulus of continuity for the cost $c$ is a function $\omega_c : \mathbb{R} \to \mathbb{R}$ such that

$$\forall x, x' \in \mathcal{X}, y \in \mathcal{Y}, \qquad |c(x', y) - c(x, y)| \leq \omega_c(d_{\mathcal{X}}(x, x'))$$

where $d_{\mathcal{X}}$ denotes the distance on $\mathcal{X}$. Moduli of continuity of $\psi^{c,\varepsilon}$ and $\phi^{\bar{c},\varepsilon}$ are defined in the same way (it is even simpler since there is only one variable).

**Lemma 4.8.** *For any $(\phi, \psi) \in C^0(\mathcal{X}) \times C^0(\mathcal{Y})$, any modulus of continuity for the cost $c$ is a modulus of continuity of the regularized c-transforms $\psi^{c,\varepsilon}$ and $\phi^{\bar{c},\varepsilon}$.*

*Proof.* We prove this property only for $\psi^{c,\varepsilon}$, the proof being similar for $\phi^{\bar{c},\varepsilon}$. Let $\omega_c$ be a modulus of continuity for the cost $c$. For any $x, x' \in \mathcal{X}$, we have

$$\begin{aligned}\psi^{c,\varepsilon}(x') - \psi^{c,\varepsilon}(x) &= \varepsilon\Big(\log\Big(\langle e^{\frac{\psi - c(x,\cdot)}{\varepsilon}} \mid \mu\rangle\Big) - \log\Big(\langle e^{\frac{\psi - c(x',\cdot)}{\varepsilon}} \mid \mu\rangle\Big)\Big) \\ &= \varepsilon\Big(\log\Big(\langle e^{\frac{\psi - c(x',\cdot)}{\varepsilon}} e^{\frac{c(x',\cdot) - c(x,\cdot)}{\varepsilon}} \mid \mu\rangle\Big) - \log\Big(\langle e^{\frac{\psi - c(x',\cdot)}{\varepsilon}} \mid \mu\rangle\Big)\Big) \\ &\leq \varepsilon\Big(\log\Big(\langle e^{\frac{\psi - c(x',\cdot)}{\varepsilon}} e^{\frac{\omega_c(d_{\mathcal{X}}(x,x'))}{\varepsilon}} \mid \mu\rangle\Big) - \log\Big(\langle e^{\frac{\psi - c(x',\cdot)}{\varepsilon}} \mid \mu\rangle\Big)\Big) \\ &\leq \omega_c(d_{\mathcal{X}}(x, x'))\end{aligned}$$

where $d_{\mathcal{X}}$ denotes the distance on $\mathcal{X}$. □

*Proof of Theorem 4.6.* To prove Theorem 4.6, we first notice that it is sufficient to prove its first part. Indeed, if there exists a solution $(\phi, \psi) \in C^0(\mathcal{X}) \times C^0(\mathcal{Y})$ to the dual problem, then we let $\gamma = e^{\frac{\phi \oplus \psi - c}{\varepsilon}}$ and we observe that for this choice of functions $\phi, \psi, \gamma$, equality holds in (4.13) according to Lemma 4.7. This proves that Primal $\leq$ Dual, i.e., the infimum (4.1) is upper bounded by the supremum in (4.7). Together with Lemma 4.7, this proves the other statements in Theorem 4.6.

There remains to prove the existence of a solution $(\phi, \psi) \in C^0(\mathcal{X}) \times C^0(\mathcal{Y})$ to the dual problem (4.7). The idea is to prove that the supremum can be taken over a compact subset of $C^0(\mathcal{X}) \times C^0(\mathcal{Y})$ where the potentials are uniformly continuous. Using repeatedly Proposition

4.5, which shows that for a fixed $\psi$, the concave function $\mathcal{J}^\varepsilon(\phi,\psi)$ is maximal when $\phi = \psi^{c,\varepsilon}$ (and similarly for fixed $\phi$ it is maximal when $\psi = \phi^{\overline{c},\varepsilon}$), we have

$$\begin{aligned}\sup_{\phi,\psi} \mathcal{J}^\varepsilon(\phi,\psi) = \sup_{\psi\in C^0(\mathcal{Y})} \mathcal{J}^\varepsilon(\psi^{c,\varepsilon},\psi) &= \sup_{\psi\in C^0(\mathcal{Y})} \mathcal{J}^\varepsilon(\psi^{c,\varepsilon},(\psi^{c,\varepsilon})^{\overline{c},\varepsilon}) \\ &= \sup_{\psi\in C^0(\mathcal{Y})} \mathcal{J}^\varepsilon(((\psi^{c,\varepsilon})^{\overline{c},\varepsilon})^{c,\varepsilon},(\psi^{c,\varepsilon})^{\overline{c},\varepsilon}) \\ &= \sup_{\psi\in C^{0,\omega_c}(\mathcal{Y})} \mathcal{J}^\varepsilon(\psi^{c,\varepsilon},\psi)\end{aligned}$$

where in the last line $C^{0,\omega_c}(\mathcal{Y})$ denotes the space of continuous functions on $\mathcal{Y}$ with modulus of continuity $\leq \omega_c$. Since for any constant $C\in\mathbb{R}$, one has $\mathcal{J}^\varepsilon(\phi+C,\psi-C) = \mathcal{J}^\varepsilon(\phi,\psi)$, we may impose without loss of generality that $\langle\psi \mid \mu\rangle = 0$ in the optimization problem. In particular $\psi$ takes both non-negative and non-positive values. Together with the fact that

$$\max_{\mathcal{Y}}\psi - \min_{\mathcal{Y}}\psi \leq \omega_c \mathrm{diam}(\mathcal{Y})$$

we get $\|\psi\|_\infty \leq \omega_c\mathrm{diam}(\mathcal{Y})$. We deduce that

$$\{\psi\in C^{0,\omega_c}(\mathcal{Y}) \mid \langle\psi\mid\mu\rangle = 0\}$$

is a compact subset of $C^0(\mathcal{Y})$. Since $\psi\mapsto\mathcal{J}^\varepsilon(\psi^{c,\varepsilon},\psi)$ is continuous on this set, we conclude thanks to Arzelà-Ascoli's theorem that the supremum in (4.7) is attained. □

From Theorem 4.6 and the uniqueness of the minimizer $\gamma$ in (4.1), we also get the uniqueness of the maximizer in (4.7) up to a constant: for any two maximizers $(\phi_1,\psi_1)$ and $(\phi_2,\psi_2)$ of (4.7), there exists a constant $C$ such that

$$\phi_1 = \phi_2 + C \quad \rho - \text{a.e}, \qquad \psi_1 = \psi_2 - C \quad \mu - \text{a.e.}$$

## 4.2 The regularized Kantorovich functional

### 4.2.1 Definition of the regularized Kantorovich functional

To obtain quantitative stability results for more general costs than the quadratic cost in $\mathbb{R}^d$, it seems natural to seek for an extension of the arguments presented in Section 2.5. Since they rely primarily on the Kantorovich functional, let us first reconsider its definition (we replace $\mathcal{X}$ by an open set $U$ for reasons which will become clear only in Section 4.4):

$$\mathcal{K}_\rho(\psi) = \int_U \psi^*\mathrm{d}\rho$$

where $\psi^*(x) = \sup_y\langle x,y\rangle - \psi(y)$ is the Legendre transform of $\psi$. The reason why the scalar product $\langle x,y\rangle$ appears in the definition of $\mathcal{K}_\rho$ is that the quadratic optimal transport problem can be reformulated equivalently in terms of the cost $\langle x,y\rangle$ (see (2.3) and the dual problem (2.5)). But for a general cost $c(x,y)$, the scalar product has to be replaced by $c(x,y)$ in order to generalize the Kantorovich duality theory presented in Section 2.1. This gives birth to the so-called $c$-transform

$$\psi^c(x) = \inf_{y\in\mathcal{Y}} c(x,y) - \psi(y). \tag{4.14}$$

(Note that an inf replaces here the sup of the Legendre transform, this is due to the minus sign in front of the scalar product in the identity $|x-y|^2 = |x|^2 - 2\langle x,y\rangle + |y|^2$). Therefore a natural generalization of the Kantorovich functional could be

$$\int_U \psi^c\mathrm{d}\rho.$$

However, this is not a good choice: actually, it is hopeless to try to prove its strong concavity [3] using the same tools as for the unregularized Kantorovich functional. To see why, we proceed as in Lemma 2.7 and compute the Hessian of this functional. Let $\psi_t = \psi_0 + tv$ for some $\psi_0, v \in C^0(\mathcal{Y})$. Then

$$\frac{\mathrm{d}}{\mathrm{d}t}\psi_t^c(x) = -v(y(t,x))$$

where $y(t,x)$ is the point (assumed in these formal computations to be unique) where $y \mapsto c(x,y) - \psi_t(y)$ reaches its infimum. Therefore, informally,

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_U \psi_t^c \mathrm{d}\rho = -\int_U v(y(t,x))\mathrm{d}\rho(x)$$

Taking a second derivative, we get

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\int_U \psi_t^c \mathrm{d}\rho = -\int_U \Big(\frac{\mathrm{d}y}{\mathrm{d}t}(t,x)\Big)\nabla v(y(t,x))\mathrm{d}\rho(x) \tag{4.15}$$

Let us compute formally $\frac{\mathrm{d}}{\mathrm{d}t}y(t,x)$. For this we write

$$\begin{aligned}
0 &= \frac{\mathrm{d}}{\mathrm{d}t}[\nabla_2 c(x,y(t,x)) - \nabla\psi_t(y(t,x))] \\
&= \nabla_2 \frac{\mathrm{d}}{\mathrm{d}t}c(x,y(t,x)) - \frac{\mathrm{d}}{\mathrm{d}t}[\nabla\psi_t(y(t,x))] \\
&= \frac{\mathrm{d}y}{\mathrm{d}t}(t,x)\nabla_2^2 c(x,y(t,x)) - \frac{\mathrm{d}y}{\mathrm{d}t}(t,x)\nabla^2\psi_t(y(t,x)) - \nabla v(y(t,x))
\end{aligned}$$

where $\nabla_2 c(\cdot,\cdot)$ denotes the gradient in the second variable of the cost function $c$. When $c(x,y) = -\langle x,y\rangle$ as in Section 2.3, the first term vanishes and we get

$$\frac{\mathrm{d}}{\mathrm{d}t}y(t,x) = -(\nabla^2\psi_t(y(t,x)))^{-1}\nabla v(y(t,x)).$$

Plugging into (4.15) we recover the relation of Lemma 2.7. However, for a general cost $c$, we obtain an expression for $\frac{\mathrm{d}y}{\mathrm{d}t}$ which involves $\nabla_2^2 c$. This seems to destroy the strategy to prove the strong convexity of the functional. In other words, the fact that the quadratic optimal transport problem can be reformulated in terms of the linear cost $\langle x,y\rangle$ makes the computations quite specific when computing the Hessian of the Kantorovich functional $\mathcal{K}_\rho$.

To avoid these issues, one possible idea is to replace $\psi^c$ by the regularized $c$-transform

$$\psi^{c,\varepsilon} : x \mapsto -\varepsilon \log\left(\int_{\mathcal{Y}} e^{\frac{\psi(y)-c(x,y)}{\varepsilon}}\mathrm{d}\mu(y)\right) \tag{4.16}$$

introduced in (4.8) and consider the functional

$$\int_U \psi^{c,\varepsilon}\mathrm{d}\rho.$$

Under the assumption that the support of $\mu$ coincides with $\mathcal{Y}$, we have $\psi^{c,\varepsilon} \to \psi^c$ pointwise as $\varepsilon \to 0$. This functional is almost perfect, but it depends on the measure $\mu$, since the definition of $\psi^{c,\varepsilon}$ itself depends on $\mu$. This is an important issue: the arguments in the quadratic case rely crucially on the fact that the definition of $\mathcal{K}_\rho(\psi)$ does not depend on a target measure $\mu$.

[3] and not convexity, signs are changed here since $\psi^c$ is defined as an infimum!

The argument $\psi$ of the functional $\mathcal{K}_\rho$ is "free", and the target measures are only recovered a posteriori as $(\nabla\psi^*)_\#\rho$.

Therefore, following an idea of Delalande, we finally introduce the regularized Kantorovich functional

$$\mathcal{K}_\rho^\varepsilon(\psi) = \int_U \psi^{c,\varepsilon,\sigma} \, \mathrm{d}\rho \tag{4.17}$$

where, for some probability measure $\sigma$ on $\mathcal{Y}$,

$$\psi^{c,\varepsilon,\sigma} : x \mapsto -\varepsilon \log\left(\int_{\mathcal{Y}} e^{\frac{\psi(y)-c(x,y)}{\varepsilon}} \mathrm{d}\sigma(y)\right).$$

In other words, we just replace $\mu$ by $\sigma$ in (4.16). Notice that $\mathcal{K}_\rho^\varepsilon$ is *concave*, whereas $\mathcal{K}_\rho$ was convex. This is due to the fact that $c$-transforms are defined as an infimum, whereas the Legendre transform is defined as a supremum. Of course this is just a sign convention, which seems to be the most natural one. The only property that we need to impose on $\sigma$ is that its support coincides with $\mathcal{Y}$; the existence of such $\sigma$ for any $\mathcal{Y}$ is easily seen by picking a dense and countable set in $\mathcal{Y}$ and constructing $\sigma$ as a sum of weighted Dirac deltas on this set.

### 4.2.2 Strong convexity of the regularized Kantorovich functional

Following the ideas introduced in Section 2.5, we would like to prove that $\mathcal{K}_\rho^\varepsilon$ is *strongly concave*, in order to deduce later some stability properties of Kantorovich potentials. To prove strong concavity of $\mathcal{K}_\rho^\varepsilon$, one could compute its second derivative and try to use the Brascamp-Lieb inequality (what we did in the unregularized case, in Section 2.5) or another functional inequality. We give below the expression of this second derivative, but it is not clear from there how to deduce strong concavity. Instead, we take inspiration from the proof of the Brascamp-Lieb inequality (Theorem 2.9) given in Section 2.4: we will deduce the strong concavity of $\mathcal{K}_\rho^\varepsilon$ from the concavity of the functional

$$\mathcal{I}_\rho^\varepsilon(\psi) = \log\Big(\int_U e^{\psi^{c,\varepsilon,\sigma}} \, \mathrm{d}\rho\Big)$$

This functional is an analog of

$$\mathcal{I}(\psi) = \log\Big(\int_U e^{-\psi^*} \, \mathrm{d}x\Big)$$

which appears in Section 2.4.

The following statement is given in a general Riemannian manifold $M$, since we will need it at this level of generality in Section 4.4. For $p$-costs in Section 4.3, we only need the easier Corollary 4.10 below, stated in $M = \mathbb{R}^d$.

**Theorem 4.9** (Concavity of $\mathcal{I}_\rho^\varepsilon$)**.** *Let $U$ be a geodesically convex subset of a Riemannian manifold $M$, whose Ricci tensor is denoted by* Ric. *Let also $\mathcal{Y}$ be a compact set, and $c : M \times \mathcal{Y} \to \mathbb{R}$. We assume that there exists $\lambda \in \mathbb{R}$ and $V \in C^2(U)$ such that:*

- *The cost is semi-concave: for all $y \in \mathcal{Y}$, all $x_0$, $x_1 \in U$, and any minimizing geodesic $(x_t)_{t\in[0,1]}$ connecting $x_0$ to $x_1$, one has*

$$c(x_t, y) \geq (1-t)c(x_0, y) + tc(x_1, y) - \frac{\lambda(1-t)t\mathrm{dist}(x_0, x_1)^2}{2}. \tag{4.18}$$

- *The $\infty$-Bakry–Emery tensor is lower bounded by $\lambda$ on $U$:*

$$D^2V + \mathrm{Ric} \geq \lambda. \tag{4.19}$$

*Let $\rho^V$ be the probability measure on $U$*

$$\mathrm{d}\rho^V = \frac{1}{Z}\exp(-V)\mathrm{dvol}$$

*where $Z = \int_U \exp(-V)\mathrm{dvol}$ is a normalizing constant. Then the functional $\mathcal{I}^\varepsilon_{\rho^V}$ is concave.*

*Proof of Theorem 4.9 in a simple case.* We consider the simpler case where $U \subset \mathbb{R}^d$, $\lambda = 0$, $V = 0$ and $x \mapsto c(x, y)$ is concave for any $y \in \mathcal{Y}$. For $\psi_0, \psi_1 \in C^0(\mathcal{Y})$, and $\psi_t = (1-t)\psi_0 + t\psi_1$, we have

$$c(x_t, y) - \psi_t(y) \geq (1-t)(c(x_0, y) - \psi_0(y)) + t(c(x_1, y) - \psi_1(y))$$

where $x_t = (1-t)x_0 + tx_1$. Dividing the above inequality by $-\varepsilon$, exponentiating, integrating over $\mathcal{Y}$ against $\mathrm{d}\sigma$, and finally using the convexity of the function $v \in C^0(\mathcal{Y}) \mapsto \log(\int_{\mathcal{Y}} e^{v(y)}\mathrm{d}\sigma(y))$ we obtain

$$\psi_t^{c,\varepsilon,\sigma}(x_t) \geq (1-t)\psi_0^{c,\varepsilon,\sigma}(x_0) + t\psi_1^{c,\varepsilon,\sigma}(x_1).$$

Exponentiating and using the Prékopa-Leindler inequality (Theorem 2.8) we conclude that

$$\int_U e^{\psi_t^{c,\varepsilon,\sigma}(x)}\mathrm{d}\rho^0(x) \geq \left(\int_U e^{\psi_0^{c,\varepsilon,\sigma}(x)}\mathrm{d}\rho^0(x)\right)^{1-t}\left(\int_U e^{\psi_1^{c,\varepsilon,\sigma}(x)}\mathrm{d}\rho^0(x)\right)^t, \tag{4.20}$$

which shows the concavity of $\mathcal{I}^\varepsilon_{\rho^0}$. □

Instead of proving Theorem 4.9, we have in fact proved the following particular case:

**Corollary 4.10.** *If $x \mapsto c(x, y)$ on $\mathbb{R}^d$ is concave for any $y \in \mathcal{Y}$, then $\mathcal{I}^\varepsilon_{\rho^0}$ is concave.*

To prove Theorem 4.9 in full generality (we will not do it! see [64]), one would need to apply the following weighted Prékopa-Leindler inequality, due to Cordero, McCann and Schmuckenschläger in 2006 [36]:

**Theorem 4.11** (Weighted Prékopa-Leindler inequality)**.** *Let $m = e^{-V}\mathrm{dvol}$ be a measure on a geodesically convex subset $U$ of a Riemannian manifold $M$ where*

$$D^2V + \mathrm{Ric} \geq \lambda \tag{4.21}$$

*for some $\lambda \in \mathbb{R}$. Denote by* dist *the Riemannian distance. Let $s \in [0, 1]$ and $f, g, h : U \to \mathbb{R}_+$ be such that for any $x, y \in U$ and*

$$z \in Z_s(x, y) := \{z \in U \mid \mathrm{dist}(x, z) = s\mathrm{dist}(x, y) \text{ and } \mathrm{dist}(z, y) = (1-s)\mathrm{dist}(x, y)\}$$

*there holds*

$$h(z) \geq e^{-\lambda s(1-s)\mathrm{dist}^2(x,y)/2} f^{1-s}(x) g^s(y).$$

*Then $\int_U h\mathrm{d}m \geq \left(\int_U f\mathrm{d}m\right)^{1-s}\left(\int_U g\mathrm{d}m\right)^s$.*

Let us turn to the consequences of Theorem 4.9. Since some computations are heavy and could obscure the key ideas, we will only give the final expressions; detailed computations may be found in [64]. Let us first introduce some notation for quantities which will naturally appear when we compute the derivatives of $\mathcal{K}^\varepsilon_\rho$ and $\mathcal{I}^\varepsilon_\rho$. To each potential $\psi \in C^0(\mathcal{Y})$ and any point

$x \in U$, we associate a probability density $\hat{\mu}_\varepsilon^x[\psi]$ (with respect to $\sigma$) and the corresponding probability measure $\mu_\varepsilon^x[\psi]$ on $\mathcal{Y}$:

$$\mu_\varepsilon^x[\psi] := \hat{\mu}_\varepsilon^x[\psi] \mathrm{d}\sigma, \qquad \hat{\mu}_\varepsilon^x[\psi](y) := \frac{e^{-\frac{c(x,y)-\psi(y)}{\varepsilon}}}{\int_{\mathcal{Y}} e^{-\frac{c(x,z)-\psi(z)}{\varepsilon}} \mathrm{d}\sigma(z)}.$$

Similarly, given $\psi \in C^0(\mathcal{Y})$, we consider the Gibbs density associated to $\psi^{c,\varepsilon,\sigma}$ denoted by $\hat{\rho}_\varepsilon[\psi]$, and the associated Gibbs measure $\rho_\varepsilon[\psi]$, i.e.

$$\rho_\varepsilon[\psi] := \hat{\rho}_\varepsilon[\psi] \mathrm{d}\rho, \qquad \hat{\rho}_\varepsilon[\psi](x) := \frac{e^{\psi^{c,\varepsilon,\sigma}(x)}}{\int_U e^{\psi^{c,\varepsilon,\sigma}(z)} \mathrm{d}\rho(z)}.$$

Finally, in the sequel we shall sometimes use the notation $\mathbb{E}_\alpha(v) = \int v \mathrm{d}\alpha$ for a measure $\alpha$ and a function $v$. This notation is of course redundant with the notation $\langle v \mid \alpha \rangle$ which we use most of the time, but the interpretation in terms of expectation is sometimes insightful (especially because we shall use the total variance law at some point).

To leverage the concavity of $\mathcal{I}_{\rho^V}^\varepsilon$, we compute its second derivative in direction $v \in C^0(\mathcal{Y})$:

$$\langle D^2 \mathcal{I}_\rho^\varepsilon(\psi) v, v \rangle = \mathrm{Var}_{x \sim \rho_\varepsilon[\psi]}(\mathbb{E}_{\mu_\varepsilon^x[\psi]}(v)) - \frac{1}{\varepsilon} \mathbb{E}_{x \sim \rho_\varepsilon[\psi]}(\mathrm{Var}_{\mu_\varepsilon^x[\psi]}(v)). \tag{4.22}$$

The regularized Kantorovich functional $\mathcal{K}_\rho^\varepsilon$ is twice differentiable at any $\psi \in C^0(\mathcal{Y})$. We identify $\nabla \mathcal{K}_\rho^\varepsilon(\psi)$ to a measure, given by

$$\forall v \in C^0(\mathcal{Y}), \quad \langle v \mid \nabla \mathcal{K}_\rho^\varepsilon(\psi) \rangle = -\int_U \langle v \mid \mu_\varepsilon^x[\psi] \rangle \mathrm{d}\rho(x). \tag{4.23}$$

and this is also equal, with the above notation, to $-\mathbb{E}_{x\sim\rho}(\mathbb{E}_{\mu_\varepsilon^x[\psi]}(v))$. The second derivative is given by

$$\langle D^2 \mathcal{K}_\rho^\varepsilon(\psi) v, v \rangle = -\frac{1}{\varepsilon} \int_U \mathrm{Var}_{\mu_\varepsilon^x[\psi]}(v) \mathrm{d}\rho(x) \tag{4.24}$$

and this can also be written $-\frac{1}{\varepsilon}\mathbb{E}_{x\sim\rho}(\mathrm{Var}_{\mu_\varepsilon^x[\psi]}(v))$. We notice a strong resemblance between (4.24) and the second term in (4.22), which we can leverage to prove a strong concavity estimate for $\mathcal{K}_\rho^\varepsilon$:

$$\begin{aligned} -\langle D^2 \mathcal{K}_\rho^\varepsilon(\psi) v, v \rangle &= \frac{1}{\varepsilon} \mathbb{E}_{x\sim\rho}(\mathrm{Var}_{\mu_\varepsilon^x[\psi]}(v)) \\ &\geq \frac{C_0^{-1}}{\varepsilon} \mathbb{E}_{x\sim\rho_\varepsilon[\psi]}(\mathrm{Var}_{\mu_\varepsilon^x[\psi]}(v)) \qquad (\star) \\ &\geq C_0^{-1} \mathrm{Var}_{x\sim\rho_\varepsilon[\psi]}(\mathbb{E}_{\mu_\varepsilon^x[\psi]}(v)) \qquad (\text{since } \langle D^2 \mathcal{I}_\rho^\varepsilon(\psi) v, v\rangle \leq 0) \\ &\geq C_0^{-2} \mathrm{Var}_{x\sim\rho}(\mathbb{E}_{\mu_\varepsilon^x[\psi]}(v)). \qquad (\star) \end{aligned} \tag{4.25}$$

The two lines denoted by $(\star)$ are obtained by noticing that $C_0^{-1}\rho \leq \rho_\varepsilon[\psi] \leq C_0\rho$ for some constant $C_0$ independent of $\varepsilon$ and $\psi$, due to Lemma 4.8. Following the same arguments which led to (2.28), we obtain:

**Theorem 4.12** (Strong concavity of $\mathcal{K}_\rho^\varepsilon$)**.** *Under the same assumptions as in Theorem 4.9, there holds*

$$\mathrm{Var}_{\rho^V}(\psi_1^{c,\varepsilon,\sigma} - \psi_0^{c,\varepsilon,\sigma}) \leq C_0^2 \langle \psi_1 - \psi_0 \mid \nabla \mathcal{K}_{\rho^V}^\varepsilon(\psi_0) - \nabla \mathcal{K}_{\rho^V}^\varepsilon(\psi_1) \rangle$$

*where $\nabla \mathcal{K}_{\rho^V}^\varepsilon$ is given by the formula* (4.23) *and $\langle \cdot \mid \cdot \rangle$ denotes the duality between continuous functions and Radon measures, as introduced in* (2.14).

*Proof of Theorem 4.12.* We set $\psi_t = \psi_0 + t(\psi_1 - \psi_0)$ and $v = \psi_1 - \psi_0$. The identity $\frac{\mathrm{d}}{\mathrm{d}t}\psi_t^{c,\varepsilon,\sigma}(x) = -\langle v \mid \mu_\varepsilon^x[\psi_t]\rangle$ and the above inequality yield

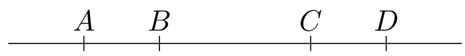
$$\begin{aligned}
\mathrm{Var}_{\rho^V}(\psi_1^{c,\varepsilon,\sigma} - \psi_0^{c,\varepsilon,\sigma}) &= \mathrm{Var}_{\rho^V}\Big(\int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\psi_t^{c,\varepsilon,\sigma}\mathrm{d}t\Big)\\
&\leq \int_0^1 \mathrm{Var}_{\rho^V}\Big(\frac{\mathrm{d}}{\mathrm{d}t}\psi_t^{c,\varepsilon,\sigma}\Big)\mathrm{d}t\\
&= \int_0^1 \mathrm{Var}_{x\sim\rho^V}(\mathbb{E}_{\mu_\varepsilon^x[\psi_t]}(v))\mathrm{d}t\\
&\leq -C_0^2 \int_0^1 \langle D^2\mathcal{K}_{\rho^V}^\varepsilon(\psi_t)v, v\rangle \mathrm{d}t\\
&= C_0^2\langle \psi_1 - \psi_0 \mid \nabla\mathcal{K}_{\rho^V}^\varepsilon(\psi_0) - \nabla\mathcal{K}_{\rho^V}^\varepsilon(\psi_1)\rangle
\end{aligned}$$

which the sought inequality. □

In the next two sections, we explain how to leverage this strong concavity of the regularized Kantorovich functional $\mathcal{K}_\rho^\varepsilon$ to establish quantitative stability estimates for the (*unregularized*) optimal transport problem, with more general costs than the quadratic cost which has been our focus until here. Indeed, the quadratic cost $c(x,y) = |x-y|^2$ in Euclidean spaces is not the only cost of interest in the theory and in applications of optimal transport: in Section 4.3, we consider $p$-costs $c(x,y) = |x-y|^p$ in $\mathbb{R}^d$, for $p > 1$; and in Section 4.4 the squared distance cost in Riemannian manifolds.

## 4.3 Quantitative stability for $p$-costs in Euclidean spaces

Monge's initial problem was formulated in his memoir [80] for the distance cost $|x-y|$ in $\mathbb{R}^d$. However, uniqueness of the optimal transport map fails for this cost, as may be easily seen on $\mathbb{R}$:




On the above picture, if $\rho$ is supported between $A$ and $B$, and $\mu$ is supported between $C$ and $D$, then all maps transporting $\rho$ to $\mu$ have the same cost, and uniqueness thus fails. Intuitively, one may understand it as follows. Let $z$ be a point between $B$ and $C$, which is fixed arbitrarily (for instance, the midpoint of the segment $[BC]$). Let $T$ be a map such that $T_\#\rho = \mu$. If $x$ is in the support of $\rho$, and $T(x)$ denotes its image in the support of $\mu$, we may decompose the path from $x$ to $T(x)$ into a path from $x$ to $z$, followed by a path from $z$ to $T(x)$. Therefore

$$\int_{\mathbb{R}} |x - T(x)|\mathrm{d}\rho(x) = \int_{\mathbb{R}} |x-z|\mathrm{d}\rho(x) + \int_{\mathbb{R}} |z - T(x)|\mathrm{d}\rho(x) = \int_{\mathbb{R}} |x-z|\mathrm{d}\rho(x) + \int_{\mathbb{R}} |z-y|\mathrm{d}\mu(y).$$

We see that this quantity does not depend on $T$, and is thus the same for any transport map $T$ from $\rho$ to $\mu$.

To recover existence and uniqueness (a Brenier type theorem) one may prefer to look at the family of costs $c(x,y) = h(x-y)$ where $h : \mathbb{R}^d \to [0,+\infty)$ is a strictly convex function.

**Theorem 4.13** (Optimal transport map for strictly convex cost)**.** *Let $\rho, \mu \in \mathcal{P}(\mathbb{R}^d)$ with compact support, and let the cost be $c(x,y) = h(x-y)$ with $h : \mathbb{R}^d \to [0,+\infty)$ strictly convex. Then the*

*Kantorovich problem* (1.2) *has a unique solution* $\gamma$*, and this solution is induced by a Borel map* $T : \mathbb{R}^d \to \mathbb{R}^d$ *such that* $\gamma = (\mathrm{Id}, T)_\# \rho$*, and solution to the Monge problem*

$$\inf_{\substack{S:\mathbb{R}^d\to\mathbb{R}^d \\ S_\#\rho=\mu}} \int_{\mathbb{R}^d} h(x - S(x))\mathrm{d}\rho(x).$$

*We have also*

$$T(x) = x - (\partial h)^{-1} (\nabla\phi(x)) \qquad \textit{for } \rho\textit{-a.e. } x, \tag{4.26}$$

*for some c-concave* $\phi$*, meaning that* $\phi$ *is the c-transform of some* $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$*:*

$$\phi(x) = \inf_{y\in\mathbb{R}^d} c(x, y) - \psi(y), \tag{4.27}$$

*and* $\psi$ *is itself the c-transform of* $\phi$*.*

The fact that $h$ is strictly convex is crucial: it ensures that the inverse of the subdifferential $\partial h$ is univalued. One may guarantee that $\phi$, called a Kantorovich potential (for the cost $c$), is Lipschitz on the support of $\rho$, thus differentiable $\rho$-almost everywhere. For given $\rho, \mu$, it is also unique up to constants as soon as $\rho$ has a connected support. When $\rho$ is fixed, as in the quadratic case, we will simply denote by $\phi_\mu$ the unique Kantorovich potential from $\rho$ to $\mu$ that satisfies $\int_\mathcal{X} \phi_\mu \mathrm{d}\rho = 0$.

The above statement applies in particular to the $p$-cost $c(x, y) = \frac{1}{p}|x-y|^p$ for $p > 1$, in which case

$$T(x) = x - (\nabla\phi(x))^{(1/(p-1))}$$

for $\rho$-almost every $x$, where we use the notation $v^{(\alpha)} = |v|^{\alpha-1}v$. For $p = 2$, we recover Brenier's theorem: the potentials $\widetilde{\phi}$ and $\widetilde{\psi}$ solutions to the Kantorovich problem (2.5) are related to the the $c$-concave functions $\phi$ and $\psi$ in Theorem 4.13 by $\widetilde{\phi} = \frac{1}{2}|\cdot|^2 - \phi$ and $\widetilde{\psi} = \frac{1}{2}|\cdot|^2 - \psi$. The reason why the formula (4.26) is natural is that if $x$ is a minimum in the $c$-transform relation $\psi(y) = \inf_{x\in\mathbb{R}^d} c(x, y) - \phi(x)$, then $\nabla_x c(x, y) = \nabla\phi(x)$, i.e., $\nabla h(x - y) = \nabla\phi(x)$, which gives (4.26).

**Theorem 4.14** (Stability of Kantorovich potentials for $C^2$ costs)**.** *Let* $\rho$ *be a log-concave probability measure with bounded convex support* $\mathcal{X} \subset \mathbb{R}^d$*, and let* $\mathcal{Y} \subset \mathbb{R}^d$ *be compact. Let also* $c : \mathcal{X} \times \mathcal{Y} \to [0, +\infty)$ *be a* $C^2$ *cost function. Then there exists* $C < +\infty$ *such that for any* $\mu, \nu \in \mathcal{P}(\mathcal{Y})$*,*

$$\|\phi_\mu - \phi_\nu\|_{L^2(\rho)} \le CW_1(\mu, \nu)^{1/2} \tag{4.28}$$

*where* $\phi_\mu$ *and* $\phi_\nu$ *are the (unique zero-mean) Kantorovich potentials associated to the optimal transport problem with cost c, from* $\rho$ *to* $\mu$ *and* $\rho$ *to* $\nu$ *respectively.*

The above theorem, due to Mischler and Trevisan [81], applies in particular to $p$-costs $c(x, y) = \frac{1}{p}|x - y|^p$, $p \ge 2$. When $p \in (1, 2]$, a similar stability estimate as (4.28) holds, but with a worse exponent. Also, for any $p > 1$, the optimal transport maps enjoy quantitative stability estimates too, but we shall not discuss this here. Notice that $\rho$ is only assumed log-concave, and thus it is not necessarily bounded below on $\mathcal{X}$. Also, it could be generalized to non-convex $\mathcal{X}$ using gluing methods, but we shall not pursue this here.

Let us sketch some proof ideas for Theorem 4.14. The cost $c$ being $C^2$ on the bounded set $\mathcal{X} \times \mathcal{Y}$, there exists $\lambda \ge 0$ such that the semi-concavity estimate (4.18) holds: indeed, for $\gamma > 0$ large enough, the cost

$$\widetilde{c}(x, y) = c(x, y) - \frac{\gamma}{2}|x|^2 \tag{4.29}$$

is concave in its first variable uniformly in $y \in \mathcal{Y}$. Now we notice that this shift of the cost results in a shift of $\mathcal{K}_\rho^\varepsilon$ of $\frac{\gamma}{2}\int_U |x|^2 \mathrm{d}\rho$, which is a quantity independent of $\psi$. Applying Theorem 4.12 with $\lambda = 0$, $\widetilde{c}$ which is concave, $V \equiv 0$, (in order to satisfy (4.19), since $\mathrm{Ric} = 0$), we get that there exists $C_0 > 0$ such that for any $\psi_0, \psi_1 \in C^0(\mathcal{Y})$,

$$\mathrm{Var}_\rho(\psi_1^{c,\varepsilon,\sigma} - \psi_0^{c,\varepsilon,\sigma}) \leq C_0 \langle \psi_1 - \psi_0 \mid \mu_\varepsilon[\psi_1] - \mu_\varepsilon[\psi_0] \rangle$$

where $\langle v \mid \mu_\varepsilon[\psi] \rangle = \int_{\mathcal{X}} \langle v \mid \mu_\varepsilon^x[\psi] \rangle \mathrm{d}\rho(x)$. Sending $\varepsilon$ to 0 and applying the above inequality to $\psi_0 = \psi_\mu$ and $\psi_1 = \psi_\nu$, we collect

$$\mathrm{Var}_\rho(\psi_\nu^c - \psi_\mu^c) \leq C_0 \langle \psi_\nu - \psi_\mu \mid \nu - \mu \rangle$$

since one can prove that $\mu_\varepsilon[\psi_\mu] \to \mu$ as $\varepsilon \to 0$, and similarly $\mu_\varepsilon[\psi_\nu] \to \nu$. The left-hand side is equal to $\|\phi_\mu - \phi_\nu\|_{L^2(\rho)}^2$, and the right-hand side is upper bounded by $CW_1(\mu, \nu)$ thanks to the Kantorovich-Rubinstein formula and Lemma 2.6, which concludes the proof of Theorem 4.14.

## 4.4 Quantitative stability in Riemannian manifolds

It seems natural to investigate the stability of optimal transport maps (and potentials) also on Riemannian manifolds. Indeed, the extension of optimal transportation theory to probability densities on general Riemannian manifolds, rather than only on Euclidean spaces, is interesting both for theory and applications. On the theoretical side, there are many links between optimal transport and Riemannian geometry, plenty of which are presented in the book [106]. On the side of applications, it often happens that data distributions are supported on Riemannian manifolds, for instance in medical imaging, where Riemannian optimal transport is used to match 3D anatomical structures, among many other applications.

The analogue of Brenier's theorem in the Riemannian context was proved by McCann in [78]:

**Theorem 4.15** (McCann's Riemannian theorem)**.** *Let $M$ be a connected, complete, smooth Riemannian manifold. Let $\rho$, $\mu$ be probability measures on $M$ with compact support, and let the cost function $c(x,y)$ be equal to $\frac{1}{2}\mathrm{dist}(x,y)^2$, where* dist *is the Riemannian distance on $M$. Further, assume that $\rho$ is absolutely continuous with respect to the volume measure on $M$. Then there is a $\rho$-a.e. unique solution of the Monge problem* (1.4) *between $\rho$ and $\mu$, and it can be written as*

$$T(x) = \exp_x(-\nabla\phi(x)) \tag{4.30}$$

*where $\phi = \psi^c$ is the c-transform, in the sense of* (4.27)*, of some $\psi : M \to \mathbb{R} \cup \{\pm\infty\}$. Here, $\nabla$ is the Riemannian gradient and $\exp_x$ denotes the exponential map with footpoint $x$: in other words, $T(x)$ is the endpoint (at time 1) of the geodesic issued from $x$, with initial tangent vector $-\nabla\phi(x)$ in the tangent space $T_xM$.*

In a recent work with Jun Kitagawa and Quentin Mérigot [64], we proved quantitative stability estimates for optimal transport maps and Kantorovich potentials when the cost is the squared distance $c(x,y) = \frac{1}{2}\mathrm{dist}(x,y)^2$ in a Riemannian manifold. These estimates are similar to those obtained in the previous chapters in the Euclidean setting (recall that John domains in general metric spaces were introduced in Definition 1.11):

**Theorem 4.16** (Quantitative stability in Riemannian manifolds)**.** *Let $M$ be a smooth and connected $d$-dimensional Riemannian manifold, endowed with the quadratic cost $c(x,y) = \frac{1}{2}\mathrm{dist}(x,y)^2$ where* dist *denotes the Riemannian distance. Let $\mathcal{X} \subset M$ be a John domain and $\mathcal{Y} \subset M$ be* *compact, and let $\rho$ be a probability measure absolutely continuous with respect to the Riemannian volume on $\mathcal{X}$, with density bounded from above and below by positive constants. Then there exists $C > 0$ such that for any $\mu, \nu \in \mathcal{P}(\mathcal{Y})$,*

$$\|\phi_\mu - \phi_\nu\|_{L^2(\rho)} \le C W_1(\mu, \nu)^{1/2} \tag{4.31}$$

*where $\phi_\mu$ denotes the (unique zero-mean) Kantorovich potential from $\rho$ to $\mu$, and similarly $\phi_\nu$ denotes the unique Kantorovich potential from $\rho$ to $\nu$. Moreover, if the boundary $\partial\mathcal{X}$ of $\mathcal{X}$ has finite $(d-1)$-dimensional Hausdorff measure, then*

$$\left(\int_M \mathrm{dist}(T_\mu(x), T_\nu(x))^2 d\rho(x)\right)^{\frac{1}{2}} \le C W_1(\mu, \nu)^{1/6} \tag{4.32}$$

*where $T_\mu$ and $T_\nu$ are the optimal transport maps from $\rho$ to $\mu$ and $\rho$ to $\nu$ respectively.*

We focus here on explaining in broad lines the stability of Kantorovich potentials (4.31). We shall not discuss at all the stability of optimal transport maps (4.32), which follows from the stability of Kantorovich potentials using integral-geometric techniques generalizing those of Section 3.8.2. For a detailed proof of Theorem 4.16, we refer to [64].

*Proof of* (4.31). The proof of (4.31) consists in applying Theorem 4.12 in small balls, before operating a gluing argument. The squared Riemannian distance cost is indeed semiconcave (but not concave!) in small balls, as we explain below. We proceed in four steps.

*Step 1: Variance inequality in small balls.* For the first step of the proof we work in a small enough ball. Let $x_0 \in \mathcal{X}$, and consider $B(x_0, r) \subset M$ such that $2r$ is smaller than the injectivity radius at $x_0$, in particular any two points in $B(x_0, r)$ are joined by a unique geodesic. It follows for instance from results by Ohta [86, Lemma 3.3] that there exists $\lambda \ge 0$ such that for any $x_0, x_1, y \in B(x_0, r)$, and for any $t \in [0, 1]$,

$$\mathrm{dist}^2(x_t, y) \ge (1-t)\mathrm{dist}^2(x_0, y) + t\mathrm{dist}^2(x_1, y) - \lambda t(1-t)\mathrm{dist}^2(x_0, x_1) \tag{4.33}$$

where $(x_s)_{s\in[0,1]}$ denotes the unique geodesic from $x_0$ to $x_1$, traveled at speed 1. This is a semiconcavity estimate for $x \mapsto \mathrm{dist}^2(x, y)$.

Besides, in $B(x_0, r)$, the potential $V_0 : x \mapsto \mathrm{dist}(x, x_0)^2$ is also strongly convex. Since the Ricci curvature in the compact set $B(x_0, r)$ is bounded below,

$$D^2 V + \mathrm{Ric} \ge \lambda$$

for $V(x) = K\mathrm{dist}(x, x_0)^2$ with $K$ chosen large enough. Therefore we may apply Theorem 4.12 in $U = B(x_0, r)$ and obtain a variance inequality

$$\mathrm{Var}_{\rho^V}(\psi_1^{c,\varepsilon,\sigma} - \psi_0^{c,\varepsilon,\sigma}) \le C_0 \langle \psi_1 - \psi_0 \mid \mu_\varepsilon^V[\psi_1] - \mu_\varepsilon^V[\psi_0] \rangle \tag{4.34}$$

where

$$\rho^V = \frac{\exp(-V)\mathbf{1}_{B(x_0,r)}}{\int_{B(x_0,r)} \exp(-V)\mathrm{dvol}}\mathrm{dvol}, \qquad \langle v \mid \mu_\varepsilon^V[\psi] \rangle = \int_{\mathcal{X}} \langle v \mid \mu_\varepsilon^x[\psi] \rangle \mathrm{d}\rho^V(x).$$

*Step 2: Uniformity.* The previous arguments were accomplished in a sufficiently small ball, and a priori the size $r$ of the ball, the positive number $K$ and the constant $C_0$ depend on the point $x_0$. But due to the boundedness of $\mathcal{X}$, it is actually possible to take all these parameters uniform (i.e., independent) in $x_0$.

*Step 3: Gluing argument.* To obtain a global variance inequality, we use the same gluing arguments as in Section 3: we cover the set $\mathcal{X}$ by a finite number of balls in which the previous steps may be applied, and then we use a gluing inequality of the form (3.3). We get

$$\mathrm{Var}_\rho(\psi_1^{c,\varepsilon,\sigma} - \psi_0^{c,\varepsilon,\sigma}) \leq C_0\langle \psi_1 - \psi_0 \mid \mu_\varepsilon[\psi_1] - \mu_\varepsilon[\psi_0]\rangle$$

where $\langle v \mid \mu_\varepsilon[\psi]\rangle = \int_{\mathcal{X}} \langle v \mid \mu_\varepsilon^x[\psi]\rangle \mathrm{d}\rho(x)$ (actually, we skip here a short argument allowing to replace $\rho^V$ by $\rho$).

*Step 4: Conclusion.* Sending $\varepsilon$ to 0 and applying the above inequality to $\psi_0 = \psi_\mu$ and $\psi_1 = \psi_\nu$, we collect

$$\mathrm{Var}_\rho(\psi_\nu^c - \psi_\mu^c) \leq C_0\langle \psi_\nu - \psi_\mu \mid \nu - \mu\rangle$$

since one can prove that $\mu_\varepsilon[\psi_\mu] \to \mu$ as $\varepsilon \to 0$, and similarly $\mu_\varepsilon[\psi_\nu] \to \nu$. Finally, we observe that the right-hand side is equal to $\|\phi_\mu - \phi_\nu\|^2_{L^2(\rho)}$ and that the right-hand side is upper bounded by $CW_1(\mu, \nu)$ thanks to the Kantorovich-Rubinstein duality formula, which concludes the proof. □

**Open question 4.17.** *The proof sketched above does not rely on the differentiable structure of Riemannian manifolds. Therefore, it would be natural to generalize it to more general metric measure spaces where the Prékopa-Leindler inequality is known to hold. We believe it would be particularly interesting to develop a theory of stability of optimal transport for two categories of metric measure spaces: the curvature-dimension spaces CD(K,N) introduced by Sturm [102] and Lott-Villani [74] to develop a synthetic theory of Ricci curvature; and sub-Riemannian manifolds, for which optimal transport has been studied for instance in [50] and [8].*

**Open question 4.18.** *Does quantitative stability of Kantorovich potentials and optimal transport maps for the squared distance cost hold when the source measure is* $\rho = e^{-V}\mathrm{dvol}$ *on a non-compact Riemannian manifold satisfying the Bakry-Emery condition* $D^2V + \mathrm{Ric} > 0$*? This would be a natural generalization of Theorem 1.10.*

## 4.5 Bibliographical notes

§4.1: This section is mostly taken from the lecture notes by Mérigot at the Institut Henri Poincaré, available on his webpage. Other introductions to entropic optimal transport may be found in the lecture notes by Nutz [85] and in the book chapters [89, Chapter 4] and [32, Chapter 3]. Entropic optimal transport is related to a minimal entropy problem considered by Schrödinger in 1931 (see the survey [70]). It has attracted a lot of interest since Cuturi's groundbreaking work [38] which showed how to solve it "at lightspeed" using the Sinkhorn algorithm. See Section 5.5 for more references.

§4.2: The regularized Kantorovich functional was introduced by Delalande in [41] and his PhD thesis [40] for the quadratic cost. It was then noticed in [81], [33] and [64] that it could be generalized to any semi-concave cost.

§4.3: Theorem 4.13 about the existence of an optimal transport for $p$-costs, $p > 1$, can be found in [2, Theorem 6.2.4]. The results of Mischler and Trevisan [81] are more general than Theorem 4.14: their statements cover the whole range of $p$-costs, $p > 1$, and hold for optimal transport maps (not only Kantorovich potentials).

§4.4: McCann's Theorem 4.15, proved in [78], was the first general optimal transport theorem on a Riemannian manifold. This result is an important ingredient in [35] and [36], which develop a Riemannian version of classical interpolation inequalities such as the Prékopa-Leindler and the Borell-Brascamp-Lieb inequality. Theorem 4.11 is proved in [36], which deals with weighted

Riemannian manifolds. Regarding optimal transport stability, an extension of Gigli's result (Theorem 1.9) to Riemannian manifolds was proved in [3, Section 3]. Riemannian optimal transport has also found nice applications, see for instance [101] and the works of Jean Feydy, Alain Trouvé, François-Xavier Vialard, and many others.

# 5 Further applications

In this last chapter, we review several applications of optimal transport stability bounds. One of the most important applications is to the statistical estimation of optimal transport maps, detailed in Section 5.1. We then discuss in Section 5.2 an application of the methods developed in Chapters 2 and 3 to the stability of Wasserstein barycenters, which define in a natural way weighted averages of probability measures, and are broadly used in applications. A final key application concerns the convergence of the Sinkhorn algorithm. This algorithm was invented by Richard Sinkhorn in 1964 to show that a positive matrix, iteratively scaled by normalizing rows and columns alternately, becomes doubly stochastic. It was discovered in 2013 by Marco Cuturi that the Sinkhorn algorithm could actually be used to solve "at lightspeed" the entropic optimal transport problem. This observation has led to tremendous progress in the field of computational optimal transport. We show in Section 5.3 the exponential convergence of Sinkhorn's algorithm in the continuous setting, and we highlight the deep connections with the quantitative stability of entropic optimal transport. We conclude this chapter with some perspectives and open problems for future research.

## 5.1 Statistical optimal transport

Stability of optimal transport has been used as a fundamental tool in statistical optimal transport. In this field, a key question is to estimate objects which arise from the optimal transport framework when the two measures are unknown but we have access to (sometimes random) samples of them. In the sequel, we focus on the following question: given samples $X_1, \dots, X_n \sim \rho$ and $Y_1, \dots, Y_m \sim \mu$, how can we estimate the optimal transport map $T$ from $\rho$ to $\mu$ via an estimator $\widehat{T}$ constructed on the basis of the samples? A possible measure of performance is the squared error

$$\int_{\mathcal{X}} |\widehat{T}(x) - T(x)|^2 \mathrm{d}\rho(x).$$

The primary goal is to design estimators that come with strong theoretical statistical guarantees—that is, they achieve a specified level of accuracy with high probability—regardless of their computational cost. This will be our main focus here. A secondary goal, which has also been explored in the literature, is to develop estimators that are not only theoretically sound but also computationally efficient.

Although many types of estimators have been used in the literature, we shall focus here on two of them, among the most important ones: plug-in estimators and semi-dual estimators. Both are defined below. These estimators are the only ones known to be minimax optimal over typical classes of source measures and of smooth optimal transport maps: this means that given a class $\mathcal{M}$ of "nice" source measures $\rho$ and a class $\mathcal{C}$ of smooth optimal transport maps, they minimize the quantity

$$\sup_{\substack{\rho \in \mathcal{M} \\ T \in \mathcal{C}}} \mathbb{E} \int_{\mathcal{X}} |\widehat{T}(x) - T(x)|^2 \mathrm{d}\rho(x)$$

where the expectation is taken with respect to the data $X_1, \dots, X_n \sim \rho$ and $Y_1, \dots, Y_m \sim T_\# \rho$.

Despite many advances in this field, some important open questions remain unsolved. One of them pertains to the assumptions that are made on the optimal transport map to be estimated: in almost all works on the topic, it is assumed that the optimal transport map is at least Lipschitz (or bi-Lipschitz, meaning that the inverse map is well-defined and also Lipschitz) to obtain good statistical guarantees. The tools presented in these notes allow to derive statistical non-guarantees for the estimation of non-smooth optimal transport maps, but they are not optimal, see Remark 5.3 below.

### 5.1.1 Plug-in estimators

We start our presentation with plug-in estimators. Their principle is particularly simple: they are defined as optimal transport maps (or optimal couplings) between measures derived from the observations, appropriately extended so that they define functions on $\mathbb{R}^d$. They are of different types.

Perhaps the most natural plug-in estimators are the *empirical estimators*. For simplicity, assume that the number of observations $X_1, \dots, X_n$ of the source measure is equal to the number of observations $Y_1, \dots, Y_m$ of the target measure, i.e., $m = n$. Let $\rho_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ and $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$, and consider the discrete optimal transport problem

$$\min_{\gamma \in \Pi(\rho_n, \mu_n)} \int |x-y|^2 \mathrm{d}\gamma(x,y). \tag{5.1}$$

In favorable cases, there is a unique solution to (5.1), induced by a transport map $T_n$. It may be computed thanks to the Hungarian algorithm or Bertsekas' auction algorithm. But in any case, a solution to (5.1) is only defined on the samples $X_1, \dots, X_n$. To extend it to the whole source space $\mathcal{X}$, different possibility have been considered, notably the 1-nearest neighbor extension: $\widehat{T}(x)$ is defined as $T_n(X_i)$, where $X_i$ is the closest sample point to $x$.

*Smooth estimators* are another category of plug-in estimators. They are also based on approximations $\rho_n$ and $\mu_m$ of $\rho$ and $\mu$, but this time they are constructed in a way that they are smooth. Such approximations may be constructed using wavelets, kernels, or simply the optimal transport map from $\rho_n$ to $\mu_m$, if it exists and it is smooth. Contrarily to empirical estimators, smooth estimators may become more accurate when $\rho$ and $\mu$ are assumed more regular.

Theoretical guarantees on plug-in estimators are based on quantitative stability bounds. These bounds require strong assumptions on the optimal transport maps, often verified using the Caffarelli regularity theory of optimal transport. We give one example of such bounds, due to Manole and Balakrishnan [6].

**Theorem 5.1** (Quantitative stability of maps under regularity assumptions)**.** *Let $\rho \in \mathcal{P}_2(\mathcal{X})$ with $\mathcal{X} \subset \mathbb{R}^d$ and $\rho$ is absolutely continuous. Assume that the Kantorovich potential $\phi_0 : \mathcal{X} \to \mathbb{R}$ from $\rho$ to $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ is $\alpha$-strongly convex ($D^2\phi_0 \geq \alpha \mathrm{Id}$) and $\beta$-smooth ($D^2\phi_0 \leq \beta \mathrm{Id}$) for some $\alpha, \beta > 0$. Let $T_0 = \nabla \phi_0$. For any $\widehat{\rho}, \widehat{\mu}$, let $\widehat{\gamma}$ denote an optimal coupling of $\widehat{\rho}$ and $\widehat{\mu}$. Then*

$$\frac{1}{\beta}\int_{\mathcal{X}\times\mathcal{Y}} |y - T_0(x)|^2 \mathrm{d}\widehat{\gamma}(x,y) \leq \frac{1}{\alpha} W_2^2(\widehat{\mu}, \mu) + \beta W_2^2(\widehat{\rho}, \rho) + 2W_2(\widehat{\rho}, \rho) W_2(\widehat{\mu}, \mu).$$

In the so-called 1-sample setting where the absolutely continuous measure $\rho \in \mathcal{P}_2(\mathcal{X})$ is known, there is an optimal map $\widehat{T}$ between $\rho$ and $\widehat{\mu}$, and Theorem 5.1 implies:

**Theorem 5.2.** *Under the same assumptions as Theorem 5.1, let $\widehat{T}$ denote the optimal transport map from $\rho$ to $\widehat{\mu}$, and $T_0$ denote the optimal transport map from $\rho$ to $\mu$. Then*

$$\frac{1}{\beta}\int_{\mathcal{X}} |\widehat{T}(x) - T_0(x)|^2 \mathrm{d}\rho \leq \frac{1}{\alpha} W_2^2(\widehat{\mu}, \mu).$$

It is worth comparing Theorem 5.2 to Theorem 1.9: Theorem 5.2 needs stronger assumptions (bi-Lipschitzness of the optimal transport map instead of Lipschitzness) but yields a $W_2^2$ bound, whereas Theorem 1.9 would only yield $W_2$ if applying the inequality $W_1 \leq W_2$.

If $\widehat{\rho}$ and/or $\widehat{\mu}$ are obtained by taking $n$ i.i.d. samples from $\rho$ and $\mu$ respectively, the bounds in terms of Wasserstein distances in Theorem 5.1 and Theorem 5.2 may be turned into bounds in expectation, e.g., an upper bound on $\mathbb{E}(\|\widehat{T}-T_0\|^2_{L^2(\rho)})$ in Theorem 5.2, thanks to classical estimates (see [51]) such as

$$\mathbb{E}(W_2^2(\widehat{\mu},\mu)) \leq C \begin{cases} n^{-1/2} & \text{if } d \leq 3 \\ n^{-1/2}\log(n) & \text{if } d = 4 \\ n^{-2/d} & \text{if } d \geq 5. \end{cases} \tag{5.2}$$

valid for any compactly supported $\mu$.

We now prove Theorem 5.2 (but not the stronger Theorem 5.1, which requires additional ideas). The proof is quite close to the proof of Theorem 1.9 in Section 2.2.

*Proof of Theorem 5.2.* We consider the dual functional

$$\mathcal{S}_{\rho,\widehat{\mu}}(\phi) = \int_{\mathcal{X}} \phi \,\mathrm{d}\rho + \int_{\mathcal{Y}} \phi^* \,\mathrm{d}\widehat{\mu}.$$

It will be convenient to write this proof with expectations instead of integrals, and more generally with a probabilistic perspective instead of the more analytic viewpoint of Section 2.2. Let $\widehat{\gamma}$ denote an optimal coupling of $\rho$ and $\widehat{\mu}$, and let $(X,Y)\sim\widehat{\gamma}$. We have

$$\mathcal{S}_{\rho,\widehat{\mu}}(\phi_0) - \mathcal{S}_{\rho,\widehat{\mu}}(\widehat{\phi}) = \mathbb{E}(\phi_0(X) + \phi_0^*(Y) - \widehat{\phi}(X) - \widehat{\phi}^*(Y)).$$

There holds $\widehat{\phi}(X) + \widehat{\phi}^*(Y) = \langle X, Y\rangle$, $\widehat{\gamma}$-almost surely. Also, the equality case in the Fenchel-Young inequality yields $\phi_0(X) = \langle X, T_0(X)\rangle - \phi_0^*(T_0(X))$ (recall that $T_0 = \nabla\phi_0$). Hence,

$$\mathcal{S}_{\rho,\widehat{\mu}}(\phi_0) - \mathcal{S}_{\rho,\widehat{\mu}}(\widehat{\phi}) = \mathbb{E}[\phi_0^*(Y) - \phi_0^*(T_0(X)) - \langle X, Y - T_0(X)\rangle].$$

Given any random variable $Z$ with law $\widehat{\mu}$, we can equivalently write

$$\mathcal{S}_{\rho,\widehat{\mu}}(\phi_0) - \mathcal{S}_{\rho,\widehat{\mu}}(\widehat{\phi}) = \mathbb{E}[\phi_0^*(Z) - \phi_0^*(T_0(X)) - \langle X, Z - T_0(X)\rangle] - \mathbb{E}\langle X, Y - Z\rangle.$$

Since $(X,Y)$ is an optimal coupling between $\rho$ and $\widehat{\mu}$, $\mathbb{E}\langle X,Y\rangle \geq \mathbb{E}\langle X,Z\rangle$, and we obtain

$$\mathcal{S}_{\rho,\widehat{\mu}}(\phi_0) - \mathcal{S}_{\rho,\widehat{\mu}}(\widehat{\phi}) \leq \mathbb{E}[\phi_0^*(Z) - \phi_0^*(T_0(X)) - \langle X, Z - T_0(X)\rangle].$$

The $X$ in the scalar product can be replaced by $\nabla\phi_0^*(T_0(X))$, since $\nabla\phi_0^*(T_0(X)) = X$ almost surely. Moreover, $\phi_0^*$ is $1/\alpha$-smooth according to Lemma 2.4, hence

$$\mathcal{S}_{\rho,\widehat{\mu}}(\phi_0) - \mathcal{S}_{\rho,\widehat{\mu}}(\widehat{\phi}) \leq \frac{1}{\alpha}\mathbb{E}\|Z - T_0(X)\|_2^2. \tag{5.3}$$

This holds for any random variable $Z\sim\widehat{\mu}$. We choose $Z$ in a way that the right-hand side is minimal: since $T_0(X)\sim\mu$, we pick $Z$ so that $(T_0(X), Z)$ is an optimal coupling between $\mu$ and $\widehat{\mu}$. The right-hand side of (5.3) is equal to $\frac{1}{\alpha}W_2^2(\mu,\widehat{\mu})$ in this case, which concludes the proof. □

**Remark 5.3.** *Let us observe that the results presented in the first chapters may be used to derive statistical guarantees on plug-in estimators* without assuming that the optimal transport map is Lipschitz. *We only need to make the same assumptions as, for instance, in Theorem 1.12, namely that $\rho$ is supported on a John domain (e.g., a compact connected set with Lipschitz boundary), has a positive density on this set, and that $\mu$ is compactly supported. In this case,* *applying Theorem 1.12 to $\nu = \widehat{\mu}$ obtained as the empirical distribution of $n$ i.i.d. samples of $\mu$, we get*

$$\mathbb{E}(\|T_\mu - \widehat{T}\|^2_{L^2(\rho)}) \lesssim \mathbb{E}(W_2(\mu, \widehat{\mu})^{1/3}) \lesssim n^{-1/3d} \tag{5.4}$$

*(for the last inequality we assumed $d \geq 5$ and used (5.2)). Notice that this bound does not reach the minimax optimal rate discussed below around Theorem 5.5.*

**Open question 5.4.** *It would be very interesting to develop statistical tools estimating optimal transport maps with minimax optimal rates, without assuming (as much) regularity of the optimal transport map.*

Not assuming any regularity of the optimal transport map, we may expect different minimax rates of convergence than those for regular optimal transport maps written in Theorem 5.5. One reason for this is that optimal transport map estimation is deeply related to the density estimation problem in Wasserstein distance. Indeed, to estimate a transport map with a plug-in estimator we need to estimate the target density, as in (5.4). And for the latter density estimation problem, the work [84] shows that for $\mathcal{X} = [0,1]^d$, the class of densities bounded from below, which corresponds to regular optimal transport maps since Caffarelli's regularity theory applies to such target measures on $[0,1]^d$, is strictly easier to estimate than the class of all densities, for which the optimal transport map is not regular in general.

### 5.1.2 Semi-dual estimators

We turn to a second type of estimators, the so-called *semi-dual estimators*. They are based on the semidual formulation of optimal transport $\phi_0 \in \operatorname{argmin}_{\phi \in C^0(\mathcal{X})} \mathcal{S}_{\rho,\mu}$ where recall that

$$\mathcal{S}_{\rho,\mu}(\phi) = \int_{\mathcal{X}} \phi \, \mathrm{d}\rho + \int_{\mathcal{Y}} \phi^* \, \mathrm{d}\mu. \tag{5.5}$$

This semi-dual formulation is trivially equivalent to (2.9). For

$$\rho_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \qquad \mu_m = \frac{1}{m} \sum_{j=1}^m \delta_{Y_j}$$

consider

$$\widehat{\phi} \in \operatorname*{argmin}_{\phi \in \mathcal{F}} \int_{\mathcal{X}} \phi \, \mathrm{d}\rho_n + \int_{\mathcal{Y}} \phi^* \, \mathrm{d}\mu_m \tag{5.6}$$

where $\mathcal{F}$ is some class of differentiable functions, possibly depending on $n, m$. The semi-dual estimator of the optimal transport map is defined as $\widehat{T} = \nabla \widehat{\phi}$. We will focus in the rest of this section the following important result due to Hütter and Rigollet [59] whose proof heavily relies on stability estimates:

**Theorem 5.5** (Minimax estimation of smooth optimal transport maps, informal statement)**.** *If the optimal transport map $T_0 = \nabla \phi_0$ is bi-Lipschitz and $C^\alpha$ for some $\alpha > 1$ (possibly not an integer), then for an appropriate choice of the family $\mathcal{F}$ (depending on $\alpha$), the semidual estimator $\nabla \widehat{\phi}$ achieves for $m = n$ the rate*

$$\mathbb{E}\|\nabla \widehat{\phi} - \nabla \phi_0\|^2_{L^2(\rho)} \lesssim n^{-\frac{2\alpha}{2\alpha - 2 + d}} (\log n)^2 \vee \frac{1}{n}. \tag{5.7}$$

*This rate is off the minimax optimal rate only by the factor $(\log n)^2$.*

The rate (5.7) suffers from the curse of dimensionality: for $s$ fixed, the exponent is equivalent to $-2/d$ as $d \to +\infty$. The choice of the family $\mathcal{F}$ appearing in (5.6) and in Theorem 5.5 is of paramount importance. Since the Kantorovich potential $\phi_0$ is assumed to be quite regular, it is natural to take for $\mathcal{F}$ some $L^2$ basis where regularity may be read in the decay of the coefficients, e.g., a wavelet basis. Actually, the basis is often truncated at a certain threshold, depending on $n, m$ and on the parameter $\alpha$ controlling the regularity of $\phi_0$, in order to reduce the variance of the estimator. This truncation introduces a small bias, and statisticians choose the tuncation threshold to minimize the sum of the variance and the bias (this is an example of a bias-variance tradeoff). Other choices of $\mathcal{F}$ exist, depending on the problem under consideration. For instance, if $\rho$ and $\mu$ are Gaussian the optimal transport is known to be linear. In this case, a clever choice for $\mathcal{F}$ is the set of convex quadratic maps, so that the estimated transport map becomes linear.

The above definition (5.6) fits exactly into the framework of empirical risk minimization (ERM). In ERM, to approximate the solution to an optimization problem where the function to be optimized (the "true risk") is unknown, one solves an optimization problem over an empirical risk, computed with a known set of training data. The analysis of these estimators is related to the theory of $M$-estimators (estimation of a maximum).

To study the theoretical guarantees associated to the estimator $\nabla\widehat{\phi}$ with $\widehat{\phi}$ given by (5.6), a fundamental role is played by the behavior of the semi-dual functional $\mathcal{S}_{\rho,\mu}$ near the solution $\phi_0$ of (5.5). Under the assumption that $T_0$ is bi-Lipschitz, it is possible to prove the existence of $C_1, C_2 > 0$ (which depend on the Lipschitz constants of $T_0$) such that

$$C_1\|\nabla\phi - \nabla\phi_0\|_{L^2(\rho)}^2 \leq \mathcal{S}_{\rho,\mu}(\phi) - \mathcal{S}_{\rho,\mu}(\phi_0) \leq C_2\|\nabla\phi - \nabla\phi_0\|_{L^2(\rho)}^2 \tag{5.8}$$

for any differentiable $\phi$. The demonstration of (5.8) is based on the same ingredients as the proofs of Theorem 1.9 and Theorem 5.2, therefore we do not detail it here. Combining (5.8) with the optimality of $\widehat{\phi}$ for $\mathcal{S}_{\rho_n,\mu_m}$, we get

$$C_1\|\nabla\widehat{\phi} - \nabla\phi_0\|_{L^2(\rho)}^2 \leq \mathcal{S}_{\rho,\mu}(\widehat{\phi}) - \mathcal{S}_{\rho_n,\mu_m}(\widehat{\phi}) - (\mathcal{S}_{\rho,\mu}(\phi_0) - \mathcal{S}_{\rho_n,\mu_m}(\phi_0)).$$

Since $\phi_0$ and $\widehat{\phi}$ are regular, it is possible to upper bound the right-hand side (in expectation) using empirical process theory, and this allows to conclude that $\mathbb{E}\|\nabla\widehat{\phi} - \nabla\phi_0\|_{L^2(\rho)}^2$ is small.

## 5.2 Stability of Wasserstein barycenters

Another important application of the methods presented in these notes is to the stability of Wasserstein barycenters. These objects are defined as means, in the Wasserstein space, of probability measures. If $\mathcal{X}$ is a Polish space, $\rho_1, \ldots, \rho_k$ are $k$ probability measures on $\mathcal{X}$, and $p_1, \ldots, p_k \geq 0$ verify $p_1 + \ldots + p_k = 1$, a Wasserstein barycenter is a "weighted average" of the $\rho_j$, with weights $p_j$. Precisely, it is a probability measure $\mu$ on $\mathcal{X}$ that minimizes the quantity

$$\frac{1}{2}\sum_{j=1}^k p_j W_2^2(\rho_j, \mu).$$

When $k = 2$, one recovers the notion of interpolation between two measures in the Wasserstein space, first studied by McCann. More generally, one may consider averages of an infinite number of probability measures on $\mathcal{X}$, with weights given by a probability measure $\mathbb{P}$ on the set of probability measures on $\mathcal{X}$: for $\mathbb{P} \in \mathcal{P}(\mathcal{P}_2(\mathcal{X}))$, a Wasserstein barycenter is a minimizer $\mu_{\mathbb{P}}$ of the functional

$$F_{\mathbb{P}} : \mu \mapsto \frac{1}{2}\int_{\mathcal{P}(\mathcal{X})} W_2^2(\rho, \mu) \mathrm{d}\mathbb{P}(\rho). \tag{5.9}$$

This definition of a barycenter as a solution to a variational problem involving the squared distance can be extended to any metric space, replacing the Wasserstein distance by the distance on the space. Wasserstein barycenters have found applications in shape interpolation [101], texture synthesis and mixing [94], color harmonization [15], etc.

Wasserstein barycenters exist and are unique under mild assumptions. We provide here an existence and uniqueness statement in a simple framework, to avoid unnecessary complications:

**Theorem 5.6** (Existence and uniqueness of Wasserstein barycenters)**.** *Let $\mathcal{X}$ be a compact subset of a connected d-dimensional Riemannian manifold, and let $\mathbb{P}$ be a probability measure on $\mathcal{P}(\mathcal{X})$. Then, the set of Wasserstein barycenters (i.e., minimizers of* (5.9)*) is non-empty. Moreover, if $\mathbb{P}(\mathcal{P}_{a.c.}(\mathcal{X})) > 0$ where $\mathcal{P}_{a.c.}(\mathcal{X})$ denotes the set of probability measures on $\mathcal{X}$ absolutely continuous with respect to the Riemannian volume, then there exists a unique Wasserstein barycenter.*

To be perfectly rigorous, we should justify that the set $\mathcal{P}_{a.c.}(\mathcal{X}) \subset \mathcal{P}(\mathcal{X})$ is measurable with respect to the weak-* topology. Since we do not want to obscure the discussion, we will only refer the reader to [63, Section 2.2]. Let us also mention that without the assumption $\mathbb{P}(\mathcal{P}_{a.c.}(\mathcal{X})) > 0$, Wasserstein barycenters are not necessarily unique: for instance, if $\mathcal{X} = \mathbb{S}^2$ and $N, S$ denote antipodal points on $\mathcal{X}$, then $\mathbb{P} = \frac{1}{2}(\delta_{\delta_N} + \delta_{\delta_S})$ has multiple barycenters: any measure supported on the equator associated to $N$ and $S$ is a Wasserstein barycenter for $\mathbb{P}$.

*Proof of Theorem 5.6.* Since $\mathcal{X}$ is compact, the set $\mathcal{P}(\mathcal{X})$ is compact by Prokhorov's theorem. Besides, the mapping $\mu \mapsto W_2^2(\rho, \mu)$ is Lipschitz, uniformly over $\rho \in \mathcal{P}(\mathcal{X})$, hence $F_{\mathbb{P}}$ is Lipschitz. It follows that it admits a minimizer, which proves the first part of the statement.

We next show that $F_{\mathbb{P}}$ is a convex function with respect to linear interpolation of measures. Let $\mu_0, \mu_1, \rho \in \mathcal{P}(\mathcal{X})$. We consider the optimal transport plans $\gamma_0$ from $\rho$ to $\mu_0$, and $\gamma_1$ from $\rho$ to $\mu_1$. We set $\mu_s = (1-s)\mu_0 + s\mu_1$ and $\gamma_s = (1-s)\gamma_0 + s\gamma_1$. The transport plan $\gamma_s$ has marginals $\rho$ and $\mu_s$, hence, denoting by dist the Riemannian distance,

$$\begin{aligned} W_2^2(\rho, \mu_s) &\le \int_{\mathcal{X}\times\mathcal{X}} \mathrm{dist}(x,y)^2 \mathrm{d}\gamma_s(x,y) \\ &= (1-s)\int_{\mathcal{X}\times\mathcal{X}} \mathrm{dist}(x,y)^2 \mathrm{d}\gamma_0(x,y) + s\int_{\mathcal{X}\times\mathcal{X}} \mathrm{dist}(x,y)^2 \mathrm{d}\gamma_1(x,y) \\ &= (1-s)W_2^2(\rho,\mu_0) + sW_2^2(\rho,\mu_1) \end{aligned} \tag{5.10}$$

which shows the convexity of $\mu \mapsto W_2^2(\rho, \mu)$, and consequently the convexity of $F_{\mathbb{P}}$, with respect to linear interpolation.

Let us show that the convexity of $F_{\mathbb{P}}$ is strict if $\mathbb{P}(\mathcal{P}_{a.c.}(\mathcal{X})) > 0$. The uniqueness statement follows immediately. We first prove that for $\rho \in \mathcal{P}_{a.c.}(\mathcal{X})$, the application $\mu \mapsto W_2^2(\rho, \mu)$ is strictly convex. Assume $W_2^2(\rho, \mu_s) = (1-s)W_2^2(\rho, \mu_0) + sW_2^2(\rho, \mu_1)$ for some $s \in (0,1)$. Then, $\gamma_s$ has to be the unique optimal transport plan from $\rho$ to $\mu_s$, according to the above inequality. By McCann's theorem (Theorem 4.15), $\gamma_s = (\mathrm{Id}, F_s)_{\#}\rho$ where $F_s$ denotes the optimal transport map from $\rho$ to $\mu_s$. In particular, $\gamma_s$ is supported on the graph of $F_s$. But we also have

$$\gamma_s = (1-s)(\mathrm{Id}, F_0)_{\#}\rho + s(\mathrm{Id}, F_1)_{\#}\rho$$

which means that $\gamma_s$ has mass on the graphs of both $F_0$ and $F_1$. This is possible only if $F_0 = F_1 = F_s$ $\rho$-almost everywhere, which in turn implies $\mu_0 = \mu_1$. We conclude that $\mu \mapsto W_2^2(\rho, \mu)$ is strictly convex when $\rho$ is absolutely continuous. Therefore $F_{\mathbb{P}}$ is strictly convex when $\mathbb{P}(\mathcal{P}_{a.c.}(\mathcal{X})) > 0$, which concludes the proof. □

Practitioners often do not have access exactly to $\mathbb{P}$ and to the measures $\rho \in \mathcal{P}(\mathcal{X})$ in the support of $\mathbb{P}$, but nevertheless they would like to compute an approximate barycenter, as close as possible to the "true" barycenter. In other words, they would like to know if Wasserstein barycenters are stable with respect to perturbations of $\mathbb{P}$.

Like for stability of optimal transport maps, it is possible to show the qualitative stability (also called "consistency") of Wasserstein barycenters using weak compactness arguments. For $p \geq 1$, we consider the Wasserstein distances on $\mathcal{P}(\mathcal{P}(\mathcal{X}))$

$$\mathcal{W}_p(\mathbb{P}, \mathbb{Q}) = \min_{\gamma \in \Pi(\mathbb{P},\mathbb{Q})} \int_{\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})} W_2(\rho, \widetilde{\rho}) \mathrm{d}\gamma(\rho, \widetilde{\rho}). \tag{5.11}$$

If $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\mathcal{X}))$ is the limit of some sequence $\mathbb{P}_n \in \mathcal{P}(\mathcal{P}(\mathcal{X}))$, then any limit of Wasserstein barycenters for $\mathbb{P}_n$ as $n \to +\infty$ is a Wasserstein barycenter for $\mathbb{P}$. A precise statement is the following:

**Theorem 5.7** (Qualitative stability of Wasserstein barycenters)**.** *Let $\mathcal{X}$ be a compact subset of a connected d-dimensional Riemannian manifold. Let $(\mathbb{P}_j)_{j\geq 1} \subset \mathcal{P}(\mathcal{P}(\mathcal{X}))$ be a sequence of probability measures on $\mathcal{P}(\mathcal{X})$ and let $\mu_j$ be a barycenter of $\mathbb{P}_j$, for all $j \in \mathbb{N}$. Assume that $\mathcal{W}_1(\mathbb{P}_j, \mathbb{P}) \to 0$ as $j \to +\infty$, for some $\mathbb{P} \in \mathcal{P}(\mathcal{P}(\mathcal{X}))$. Then, any limit of the sequence $(\mu_j)_{j\geq 1}$ is a Wasserstein barycenter of $\mathbb{P}$.*

The statement is given with a convergence assumption in $\mathcal{W}_1$-norm for coherence with Proposition 5.8 below. We could have assumed equivalently $\mathcal{W}_2(\mathbb{P}_j, \mathbb{P}) \to 0$ since $\mathcal{P}(\mathcal{X})$ is compact.

*Proof of Theorem 5.7.* Let $\mu$ be a limit of the sequence $(\mu_j)_{j\geq 1}$. Up to extraction of a subsequence, which we omit in the notation, we may assume that the full sequence converges to $\mu$. Let $\widetilde{\mu}_j$ be a random measure of distribution $\mathbb{P}_j$, and let $\widetilde{\mu}$ be a random measure of distribution $\mathbb{P}$. The Skorokhod representation theorem allows to choose $\widetilde{\mu}_j$ and $\widetilde{\mu}$ on the same probability space, and in a way that $\widetilde{\mu}_j$ converges almost surely to $\widetilde{\mu}$. For $\nu \in \mathcal{P}(\mathcal{X})$, there holds

$$\begin{aligned}
\mathbb{E} W_2^2(\nu, \widetilde{\mu}) &= \mathcal{W}_2^2(\delta_\nu, \mathbb{P}) \\
&= \lim_{j \to +\infty} \mathcal{W}_2^2(\delta_\nu, \mathbb{P}_j) \qquad \text{since } \mathcal{W}_2(\mathbb{P}_j, \mathbb{P}) \to 0 \\
&= \lim_{j \to +\infty} \mathbb{E} W_2^2(\nu, \widetilde{\mu}_j) \\
&\geq \lim_{j \to +\infty} \mathbb{E} W_2^2(\mu_j, \widetilde{\mu}_j) \qquad \text{since } \mu_j \text{ is a barycenter for } \mathbb{P}_j \\
&\geq \mathbb{E} \liminf_{j \to +\infty} W_2^2(\mu_j, \widetilde{\mu}_j) \qquad \text{using Fatou's lemma} \\
&\geq \mathbb{E} W_2^2(\mu, \widetilde{\mu}).
\end{aligned}$$

In the last line we used the lower semi-continuity of $W_2$, together with the convergence of $\mu_j$ to $\mu$, and the almost sure convergence of $\widetilde{\mu}_j$ to $\widetilde{\mu}$. This proves that $\mu$ is a Wasserstein barycenter for $\mathbb{P}$. □

To go beyond the qualitative result given by Theorem 5.7 and prove stability in a quantitative form, one may first consider the 1-dimensional case. Although $\mathbb{R}$ is not compact, our proof below shows that Wasserstein barycenters on $\mathbb{R}$ exist and are unique, without assuming anything on $\mathbb{P}$.

**Proposition 5.8.** *Let $\mathbb{P}, \mathbb{Q} \in \mathcal{P}(\mathcal{P}_2(\mathbb{R}))$, then there exist unique Wasserstein barycenters $\mu_\mathbb{P}, \mu_\mathbb{Q}$ associated to $\mathbb{P}$ and $\mathbb{Q}$, and*

$$W_2(\mu_\mathbb{P}, \mu_\mathbb{Q}) \leq \mathcal{W}_1(\mathbb{P}, \mathbb{Q}).$$

*Proof.* The optimal transport map $T_\rho$ from the Lebesgue measure $\lambda_{[0,1]}$ on $[0,1]$ to any measure $\rho$ on $\mathbb{R}$ is explicit and verifies $W_2(\rho,\mu) = \|T_\rho - T_\mu\|_{L^2([0,1])}$ according to (1.17), where $L^2([0,1])$ is taken with respect to $\lambda_{[0,1]}$. Hence

$$F_{\mathbb{P}}(\mu) = \frac{1}{2}\int_{\mathcal{P}(\mathbb{R})} \|T_\rho - T_\mu\|^2_{L^2([0,1])} \mathrm{d}\mathbb{P}(\rho) \tag{5.12}$$

Thus, existence and uniqueness of the Wasserstein barycenter come from the existence and uniqueness of barycenters in the Hilbert space $L^2([0,1])$ (which follows from the strict convexity of the squared norm). The right-hand side in (5.12) is minimal for $T_\mu = T_{\mu_{\mathbb{P}}}$ given by

$$x \in (0,1) \mapsto T_{\mu_{\mathbb{P}}}(x) = \int_{\mathcal{P}(\mathbb{R})} T_\rho(x) \mathrm{d}\mathbb{P}(\rho)$$

and therefore $F_{\mathbb{P}}$ is minimal for $\mu = \mu_{\mathbb{P}} = (T_{\mu_{\mathbb{P}}})_{\#}\lambda_{[0,1]}$. For any $\gamma \in \Pi(\mathbb{P},\mathbb{Q})$ we have, using Minkowski's inequality,

$$\begin{aligned} W_2(\mu_{\mathbb{P}},\mu_{\mathbb{Q}}) &= \|T_{\mu_{\mathbb{P}}} - T_{\mu_{\mathbb{Q}}}\|_{L^2([0,1])} \\ &= \|\int_{\mathcal{P}(\mathbb{R})} T_\rho \, \mathrm{d}\mathbb{P}(\rho) - \int_{\mathcal{P}(\mathbb{R})} T_\rho \, \mathrm{d}\mathbb{Q}(\rho)\|_{L^2([0,1])} \\ &= \|\int_{\mathcal{P}(\mathbb{R})\times\mathcal{P}(\mathbb{R})} (T_\rho - T_{\rho'}) \mathrm{d}\gamma(\rho,\rho')\|_{L^2([0,1])} \\ &\leq \int_{\mathcal{P}(\mathbb{R})\times\mathcal{P}(\mathbb{R})} \|T_\rho - T_{\rho'}\|_{L^2([0,1])} \mathrm{d}\gamma(\rho,\rho') \\ &= \int_{\mathcal{P}(\mathbb{R})\times\mathcal{P}(\mathbb{R})} W_2(\rho,\rho') \mathrm{d}\gamma(\rho,\rho'). \end{aligned}$$

Taking the infimum over $\gamma \in \Pi(\mathbb{P},\mathbb{Q})$, this concludes the proof. □

To prove the quantitative stability of Wasserstein barycenters in dimension $d \geq 2$, we assume for simplicity that we are in the Euclidean setting and consider $\mathcal{X}$ a compact subset of $\mathbb{R}^d$. It also seems natural to require uniqueness of the minimizer (otherwise it is hard to give a meaning to "quantitative stability"). Therefore, following Theorem 5.6, we assume $\mathbb{P}(\mathcal{P}_{a.c.}(\mathcal{X})) > 0$, which guarantees uniqueness of the Wasserstein barycenter. Actually, we make the following set of assumptions, which we comment below:

**Assumption 5.9.** *There exist $\alpha_{\mathbb{P}}, c_{\mathbb{P}}, m_{\mathbb{P}}, M_{\mathbb{P}}, \mathrm{per}_{\mathbb{P}} \in (0,+\infty)$, and a measurable set $S_{\mathbb{P}} \subset \mathcal{P}(\mathcal{X})$ such that $\mathbb{P}(S_{\mathbb{P}}) = \alpha_{\mathbb{P}}$ and for all $\rho \in S_{\mathbb{P}}$,*

1. *$\rho \in \mathcal{P}_{a.c.}(\mathcal{X})$,*
2. *the density of $\rho$ on its support is bounded below by $m_{\mathbb{P}}$, and above by $M_{\mathbb{P}}$,*
3. *the $(d-1)$-dimensional Hausdorff measure of the boundary of the support of $\rho$ is bounded above by $\mathrm{per}_{\mathbb{P}}$,*
4. *for any $\psi_0, \psi_1 \in C^0(\mathcal{X})$,*

$$c_{\mathbb{P}} \mathrm{Var}_\rho(\psi_1^* - \psi_0^*) \leq \mathcal{K}_\rho(\psi_1) - \mathcal{K}_\rho(\psi_0) - \langle \psi_1 - \psi_0 \mid -(\nabla\psi_0^*)_{\#}\rho\rangle. \tag{5.13}$$

Point 4. is simply a strong convexity estimate on $\mathcal{K}_\rho$ (or almost, since again in the left-hand side we have $\psi_0^*$ and $\psi_1^*$, and not $\psi_0$ and $\psi_1$). It is written in a different form than the ones we encountered previously. In some sense, instead of $|x-y|^2 \lesssim \langle x-y, \nabla f(x) - \nabla f(y)\rangle$ as in (2.17), we have here $|x-y|^2 \lesssim f(y) - f(x) - \langle y - x \mid \nabla f(x)\rangle$. In finite dimension, these two forms of strong convexity are equivalent; but in infinite dimension (here, in the space of measures) it is less clear, only one implication is obviously true.

It is possible to show that any absolutely continuous $\rho$ with density bounded above and below on its support, and whose support is a John domain, satisfies (5.13). The case where the John domain is actually convex is covered in [26, Appendix B.1], and the case of a general John domain may be shown using in addition the gluing techniques of Chapter 3. However, to avoid discussions that may not be particularly useful and that would divert us from our main purpose, we will not pursue this here. Instead, we state the main result of this section, conditionally on Assumption 5.9.

**Theorem 5.10** (Quantitative stability of Wasserstein barycenters)**.** *Let $\mathbb{P}, \mathbb{Q} \in \mathcal{P}(\mathcal{P}(\mathcal{X}))$ and assume that $\mathbb{P}$ satisfies Assumption 5.9. Let $\mu_\mathbb{P}$ be the Wasserstein barycenter of $\mathbb{P}$, and $\mu_\mathbb{Q}$ be a barycenter of $\mathbb{Q}$. Then there exists a constant $C < +\infty$ depending only on $\alpha_\mathbb{P}, c_\mathbb{P}, m_\mathbb{P}, M_\mathbb{P}, \mathrm{per}_\mathbb{P}$ and the dimension $d$ such that*

$$W_2(\mu_\mathbb{P}, \mu_\mathbb{Q}) \leq C \mathcal{W}_1(\mathbb{P}, \mathbb{Q})^{1/6}. \tag{5.14}$$

*Proof of Theorem 5.10.* We rely on a dual formulation of the minimization problem (5.9). By Kantorovich duality (see Section 2.1.1),

$$\frac{1}{2} W_2^2(\rho, \mu) = \langle \frac{1}{2}|\cdot|^2 \mid \rho\rangle + \langle \frac{1}{2}|\cdot|^2 \mid \mu\rangle - \min_{\psi \in C^0(\mathcal{X})} \mathcal{T}(\psi) \tag{5.15}$$

where

$$\mathcal{T}(\psi) = \int_\mathcal{X} \psi^* \mathrm{d}\rho + \int_\mathcal{Y} \psi \, \mathrm{d}\mu.$$

In the sequel, the subdifferential of a functional $\mathcal{A} : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$ at $\mu \in \mathcal{P}(\mathcal{X})$ is the set $\partial\mathcal{A}(\mu)$ of functions $f \in C^0(\mathcal{X})$ such that

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \qquad \mathcal{A}(\nu) - \mathcal{A}(\mu) \geq \langle f \mid \nu - \mu\rangle.$$

Fixing $\rho$, the function $W_2^2(\rho, \cdot)$ is convex (see (5.10)) and its subdifferential at $\mu \in \mathcal{P}(\mathcal{X})$ is given by

$$\partial \left[\frac{1}{2} W_2^2(\rho, \cdot)\right](\mu) = \left\{\frac{1}{2}|\cdot|^2 - \psi_{\rho\to\mu} \mid \psi_{\rho\to\mu} \in \operatorname*{argmin}_{\psi \in C^0(\mathcal{X})} \mathcal{T}(\psi)\right\}$$

(to see this, use (5.15) and an envelope theorem, particularly easy since each function in the minimum in (5.15) is linear). Integrating against $\mathbb{P}$, the subdifferential of $F_\mathbb{P}$ at $\mu$ verifies:

$$\left\{\int_{\mathcal{P}(\mathcal{X})} \left(\frac{1}{2}|\cdot|^2 - \psi_{\rho\to\mu}\right) \mathrm{d}\mathbb{P}(\rho) \mid \psi_{\rho\to\mu} \in \operatorname*{argmin}_{\psi \in C^0(\mathcal{X})} \mathcal{T}(\psi) \text{ for } \mathbb{P}\text{-a.e. } \rho\right\} \subset \partial F_\mathbb{P}(\mu) \tag{5.16}$$

(looking for a reverse inclusion is useless here, and not immediate since we are in infinite dimension). Thus for any $\mu, \nu$, and for any collection $(\psi_{\rho\to\mu})_{\rho \in \mathcal{P}(\mathcal{X})}$ of Kantorovich potentials we get by definition of the subdifferential of $F_\mathbb{P}$

$$F_\mathbb{P}(\nu) - F_\mathbb{P}(\mu) - \langle \int_{\mathcal{P}(\mathcal{X})} \left(\frac{1}{2}|\cdot|^2 - \psi_{\rho\to\mu}\right) \mathrm{d}\mathbb{P}(\rho), \nu - \mu\rangle \geq 0. \tag{5.17}$$

This is a convexity inequality for $F_{\mathbb{P}}$. We will strengthen it into the following strong convexity estimate:

$$F_{\mathbb{P}}(\nu) - F_{\mathbb{P}}(\mu) - \langle \int_{\mathcal{P}(\mathcal{X})} \left(\frac{1}{2}|\cdot|^2 - \psi_{\rho\to\mu}\right) \mathrm{d}\mathbb{P}(\rho), \nu - \mu\rangle \geq CW_2^6(\mu,\nu). \tag{5.18}$$

Let us show (5.18). For any $\rho$ satisfying the four points in Assumption 5.9, and any $c \in \mathbb{R}$, we have

$$W_2^6(\mu,\nu) \leq \|T_{\rho\to\mu} - T_{\rho\to\nu}\|_{L^2(\rho)}^6 \leq C\|\psi^*_{\rho\to\mu} - \psi^*_{\rho\to\nu} - c\|_{L^2(\rho)}^2.$$

The first inequality is simply (1.15), and the second inequality comes from Lemma 3.16 since $\psi^*_{\rho\to\mu}$ and $\psi^*_{\rho\to\nu}$ are $R_{\mathcal{Y}}$-Lipschitz with $R_{\mathcal{Y}} = \sup_{y\in\mathcal{Y}}|y|$ (see Lemma 2.6). The constant $C$ depends only on the constants in Assumption 5.9. Taking the infimum over $c \in \mathbb{R}$, we recover $\mathrm{Var}_\rho(\psi^*_{\rho\to\mu} - \psi^*_{\rho\to\nu})$ in the right-hand side. From (5.13) we get

$$W_2^6(\mu,\nu) \lesssim \mathcal{K}_\rho(\psi_{\rho\to\mu}) - \mathcal{K}_\rho(\psi_{\rho\to\nu}) - \langle \psi_{\rho\to\mu} - \psi_{\rho\to\nu} \mid \nu\rangle.$$

Due to (5.15), the right-hand side is also equal to

$$\frac{1}{2}W_2^2(\nu,\rho) - \frac{1}{2}W_2^2(\mu,\rho) - \langle \frac{1}{2}|\cdot|^2 - \psi_{\rho\to\mu} \mid \nu - \mu\rangle,$$

and finally integrating against $\mathbb{P}$ we obtain (5.18).

Let $\mu = \mu_{\mathbb{P}}$ and $\nu = \mu_{\mathbb{Q}}$ be the Wasserstein barycenters with respect to $\mathbb{P}$ and $\mathbb{Q}$. Notice that for an appropriate choice of Kantorovich potentials $(\psi_{\rho\to\mu})_{\rho\in\mathcal{P}(\mathcal{X})}$, the integral in the left-hand side of (5.17) vanishes at $\mu = \mu_{\mathbb{P}}$. Indeed, $\mu_{\mathbb{P}}$ is a minimizer of the convex functional $F_{\mathbb{P}}$, thus $0 \in \partial F_{\mathbb{P}}(\mu_{\mathbb{P}})$, given by (5.16). Therefore, applying the strong convexity of $F_{\mathbb{P}}$ at $\mu_{\mathbb{P}}$, we get

$$CW_2^6(\mu_{\mathbb{P}},\mu_{\mathbb{Q}}) \leq F_{\mathbb{P}}(\mu_{\mathbb{Q}}) - F_{\mathbb{P}}(\mu_{\mathbb{P}}).$$

By definition of $\mu_{\mathbb{Q}}$ as a minimizer of $F_{\mathbb{Q}}$, we have $F_{\mathbb{Q}}(\mu_{\mathbb{P}}) - F_{\mathbb{Q}}(\mu_{\mathbb{Q}}) \geq 0$, hence

$$CW_2^6(\mu_{\mathbb{P}},\mu_{\mathbb{Q}}) \leq F_{\mathbb{P}}(\mu_{\mathbb{Q}}) - F_{\mathbb{Q}}(\mu_{\mathbb{Q}}) + F_{\mathbb{Q}}(\mu_{\mathbb{P}}) - F_{\mathbb{P}}(\mu_{\mathbb{P}}) = \langle \frac{1}{2}(W_2^2(\cdot,\mu_{\mathbb{Q}}) - W_2^2(\cdot,\mu_{\mathbb{P}})) \mid \mathbb{P} - \mathbb{Q}\rangle.$$

Since $\frac{1}{2}(W_2^2(\cdot,\mu_{\mathbb{Q}}) - W_2^2(\cdot,\mu_{\mathbb{P}}))$ is $2\mathrm{diam}(\mathcal{X})$-Lipschitz, the Kantorovich-Rubinstein duality formula (2.12) (since $\mathcal{P}(\mathcal{X})$ is a Polish space) yields (5.14). □

## 5.3 Stability of entropic optimal transport and convergence of the Sinkhorn algorithm

Our final application pertains to the numerical resolution of the entropic optimal transport problem

$$\inf_{\gamma\in\Pi(\rho,\mu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\mathrm{d}\gamma(x,y) + \varepsilon H(\gamma \mid \rho\otimes\mu) \tag{5.19}$$

introduced in Section 4.1. We showed in Theorem 4.6 that it is equivalent to the following dual problem:

$$\sup_{(\phi,\psi)\in C^0(\mathcal{X})\times C^0(\mathcal{Y})} \mathcal{J}^\varepsilon(\phi,\psi) \tag{5.20}$$

where

$$\mathcal{J}^\varepsilon(\phi,\psi) = \langle \phi \mid \rho\rangle + \langle \psi \mid \mu\rangle - \varepsilon\langle e^{\frac{\phi\oplus\psi - c}{\varepsilon}} \mid \rho\otimes\mu\rangle.$$

**Sinkhorn's algorithm.** To solve this dual problem, the *Sinkhorn algorithm* proceeds by alternate maximization: starting at an arbitrary $\psi_0 \in C^0(\mathcal{Y})$, it replaces alternatively the first and the second argument of $\mathcal{J}^\varepsilon$ by the maximizers of $\phi \mapsto \mathcal{J}^\varepsilon(\phi, \psi)$ and $\psi \mapsto \mathcal{J}^\varepsilon(\phi, \psi)$ respectively. In other words, it performs the following sequence of iterations:

$$\begin{cases} \phi_{t+1} & \in \underset{\phi \in C^0(\mathcal{X})}{\operatorname{argmax}} \ \mathcal{J}^\varepsilon(\phi, \psi_t) \\ \psi_{t+1} & \in \underset{\psi \in C^0(\mathcal{Y})}{\operatorname{argmax}} \ \mathcal{J}^\varepsilon(\phi_{t+1}, \psi). \end{cases} \tag{5.21}$$

The maxima in the above equations are actually explicit: it follows from Proposition 4.5 that the maximizer in the first line of (5.21) is $\psi_t^{c,\varepsilon}$ where

$$\psi^{c,\varepsilon}(x) = -\varepsilon \log \left( \int_{\mathcal{Y}} e^{\frac{\psi(y) - c(x,y)}{\varepsilon}} \mathrm{d}\mu(y) \right),$$

and the maximizer in the second line is $\phi_{t+1}^{\bar{c},\varepsilon}$ where

$$\phi^{\bar{c},\varepsilon}(y) = -\varepsilon \log \left( \int_{\mathcal{X}} e^{\frac{\phi(x) - c(x,y)}{\varepsilon}} \mathrm{d}\rho(x) \right).$$

Therefore, the sequence of iterations (5.21) is equivalent to

$$\phi_{t+1} = \psi_t^{c,\varepsilon}, \qquad \psi_{t+1} = \phi_{t+1}^{\bar{c},\varepsilon}. \tag{5.22}$$

We saw in Theorem 4.6 that (5.20) admits a solution. Actually, this solution is $(\rho \otimes \mu)$-a.e. unique by the strict concavity of $\mathcal{J}^\varepsilon$, which is a consequence of the strict convexity of the exponential. Moreover, it is a fixed point $(\phi, \psi)$ of the iterations (5.22) (due to Proposition 4.5 again), and therefore it is of the form $(\psi^{c,\varepsilon}, \psi)$, for some $\psi$ satisfying $\psi = (\psi^{c,\varepsilon})^{\bar{c},\varepsilon}$.

**Relation to the strong concavity of $\mathcal{J}^\varepsilon$.** Along the sequence of iterations (5.21) (or equivalently (5.22)), the value of $\mathcal{J}^\varepsilon$ progressively increases. Since $\mathcal{J}^\varepsilon$ is concave, it is expected that these iterations converge to the solution $(\phi, \psi)$ to (5.20) as $t \to +\infty$. The question we will address is the speed of convergence of this algorithm as $t \to +\infty$. As always, convergence speeds of optimization algorithms depend on the convexity/concavity properties near the optimizer of the function which is optimized. Here, it will be related to the *strong concavity* properties of $\mathcal{J}^\varepsilon$.

This is in turn directly related to the stability of Kantorovich potentials with respect to perturbations of the marginals. Let us explain why. Observe that $\nabla_\phi \mathcal{J}^\varepsilon(\phi_{t+1}, \psi_t) = 0$ for any $t$, since $\phi_{t+1}$ is a maximizer of $\phi \mapsto \mathcal{J}^\varepsilon(\phi, \psi_t)$. We saw in (4.11) that this implies $\Pi_{\mathcal{X}\#}\gamma_t = \rho$, where $\gamma_t = e^{\frac{\phi_{t+1} \oplus \psi_t - c}{\varepsilon}} \rho \otimes \mu$ is the optimal entropic transport plan associated to the pair of potentials $(\phi_{t+1}, \psi_t)$. However, $\Pi_{\mathcal{Y}\#}\gamma_t$ is not equal in general to $\mu$ since the maximizer of $\psi \mapsto \mathcal{J}^\varepsilon(\phi_{t+1}, \psi)$ is $\psi_{t+1}$, and not $\psi_t$. In other words, $\gamma_t \notin \Pi(\rho, \mu)$. But the first marginal of $\gamma_t$ is nevertheless equal to $\rho$, and this is where we reconnect with our leitmotif: using this observation, some proofs of convergence of Sinkhorn's algorithm (for instance in [29]) implement an argument based on the stability of the optimal entropic plan with respect to the second marginal. This stability is directly related to the strong concavity of $\mathcal{J}^\varepsilon$.

The arguments presented here are slightly different: we will leverage directly the strong concavity of $\mathcal{J}^\varepsilon$, and will not need to establish the stability of the optimal entropic plan with respect to perturbations of the marginals. However, all these phenomena are totally related!

There exists a vast literature on the convergence of Sinkhorn's algorithm. We chose to focus only on one recent result, for which the techniques presented in the previous chapters are particulary well suited; other approaches exist, and we shall mention some of them in the bibliographical notes of Section 5.5.

**Main result.** To study the convergence of Sinkhorn's algorithm, we introduce the following quantity:

$$E(\psi) = \max_{\phi \in C^0(\mathcal{X})} \mathcal{J}^\varepsilon(\phi, \psi).$$

By definition, $E(\psi_t) = \mathcal{J}^\varepsilon(\phi_{t+1}, \psi_t)$. The function $E$ admits a $\mu$-a.e. unique maximizer denoted by $\psi_{\text{opt}}$, due to the $(\rho \otimes \mu)$-a.e. uniqueness of the maximizer of $\mathcal{J}^\varepsilon$. The main result of this section, due to [33], is the following:

**Theorem 5.11.** *Assume that $\mathcal{X}$ is convex, and that there exists $\xi \in \mathbb{R}_+$ such that for all $y \in \mathcal{Y}$, $x \mapsto c(x, y)$ is $\xi$-semi-concave, and*

$$\|c\|_{\text{osc}} := \sup_{x \in \mathcal{X}, y \in \mathcal{Y}} c(x, y) - \inf_{x \in \mathcal{X}, y \in \mathcal{Y}} c(x, y) < +\infty.$$

*Then for any integer $t \geq 0$, the Sinkhorn iterates $(\psi_t)_{t \geq 0}$ defined in* (5.21) *satisfy*

$$E(\psi_{\text{opt}}) - E(\psi_{t+1}) \leq (1 - \alpha\varepsilon^2)(E(\psi_{\text{opt}}) - E(\psi_t)) \tag{5.23}$$

*for some $\alpha > 0$ (independent of $\varepsilon$) provided either one of the following additional assumptions holds:*

- *The domain $\mathcal{X}$ is compact, the measure $\rho$ admits a density $f$ bounded above and below by positive constants on $\mathcal{X}$.*
- *There exists a convex function $V : \mathcal{X} \to \mathbb{R}$ such that $\rho = e^{-V(x)}$ and $D^2 V \geq \xi \mathrm{Id}$.*

The rest of this section is devoted to the proof of Theorem 5.11, following [33]. The main ingredients are Propositions 5.12 and 5.13. We denote by

$$\delta_t = E(\psi_{\text{opt}}) - E(\psi_t)$$

the quantity appearing in (5.23). The first ingredient is a lower bound on $\delta_t - \delta_{t+1}$, in terms of a variance with respect to the *target* measure $\mu$ (whereas in the previous chapters the variances were usually taken with respect to $\rho$):

**Proposition 5.12.** *For any $t \geq 0$, the Sinkhorn iterates $(\psi_s)_{s \geq 0}$ defined in* (5.21) *satisfy*

$$\delta_t \leq 2\sqrt{\varepsilon^{-1}\mathrm{Var}_\mu(\psi_{\text{opt}} - \psi_t)(\delta_t - \delta_{t+1})} + \frac{14c_\infty}{3}\varepsilon^{-1}(\delta_t - \delta_{t+1}). \tag{5.24}$$

The above lower bound on $\delta_t - \delta_{t+1}$ can be "closed" thanks to the following upper bound on the variance in terms of $\delta_t$, which expresses the strong concavity of $E$:

**Proposition 5.13.** *Under the assumptions of Theorem 5.11, for any $t \geq 0$, the Sinkhorn iterates $(\psi_s)_{s \geq 0}$ defined in* (5.21) *satisfy*

$$\mathrm{Var}_\mu(\psi_{\text{opt}} - \psi_t) \leq C_1 \varepsilon^{-1} \delta_t \tag{5.25}$$

*for some $C_1 > 0$ independent of $\varepsilon$ (taken $< 1$).*

Before showing Propositions 5.12 and 5.13, we explain how to prove Theorem 5.11 using these two propositions. Depending on which one of the two terms in the right-hand side of (5.24) is the largest, we are in one of the two following cases:

*Case 1:* If $\delta_t \leq 4\sqrt{\varepsilon^{-1}\mathrm{Var}_\mu(\psi_{\text{opt}} - \psi_t)(\delta_t - \delta_{t+1})}$, then we plug (5.25) into this bound, we square both sides, divide by $\delta_t$ and rearrange the terms. We get $\delta_{t+1} \leq (1 - \alpha\varepsilon^2)\delta_t$.

*Case 2:* If $\delta_t \leq \frac{28c_\infty}{3}\varepsilon^{-1}(\delta_t - \delta_{t+1})$, then rearranging the terms yields $\delta_{t+1} \leq (1-\alpha\varepsilon^2)\delta_t$. In both cases, we have established Theorem 5.11.

We turn to the proof of the above two propositions. Recall that in Section 4.2, we introduced the regularized Kantorovich functional $\mathcal{K}_\rho^\varepsilon$, which depends on an auxiliary measure $\sigma$. Here, we use this same functional, with $\mu$ in place of $\sigma$. In other words, in this section,

$$\mathcal{K}_\rho^\varepsilon(\psi) = \int_\mathcal{X} \psi^{c,\varepsilon} \mathrm{d}\rho$$

where recall that

$$\psi^{c,\varepsilon}(x) = -\varepsilon \log\left(\int_\mathcal{Y} e^{\frac{\psi(y)-c(x,y)}{\varepsilon}} \mathrm{d}\mu(y)\right).$$

The functional $\mathcal{K}_\rho^\varepsilon$ is concave. We reintroduce the notation

$$\mu_\varepsilon^x[\psi] := \hat{\mu}_\varepsilon^x[\psi]\mathrm{d}\mu, \qquad \hat{\mu}_\varepsilon^x[\psi](y) := \frac{e^{\frac{\psi(y)-c(x,y)}{\varepsilon}}}{\int_\mathcal{Y} e^{\frac{\psi(z)-c(x,z)}{\varepsilon}} \mathrm{d}\mu(z)}$$

and $\mu_\varepsilon[\psi] = \int_\mathcal{X} \mu_\varepsilon^x[\psi]\mathrm{d}\rho(x)$, i.e.,

$$\langle v \mid \mu_\varepsilon[\psi]\rangle = \int_\mathcal{X} \langle v \mid \mu_\varepsilon^x[\psi]\rangle \mathrm{d}\rho(x)$$

for any $v \in C^0(\mathcal{X})$. This notation already appeared in Section 4.2, with $\sigma$ in place of $\mu$. Notice that $\mu_\varepsilon[\psi]$ is the second marginal of the probability measure

$$\exp\left(\frac{\psi^{c,\varepsilon}(x) + \psi(y) - c(x,y)}{\varepsilon}\right) \rho \otimes \mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}).$$

In particular, $\mu_\varepsilon[\psi_{\mathrm{opt}}] = \mu$ (see Theorem 4.6). The family $(\mu_\varepsilon^x[\psi])_{x\in\mathcal{X}}$ is simply the disintegration of $\mu_\varepsilon[\psi]$ with respect to $\rho$.

Finally, in the sequel,

$$\mathrm{KL}(\alpha \mid \beta) = \int_\mathcal{X} \frac{\mathrm{d}\alpha}{\mathrm{d}\beta} \log\Big(\frac{\mathrm{d}\alpha}{\mathrm{d}\beta}\Big) \mathrm{d}\beta$$

denotes the Kullback-Leibler divergence. Compared to $H$ introduced in (4.2), it satisfies

$$\mathrm{KL}(\alpha \mid \beta) = 1 + H(\alpha \mid \beta)$$

for any probability measures $\alpha, \beta$ on $\mathcal{X}$. Let us recall the Donsker-Varadhan variational representation of the Kullback-Leibler divergence: for any probability measures $\alpha, \beta$,

$$\mathrm{KL}(\alpha \mid \beta) = \sup_{h \in C_b} \{\langle h \mid \alpha\rangle - \log\langle e^h \mid \beta\rangle\} \tag{5.26}$$

or equivalently by the Fenchel-Moreau theorem since KL is convex:

$$\forall h, \qquad \log\langle e^h \mid \beta\rangle = \sup_{\alpha \ll \beta} \{\langle h \mid \alpha\rangle - \mathrm{KL}(\alpha \mid \beta)\} \tag{5.27}$$

The strong concavity of $\mathcal{K}_\rho^\varepsilon$ is proved (and used) only in the proof of Proposition 5.13. However, we start with the proof of Proposition 5.12.

*Proof of Proposition 5.12.* One may check that for any $\psi$,

$$(\psi^{c,\varepsilon})^{\bar{c},\varepsilon} - \psi = \varepsilon \log \frac{\mathrm{d}\mu}{\mathrm{d}\mu_\varepsilon[\psi]}. \tag{5.28}$$

As a consequence

$$\mathrm{KL}(\mu \mid \mu_\varepsilon[\psi]) = \frac{1}{\varepsilon}\langle (\psi^{c,\varepsilon})^{\bar{c},\varepsilon} - \psi \mid \mu\rangle.$$

In particular,

$$\delta_t - \delta_{t+1} = E(\psi_{t+1}) - E(\psi_t) \geq \mathcal{J}^\varepsilon(\phi_{t+1}, \psi_{t+1}) - \mathcal{J}^\varepsilon(\phi_{t+1}, \psi_t) = \varepsilon \mathrm{KL}(\mu \mid \mu_\varepsilon[\psi_t]). \tag{5.29}$$

Since $E$ is concave, letting $v = \psi_{\mathrm{opt}} - \psi_t$ we have $E(\psi_{\mathrm{opt}}) \leq E(\psi_t) + \frac{\mathrm{d}}{\mathrm{d}r} E(\psi_t + rv)\Big|_{r=0}$. Since

$$\frac{\mathrm{d}}{\mathrm{d}r} E(\psi_t + rv)\Big|_{r=0} = \langle v \mid \mu\rangle + \frac{\mathrm{d}}{\mathrm{d}r}\mathcal{K}^\varepsilon_\rho(\psi_t + rv)\Big|_{r=0} = \langle v \mid \mu - \mu_\varepsilon[\psi_t]\rangle$$

according to (4.23), we get

$$\delta_t \leq \langle \psi_{\mathrm{opt}} - \psi_t \mid \mu - \mu_\varepsilon[\psi_t]\rangle.$$

We deduce that for any $\eta > 0$,

$$\begin{aligned}
\delta_t &\leq \eta^{-1}\Big(\langle \eta(\psi_{\mathrm{opt}} - \psi_t) \mid \mu - \mu_\varepsilon[\psi_t]\rangle - \mathrm{KL}(\mu \mid \mu_\varepsilon[\psi_t])\Big) + \eta^{-1}\mathrm{KL}(\mu \mid \mu_\varepsilon[\psi_t]) \\
&\leq \eta^{-1} \sup_{\mu' \in \mathcal{P}(\mathcal{X})} \Big(\langle \eta(\psi_{\mathrm{opt}} - \psi_t) \mid \mu' - \mu_\varepsilon[\psi_t]\rangle - \mathrm{KL}(\mu' \mid \mu_\varepsilon[\psi_t])\Big) + \eta^{-1}\mathrm{KL}(\mu \mid \mu_\varepsilon[\psi_t]) \\
&= \eta^{-1}\log \mathbb{E}_{\mu_\varepsilon[\psi_t]}(\exp(\eta f)) + \eta^{-1}\mathrm{KL}(\mu \mid \mu_\varepsilon[\psi_t])
\end{aligned} \tag{5.30}$$

where $f(y) = \psi_{\mathrm{opt}} - \psi_t - \mathbb{E}_{\mu_\varepsilon[\psi_t]}(\psi_{\mathrm{opt}} - \psi_t)$. The last line follows from (5.27) and the fact that

$$\langle \eta(\psi_{\mathrm{opt}} - \psi_t) \mid \mu' - \mu_\varepsilon[\psi_t]\rangle = \langle \eta f \mid \mu'\rangle.$$

Recall the Bernstein inequality: if $Z \leq b$ almost surely, then

$$\mathbb{E}(e^{\lambda(Z - \mathbb{E}(Z))}) \leq \exp\left(\frac{\frac{\lambda^2}{2}\mathbb{E}(Z^2)}{1 - \frac{b\lambda}{3}}\right) \qquad \text{for all } \lambda \in [0, 3/b).$$

Since $f$ is contained in an interval of length at most $2\|c\|_{\mathrm{osc}}$ (due to Lemma 4.8), we get that for any $\eta \in (0, \frac{3}{2\|c\|_{\mathrm{osc}}})$,

$$\log \mathbb{E}_{\mu_\varepsilon[\psi_t]}(\exp(\eta f)) \leq \frac{\eta^2 \mathrm{Var}_{\mu_\varepsilon[\psi_t]}(\psi_{\mathrm{opt}} - \psi_t)}{2(1 - \eta\frac{2\|c\|_{\mathrm{osc}}}{3})}.$$

Plugging into (5.30) we find

$$\begin{aligned}
\delta_t &\leq \inf_{\eta \in (0, \frac{3}{2\|c\|_{\mathrm{osc}}})} \left\{ \frac{\eta \mathrm{Var}_{\mu_\varepsilon[\psi_t]}(\psi_{\mathrm{opt}} - \psi_t)}{2(1 - \eta\frac{2\|c\|_{\mathrm{osc}}}{3})} + \eta^{-1}\mathrm{KL}(\mu \mid \mu_\varepsilon[\psi_t]) \right\} \\
&\leq \sqrt{2\mathrm{Var}_{\mu_\varepsilon[\psi_t]}(\psi_{\mathrm{opt}} - \psi_t)\mathrm{KL}(\mu \mid \mu_\varepsilon[\psi_t])} + \frac{2\|c\|_{\mathrm{osc}}}{3}\mathrm{KL}(\mu \mid \mu_\varepsilon[\psi_t])
\end{aligned}$$

where the final inequality follows by optimizing over $\eta$. And combining with (5.29) we get

$$\delta_t \leq \sqrt{2\varepsilon^{-1}\mathrm{Var}_{\mu_\varepsilon[\psi_t]}(\psi_{\mathrm{opt}} - \psi_t)(\delta_t - \delta_{t+1})} + \frac{2\|c\|_{\mathrm{osc}}}{3}\varepsilon^{-1}(\delta_t - \delta_{t+1}). \tag{5.31}$$

In order to complete the proof, we need to replace the variance with respect to $\mu_\varepsilon[\psi_t]$ by a variance with respect to $\mu$. To make this change, we use the following inequality, which follows from the variational representation for the Hellinger distance and its comparison to the Kullback-Leibler divergence (see [33, Appendix A.4] for a proof): for any mutually absolutely continuous measures $\nu_1, \nu_2$ on $\mathcal{X}$, and for any function $f : \mathcal{X} \to [a, b]$, there holds

$$\mathrm{Var}_{\nu_1}(f) \geq \frac{1}{2}\mathrm{Var}_{\nu_2}(f) - (b-a)^2 \min(\mathrm{KL}(\nu_1 \mid \nu_2), \mathrm{KL}(\nu_2 \mid \nu_1)). \tag{5.32}$$

Since $\|\psi_{\mathrm{opt}} - \psi_t\|_{\mathrm{osc}} \leq 2\|c\|_{\mathrm{osc}}$ (again due to Lemma 4.8), we have

$$\begin{aligned}\mathrm{Var}_{\mu_\varepsilon[\psi_t]}(\psi_{\mathrm{opt}} - \psi_t) &\leq 2\mathrm{Var}_\mu(\psi_{\mathrm{opt}} - \psi_t) + 8\|c\|_{\mathrm{osc}}^2 \mathrm{KL}(\mu \mid \mu_\varepsilon[\psi_t]) \\ &\leq 2\mathrm{Var}_\mu(\psi_{\mathrm{opt}} - \psi_t) + 8\|c\|_{\mathrm{osc}}^2 \varepsilon^{-1}(\delta_t - \delta_{t+1}).\end{aligned}$$

Plugging into (5.31) and using $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, we get (5.24). □

*Proof of Proposition 5.13.* We have, as in (4.23),

$$\frac{\mathrm{d}}{\mathrm{d}r}\mathcal{K}_\rho^\varepsilon(\psi + rv) = -\langle v \mid \mu_\varepsilon[\psi + rv]\rangle$$

Choosing $v = \psi_{\mathrm{opt}} - \psi_t$ and $r = 1$, we have $\mu_\varepsilon[\psi + rv] = \mu_\varepsilon[\psi_{\mathrm{opt}}] = \mu$, hence

$$\begin{aligned}\delta_t &= E(\psi_{\mathrm{opt}}) - E(\psi_t) \\ &= \mathcal{J}^\varepsilon(\psi_{\mathrm{opt}}^{c,\varepsilon}, \psi_{\mathrm{opt}}) - \mathcal{J}^\varepsilon(\psi_t^{c,\varepsilon}, \psi_t) \\ &= \langle \psi_{\mathrm{opt}}^{c,\varepsilon} \mid \rho\rangle + \langle \psi_{\mathrm{opt}} \mid \mu\rangle - \langle \psi_t^{c,\varepsilon} \mid \rho\rangle - \langle \psi_t \mid \mu\rangle \\ &= \mathcal{K}_\rho^\varepsilon(\psi_{\mathrm{opt}}) - \mathcal{K}_\rho^\varepsilon(\psi_t) - \frac{\mathrm{d}}{\mathrm{d}r}\mathcal{K}_\rho^\varepsilon(\psi_t + rv)\Big|_{r=1} \\ &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}r}\mathcal{K}_\rho^\varepsilon(\psi_t + rv)\mathrm{d}r - \frac{\mathrm{d}}{\mathrm{d}r}\mathcal{K}(\psi_t + rv)\Big|_{r=1} \\ &= -\int_0^1 \int_r^1 \frac{\mathrm{d}^2}{\mathrm{d}s^2}\mathcal{K}_\rho^\varepsilon(\psi_t + sv)\mathrm{d}s\mathrm{d}r.\end{aligned} \tag{5.33}$$

From second to third line we used the fact that the term $\langle e^{\frac{\phi \oplus \psi - c}{\varepsilon}} \mid \rho \otimes \mu\rangle$ is equal to 1 both for $(\psi, \psi) = (\psi_{\mathrm{opt}}^{c,\varepsilon}, \psi_{\mathrm{opt}})$ and $(\phi, \psi) = (\psi_t^{c,\varepsilon}, \psi_t)$, due to the first-order optimality condition (4.11). We now prove the key concavity estimate

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}\mathcal{K}_\rho^\varepsilon(\psi + sv) \leq -C\mathrm{Var}_{\mu_\varepsilon[\psi+sv]}(v) \tag{5.34}$$

where $C > 0$ is some constant independent of $\varepsilon \in (0, 1)$. We recall from (4.24) the identity

$$\begin{aligned}-\frac{\mathrm{d}^2}{\mathrm{d}s^2}\mathcal{K}_\rho^\varepsilon(\psi + sv) &= \frac{1}{\varepsilon}\mathbb{E}_{x\sim\rho}(\mathrm{Var}_{\mu_\varepsilon^x[\psi+sv]}(v)) \\ &= \frac{1}{\varepsilon}(\mathrm{Var}_{\mu_\varepsilon[\psi+sv]}(v) - \mathrm{Var}_{x\sim\rho}(\mathbb{E}_{\mu_\varepsilon^x[\psi+sv]}(v))\end{aligned}$$

(in the second line we use the law of total variance). We plug this into the inequality (4.25):

$$\begin{aligned}-C_0^2\frac{\mathrm{d}^2}{\mathrm{d}s^2}\mathcal{K}_\rho^\varepsilon(\psi + sv) &\geq \mathrm{Var}_{x\sim\rho}(\mathbb{E}_{\mu_\varepsilon^x[\psi+sv]}(v)) \\ &= \varepsilon\frac{\mathrm{d}^2}{\mathrm{d}s^2}\mathcal{K}_\rho^\varepsilon(\psi + sv) + \mathrm{Var}_{\mu_\varepsilon[\psi+sv]}(v).\end{aligned}$$

This proves (5.34).

In (5.34), the variance is taken with respect to $\mu_\varepsilon[\psi + sv]$, whereas we are looking for a variance with respect to $\mu$. Again, we rely on (5.32). Plugging into (5.33), we get

$$\begin{aligned} C^{-1}\delta_t &\geq \int_0^1 \int_r^1 \frac{1}{2}\mathrm{Var}_\mu(v) - (2\|c\|_{\mathrm{osc}})^2 \mathrm{KL}(\mu \mid \mu_\varepsilon[\psi_t + sv])\mathrm{d}s\mathrm{d}r \\ &= \frac{1}{4}\mathrm{Var}_\mu(v) - 4\|c\|_{\mathrm{osc}}^2 \int_0^1 \int_r^1 \mathrm{KL}(\mu \mid \mu_\varepsilon[\psi_t + sv])\mathrm{d}s\mathrm{d}r. \end{aligned} \tag{5.35}$$

Next we handle the double integral as follows: using (5.28), the optimality of $\psi_{\mathrm{opt}}$ and the concavity of $E$, we get

$$\begin{aligned} \mathrm{KL}(\mu \mid \mu[\psi_t + sv]) &= \frac{1}{\varepsilon}\langle ((\psi_t + sv)^{c,\varepsilon})^{\bar{c},\varepsilon} - (\psi_t + sv), \mu\rangle \\ &= \frac{1}{\varepsilon}\left(\mathcal{J}^\varepsilon\left((\psi_t + sv)^{c,\varepsilon}, ((\psi_t + sv)^{c,\varepsilon})^{\bar{c},\varepsilon}\right) - \mathcal{J}^\varepsilon((\psi_t + sv)^{c,\varepsilon}, \psi_t + sv)\right) \\ &\leq \frac{1}{\varepsilon}\left(\mathcal{J}^\varepsilon(\psi_{\mathrm{opt}}^{c,\varepsilon}, \psi_{\mathrm{opt}}) - \mathcal{J}^\varepsilon((\psi_t + sv)^{c,\varepsilon}, \psi_t + sv)\right) \\ &= \frac{1}{\varepsilon}(E(\psi_{\mathrm{opt}}) - E((1-s)\psi_t + s\psi_{\mathrm{opt}})) \\ &\leq \frac{1-s}{\varepsilon}(E(\psi_{\mathrm{opt}}) - E(\psi_t)) \\ &= \frac{1-s}{\varepsilon}\delta_t. \end{aligned}$$

Plugging into (5.35) we get (5.25). □

**Remark 5.14** (Relation to matrix factorization)**.** *When $\rho$ and $\mu$ are discrete measures, the algorithm* (5.21) *is in fact a reformulation, using a logarithmic change of variables, of the original Sinkhorn algorithm [98]. Originally, Sinkhorn's algorithm takes as input a square matrix $A$ with non-negative entries, and returns a factorization $A = D_1 B D_2$, where $D_1$ and $D_2$ are diagonal matrices, and $B$ is doubly stochastic. The algorithm runs by normalizing alternatively the rows and the columns of $A$ so that after each step, either each row sums to* 1*, or each column sums to* 1*. For this, one simply divides each element by the sum of the elements in its row (and the step after, one divides each element by the sum of the elements in its column).*

*Let us explain the relation of the iterations* (5.21) *with this algorithm. Let $\rho = \sum_{i=1}^N \rho_i \delta_{x_i}$ and $\mu = \sum_{j=1}^N \mu_j \delta_{y_j}$ be two non-negative discrete measures, and let $c_{ij} = c(x_i, y_j)$ be a cost function. Then the solution to the (primal) entropic optimal transport* (5.19) *is given by*

$$\gamma_{ij} = e^{\frac{\phi_i + \psi_j - c_{ij}}{\varepsilon}} \rho_i \mu_j$$

*for some $\phi, \psi \in \mathbb{R}^N$ (see Theorem 4.6). The iterates* (5.22) *read*

$$\phi_i^{(t+1)} = -\varepsilon \log\left(\sum_{j=1}^N e^{\frac{\psi_j^{(t)} - c_{ij}}{\varepsilon}} \mu_j\right), \qquad \psi_j^{(t+1)} = -\varepsilon \log\left(\sum_{i=1}^N e^{\frac{\phi_i^{(t)} - c_{ij}}{\varepsilon}} \rho_i\right).$$

*and the associated transport plan is*

$$\gamma^{(t)} = e^{\frac{\phi_i^{(t)} + \psi_j^{(t)} - c_{ij}}{\varepsilon}} \rho_i \mu_j.$$

*Performing the change of variables* $u_i^{(t)} = e^{\frac{\phi_i^{(t)}}{\varepsilon}}\rho_i$, $v_j^{(t)} = e^{\frac{\psi_j^{(t)}}{\varepsilon}}\mu_j$ *and* $K_{ij} = e^{-\frac{c_{ij}}{\varepsilon}}$, *the iterations become*

$$\begin{cases} u_i^{(t+1)} &= \frac{\rho_i}{(Kv^{(t)})_i} \\ v_j^{(t+1)} &= \frac{\mu_j}{(K^\top u^{(t+1)})_j} \\ \gamma^{(t)} &= \mathrm{diag}(v^{(t)})K\mathrm{diag}(u^{(t)}), \end{cases}$$

*where* $\mathrm{diag}(x)$ *denotes the square diagonal matrix with entries* $x_i$. *The transport plan* $\gamma^{(t)}$ *converges toward a matrix with satisfies* $\sum_{j=1}^N \gamma_{ij} = \rho_i$ *for any* $i$, *and* $\sum_{i=1}^N \gamma_{ij} = \mu_j$ *for any* $j$. *In particular, if* $\rho_i = \mu_j = 1$ *for any* $i, j$, *then the limits of the diagonal matrices* $\mathrm{diag}(v^{(t)})$ *and* $\mathrm{diag}(u^{(t)})$ *as* $t \to +\infty$ *provide a factorization of the matrix* $K$ *into a doubly stochastic matrix* $\gamma$.

## 5.4 Perspectives

Beside the open problems already pointed out in this document, we mention two other open directions of research that we find relevant. Many others exist!

First, it would be interesting to extend qualitative and quantitative stability results to the many variants of optimal transport which have been studied over the years (and used in applications!), in particular:

- Multi-marginal optimal transport. It is a generalization of optimal transport to more than two probability distributions, relevant in some physical applications such as density functional theory. We refer to the survey [88].
- Unbalanced optimal transport compares arbitrary positive measures, not restricted to probability distributions. It allows mass to be created or destroyed at some cost, which is much more realistic in many scenarios than classical optimal transport. For instance, it can match biological cell populations with different sample sizes. This generalization of optimal transport makes it more robust to outliers and missing data, and thus relevant for applications. We refer to one of the first papers on the topic [30], and to the survey [97].
- Partial optimal transport. It is a generalization of optimal transport where only a fixed fraction of the mass is transported (and it is thus related to unbalanced optimal transport). More precisely, given non-negative $f, g \in L^1$, one seeks to transport a fraction $m \leq \min(\|f\|_{L^1}, \|g\|_{L^1})$ of the mass of $f$ onto $g$ as cheaply as possible. Introduced in the mathematical community by Caffarelli and McCann in [24], it has found applications in imaging, for instance in image retrieval [95].
- Sliced optimal transport has been introduced in [94] and is based on the computation of the Wasserstein distances between all 1-dimensional projections of the two measures under consideration. The associated distance, called sliced Wasserstein distance, is given by

  $$SW_2(\rho,\mu) = \left( \fint_{\mathbb{S}^{d-1}} W_2^2((\pi_e)_\#\rho, (\pi_e)_\#\mu) \mathrm{d}\mathscr{H}^{d-1}(e) \right)^{1/2}$$

  where $\pi_e$ is the orthogonal projection onto the line $\mathbb{R}e$, and $\mathscr{H}^{d-1}$ denotes the Hausdorff measure on the sphere $\mathbb{S}^{d-1}$. It is particularly simple to approximate numerically, and it compares well to the Wasserstein distance:

  $$\mathrm{SW}_2 \leq W_2 \leq C\mathrm{SW}_2^\beta$$

  for $\beta = 1/(2(d+1))$ and some $C > 0$ depending on the dimension and on the radius of a large ball containing the support of the measures (see [16, Chapter 5]).

One could also wonder why the transport between two measures would always need to be optimal in some sense! And indeed, there exist many well-known non-optimal couplings between pairs of probability measures, whose study even preceded that of optimal transport. To name only two:

- the Knothe-Rosenblatt rearrangement. For two probability measures in $\mathbb{R}^d$, it consists in rearranging monotonically the marginal distributions of the first coordinate, and then the conditional distributions, iteratively. We refer for instance to [106, Chapter 1] for a precise description. One advantage of the Knothe-Rosenblatt rearrangement is its low computational cost, due to its simple 1-dimensional structure. However, it is unsuitable for many applications due to its strong dependency on the order in which the dimensions are handled.

- The Moser coupling. It has been used by Moser in [83] to show that between any two manifolds endowed with normalized volume forms there exists a differentiable homeomorphism matching the volume forms. Interpreting the normalized volumes as probability densities, the homeomorphism can be equivalently seen as a transport map. It has the advantage of being obtained as the solution to an elliptic equation, with the associated regularity properties. We refer to [105, Chapter 1, Appendix] for a precise description.

The stability of these couplings with respect to perturbations of the marginal measures is an open problem too. Interest in this problem is motivated by the fact that non-optimal transport maps between measures are a fundamental tool in many recent developments in machine learning, sampling and imaging.

## 5.5 Bibliographical notes

§5.1: Statistical optimal transport is a very active field, and giving an exhaustive list of references would be impossible. We prefer to refer to the nice book recently written on the topic by Sinho Chewi, Jonathan Niles-Weed and Philippe Rigollet [32]. We also recommand the survey [7]. One of the first papers on the subject, to which we borrowed Theorem 5.5, is [59]. The semidual estimation approach was then systematically explored in [47]. An important reference regarding plug-in estimators is [75]. Theorem 5.1 and Theorem 5.2 are proved in [6], which also discusses the importance of stability bounds in this field. Regarding computationally tractable estimators of the optimal transport map, a good reference is [104]. Another family of estimators which we did not discuss are those derived from the entropic optimal transport map, see for instance [90]. Such estimators have also been used in [91] which is one of very few works dealing with the estimation of discontinuous optimal transport maps.

§5.2: Wasserstein barycenters were introduced in the work of Agueh and Carlier [1], as a generalization of McCann's theory of interpolation [77] to more than two measures. They established in this work foundational results such as existence and duality.

The computational aspects of Wasserstein barycenters were first considered in [94], where Rabin, Peyré, Delon and Bernot introduced the sliced Wasserstein to compute them fastly, at least in low dimensions. Slightly later, Cuturi and Doucet [39] applied entropic optimal transport to barycenters, making them computationally practical. We refer to [89, Chapter 9.2] for an account on Wasserstein barycenters from the numerical point of view.

More general existence and uniqueness statements than Theorem 5.6 exist, see for instance [68, Theorem 2]. Theorem 5.7 is a weak version of Theorem 3 in [68]. Theorem 5.6 is Theorem 3.1 in [63]. The statement of Proposition 5.8 comes from [26, Section 1.2.2].

The book chapter [32, Chapter 8] is devoted to statistical perspectives on Wasserstein barycenters and references gathered in [32, Chapter 8.4].

§5.3: The Sinkhorn algorithm was invented in 1964 [98] in the context of matrix factorization. In the literature, it bears various names, for instance the iterative proportional fitting procedure (IPFP), the RAS algorithm in economics, or the matrix scaling algorithm in computer science. Cuturi discovered in 2013 [38] that it could be used to solve fastly the entropic optimal transport problem, and thus to infer an approximate solution to the *unregularized* optimal transport problem. To address the convergence of Sinkhorn's algorithm, one possibility is to use the Hilbert projective metric, see [89, Section 4.2]. Another method relying on couplings and stochastic control has been recently developed in [29]. The approach presented here, due to [33], relies on the convexity properties of the (regularized) Kantorovich functional. There is actually a very large corpus on the convergence of Sinkhorn's algorithm, see the detailed literature reviews in [33, Section 1.1] and [29, Section 1.2]. Sinkhorn's algorithm iterates over entropic Kantorovich *potentials*, but there is also some literature (not related to Sinkhorn's algorithm) on the stability of entropic optimal transport *maps*. These maps, which are defined as barycentric projections of optimal entropic plans, are not necessarily admissible transport maps (they do not transport $\rho$ to $\mu$) but they are nevertheless natural objects. They converge to the unregularized transport map when it exists, as the regularization parameter tends to 0 (under mild additional assumptions). We refer to [46] for interesting results in this direction.

We did not mention in this chapter other interesting applications of optimal transport stability theory. For instance, in [58], stability estimates have been used to derive precise asymptotics for the random matching problem with quadratic cost in dimension $d \geq 3$. They have also been helpful to give a constructive proof of the existence of global-in-time weak solutions of the 3-dimensional incompressible semi-geostrophic equations [17].

## References


[1] Martial Agueh, and Guillaume Carlier. "Barycenters in the Wasserstein space." SIAM Journal on Mathematical Analysis 43.2 (2011): 904-924.

[2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. Gradient flows: in metric spaces and in the space of probability measures. Springer Science & Business Media, 2008.

[3] Luigi Ambrosio, Federico Glaudo, and Dario Trevisan. "On the optimal map in the 2-dimensional random matching problem." Discrete & Continuous Dynamical Systems: Series A 39.12 (2019).

[4] Charles J. Amick. "Some remarks on Rellich's theorem and the Poincaré inequality." Journal of the London Mathematical Society 2.1 (1978): 81-93.

[5] Alexandr Andoni, Assaf Naor, and Ofer Neiman. "Snowflake universality of Wasserstein spaces." Annales Scientifiques de l'Ecole Normale Superieure. Vol. 51. No. 3. 2018.

[6] Sivaraman Balakrishnan, and Tudor Manole. "Stability bounds for smooth optimal transport maps and their statistical implications." arXiv preprint arXiv:2502.12326 (2025).

[7] Sivaraman Balakrishnan, Tudor Manole, and Larry Wasserman. "Statistical inference for optimal transport maps: Recent advances and perspectives." arXiv preprint arXiv:2506.19025 (2025).

[8] Davide Barilari, and Luca Rizzi. "Sub-Riemannian interpolation inequalities." Inventiones mathematicae 215.3 (2019): 977-1038.

[9] Saurav Basu, Soheil Kolouri, and Gustavo K. Rohde. "Detecting and visualizing cell phenotype differences from microscopy images using transport-based morphometry." Proceedings of the National Academy of Sciences 111.9 (2014): 3448-3453.

[10] Frank Bauer, Matthias Keller, and Radosław K. Wojciechowski. "Cheeger inequalities for unbounded graph Laplacians." Journal of the European Mathematical Society 17.2 (2015): 259-271.

[11] Robert J. Berman. "Convergence rates for discretized Monge–Ampère equations and quantitative stability of optimal transport." Foundations of Computational Mathematics 21.4 (2021): 1099-1140.

[12] Sergei Bobkov and Michel Ledoux. "From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities." Geometric and Functional Analysis, Vol. 10 (2000) 1028-1052.

[13] Bogdan Bojarski. "Remarks on Sobolev imbedding inequalities." Complex Analysis Joensuu 1987: Proceedings of the XIIIth Rolf Nevanlinna-Colloquium, held in Joensuu, Finland, Aug. 10–13, 1987. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.

[14] Jan Boman. "$L^p$-estimates for very strongly elliptic systems". Department of Mathematics, University of Stockholm, Sweden, 1982. Available at `https://mathoverflow.net/questions/320172/famous-but-unavailable-paper-of-jan-boman`

[15] Nicolas Bonneel, Julien Rabin, Gabriel Peyré, Hanspeter Pfister. "Sliced and Radon Wasserstein barycenters of measures." Journal of Mathematical Imaging and Vision 51 (2015): 22-45.

[16] Nicolas Bonnotte. Unidimensional and evolution methods for optimal transportation. Diss. Université Paris Sud-Paris XI; Scuola normale superiore (Pise, Italie), 2013.

[17] David P. Bourne, Charlie P. Egan, Beatrice Pelloni, and Mark Wilkinson. "Semi-discrete optimal transport methods for the semi-geostrophic equations." Calculus of Variations and Partial Differential Equations 61.1 (2022): 39.

[18] Herm Jan Brascamp, and Elliott H. Lieb. "On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation." Journal of Functional Analysis 22.4 (1976): 366-389.

[19] Yann Brenier. "Décomposition polaire et réarrangement monotone des champs de vecteurs." CR Acad. Sci. Paris Sér. I Math. 305 (1987): 805-808.

[20] Yann Brenier. "Polar factorization and monotone rearrangement of vector-valued functions." Communications on Pure and Applied Mathematics 44.4 (1991): 375-417.

[21] Stephen M. Buckley, and Pekka Koskela. "Sobolev-Poincaré implies John." Mathematical Research Letters 2.5 (1995): 577-593.

[22] Stephen M. Buckley, Pekka Koskela, and Guozhen Lu. "Boman equals John." Proceedings of the XVI Rolf Nevanlinna Colloquium. W. de Gruyter, 1996.

[23] Luis A. Caffarelli. "The regularity of mappings with a convex potential." Journal of the American Mathematical Society 5.1 (1992): 99-104.

[24] Luis A. Caffarelli, and Robert J. McCann. "Free boundaries in optimal transport and Monge-Ampère obstacle problems." Annals of mathematics (2010): 673-730.

[25] Guillaume Carlier, Alex Delalande, and Quentin Mérigot. "Quantitative stability of the pushforward operation by an optimal transport map." Foundations of Computational Mathematics (2024): 1-28.

[26] Guillaume Carlier, Alex Delalande and Quentin Mérigot. "Quantitative stability of barycenters in the Wasserstein space." Probability Theory and Related Fields 188.3 (2024): 1257-1286.

[27] Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. "Boundary measures for geometric inference". Foundations of Computational Mathematics, 10(2):221–240, 2010.

[28] Jeff Cheeger. "A lower bound for the smallest eigenvalue of the Laplacian." Problems in Analysis (1970): 195-199.

[29] Alberto Chiarini, Giovanni Conforti, Giacomo Greco, and Luca Tamanini. "A semiconcavity approach to stability of entropic plans and exponential convergence of Sinkhorn's algorithm." arXiv preprint arXiv:2412.09235 (2024).

[30] Lénaïc Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. "Unbalanced optimal transport: Dynamic and Kantorovich formulations." Journal of Functional Analysis 274.11 (2018): 3090-3123.

[31] Victor Chernozhukov, Alfred Galichon, Marc Hallin, and Marc Henry. "Monge-Kantorovich depth, quantiles, ranks and signs." Annals of Statistics 45.1 (2017): 223-256.

[32] Sinho Chewi, Jonathan Niles-Weed, and Philippe Rigollet. "Statistical optimal transport." arXiv preprint arXiv:2407.18163 (2024).

[33] Lénaïc Chizat, Alex Delalande, Tomas Vaškevičius. "Sharper exponential convergence rates for Sinkhorn's algorithm in continuous settings." Mathematical Programming (2025).

[34] Fan RK Chung. Spectral graph theory. Vol. 92. American Mathematical Soc., 1997.

[35] Dario Cordero-Erausquin, Robert J. McCann, and Michael Schmuckenschläger. "A Riemannian interpolation inequality à la Borell, Brascamp and Lieb." Inventiones mathematicae 146.2 (2001): 219-257.

[36] Dario Cordero-Erausquin, Robert J. McCann, and Michael Schmuckenschläger. "Prékopa–Leindler type inequalities on Riemannian manifolds, Jacobi fields, and optimal transport." Annales de la Faculté des sciences de Toulouse: Mathématiques. Vol. 15. No. 4. 2006.

[37] Richard Courant, and David Hilbert. Methoden der mathematischen Physik, Vol. 2, Springer, Berlin, 1937.

[38] Marco Cuturi. "Sinkhorn distances: Lightspeed computation of optimal transport." Advances in neural information processing systems 26 (2013).

[39] Marco Cuturi, and Arnaud Doucet. "Fast computation of Wasserstein barycenters." International conference on machine learning. PMLR, 2014.

[40] Alex Delalande. Quantitative stability in quadratic optimal transport. Diss. Université Paris-Saclay, 2022.

[41] Alex Delalande. "Nearly tight convergence bounds for semi-discrete entropic optimal transport." International Conference on Artificial Intelligence and Statistics. PMLR, 2022.

[42] Alex Delalande and Quentin Mérigot. "Quantitative stability of optimal transport maps under variations of the target measure." Duke Mathematical Journal 172.17 (2023): 3321-3357.

[43] George Deligiannidis, Valentin De Bortoli, and Arnaud Doucet. "Quantitative uniform stability of the iterative proportional fitting procedure." The Annals of Applied Probability 34.1A (2024): 501-516.

[44] Guido De Philippis, and Alessio Figalli. "The Monge–Ampère equation and its link to optimal transportation." Bulletin of the American Mathematical Society 51.4 (2014): 527-580.

[45] Emmanuele DiBenedetto. Real analysis. Springer Science+Business Media, LLC. Vol. 13. Boston: Birkhäuser, 2002.

[46] Vincent Divol, Jonathan Niles-Weed, and Aram-Alexandre Pooladian. "Tight stability bounds for entropic Brenier maps." International Mathematics Research Notices 2025.7 (2025): rnaf078.

[47] Vincent Divol, Jonathan Niles-Weed, and Aram-Alexandre Pooladian. "Optimal transport map estimation in general function spaces." The Annals of Statistics 53.3 (2025): 963-988.

[48] Ronen Eldan. "Thin shell implies spectral gap up to polylog via a stochastic localization scheme." Geometric and Functional Analysis 23.2 (2013): 532-569.

[49] Herbert Federer. Geometric measure theory. Springer-Verlag Berlin Heidelberg New York, 1969.

[50] Alessio Figalli, and Ludovic Rifford. "Mass transportation on sub-Riemannian manifolds." Geometric and functional analysis 20.1 (2010): 124-159.

[51] Nicolas Fournier, and Arnaud Guillin. "On the rate of convergence in Wasserstein distance of the empirical measure." Probability theory and related fields 162.3 (2015): 707-738.

[52] Anatole Gallouët, Quentin Mérigot, and Boris Thibert. "Strong c-concavity and stability in optimal transport." arXiv preprint arXiv:2207.11042 (2022).

[53] 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 (2011): 401-409.

[54] Piotr Hajłasz, and Pekka Koskela. "Isoperimetric inequalities and imbedding theorems in irregular domains." Journal of the London Mathematical Society 58.2 (1998): 425-450.

[55] Piotr Hajłasz, and Pekka Koskela. Sobolev met Poincaré. Vol. 688. American Mathematical Soc., 2000.

[56] Rainer Hempel, Luis A. Seco, and Barry Simon. "The essential spectrum of Neumann Laplacians on some bounded singular domains." Journal of Functional Analysis 102.2 (1991): 448-483.

[57] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. Fundamentals of convex analysis'. Springer, 2004.

[58] Martin Huesmann, Michael Goldman, and Dario Trevisan. "Asymptotics for random quadratic transportation costs." arXiv preprint arXiv:2409.08612 (2024).

[59] Jan-Christian Hütter, and Philippe Rigollet. "Minimax estimation of smooth optimal transport maps." The Annals of Statistics, 49(2) (2021): 1166-1194.

[60] Fritz John. "Rotation and strain." Communications on Pure and Applied Mathematics 14.3 (1961): 391-413.

[61] Leonid Kantorovich. "On the translocation of masses." Dokl. Akad. Nauk. USSR 37 (1942), 199–201. English translation in Management science 5.1 (1958): 1-4.

[62] Matthias Keller, Daniel Lenz, and Radoslaw K. Wojciechowski. Graphs and discrete Dirichlet spaces. Vol. 358. Berlin, Germany: Springer, 2021.

[63] Young-Heon Kim, and Brendan Pass. "Wasserstein barycenters over Riemannian manifolds." Advances in Mathematics 307 (2017): 640-683.

[64] Jun Kitagawa, Cyril Letrouit, and Quentin Mérigot. "Stability of optimal transport maps on Riemannian manifolds." ArXiv preprint arXiv:2504.05412.

[65] Boaz Klartag. Regularity through convexity in high dimensions. Lecture notes available at `https://www.weizmann.ac.il/math/klartag/sites/math.klartag/files/uploads/klartag-notes-by-brazitikos.pdf`

[66] Alexander V. Kolesnikov, and Emanuel Milman. "Brascamp–Lieb-type inequalities on weighted Riemannian manifolds with boundary." The Journal of Geometric Analysis 27.2 (2017): 1680-1702.

[67] Soheil Kolouri, and Gustavo K. Rohde. "Transport-based single frame super resolution of very low resolution face images." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[68] Thibaut Le Gouic, and Jean-Michel Loubes. "Existence and consistency of Wasserstein barycenters." Probability Theory and Related Fields 168 (2017): 901-917.

[69] László Leindler. "On a certain converse of Hölder's inequality." In Linear Operators and Approximation (Proc. Conf., Oberwolfach 1971). Birkhäuser, Basel, 1972.

[70] Christian Léonard. "A survey of the Schrödinger problem and some of its connections with optimal transport." Discrete and Continuous Dynamical Systems 34.4 (2013): 1533-1574.

[71] Dorian Le Peutrec. "On Witten Laplacians and Brascamp–Lieb's inequality on manifolds with boundary." Integral Equations and Operator Theory 87.3 (2017): 411-434.

[72] Cyril Letrouit, and Quentin Mérigot. "Gluing methods for quantitative stability of optimal transport maps." arXiv preprint arXiv:2411.04908 (2024).

[73] Wenbo Li, and Ricardo H. Nochetto. "Quantitative stability and error estimates for optimal transport plans." IMA Journal of Numerical Analysis 41.3 (2021): 1941-1965.

[74] John Lott, and Cédric Villani. "Ricci curvature for metric-measure spaces via optimal transport." Annals of Mathematics (2009): 903-991.

[75] Tudor Manole, Sivaraman Balakrishnan, Jonathan Niles-Weed and Larry Wasserman. "Plugin estimation of smooth optimal transport maps." The Annals of Statistics 52.3 (2024): 966-998.

[76] Olli Martio, and Jukka Sarvas. "Injectivity theorems in plane and space." Annales Fennici Mathematici 4.2 (1979): 383-401.

[77] Robert J. McCann. "A convexity principle for interacting gases." Advances in mathematics 128.1 (1997): 153-179.

[78] Robert J. McCann. "Polar factorization of maps on Riemannian manifolds." Geometric & Functional Analysis 11.3 (2001): 589-608.

[79] Quentin Mérigot, Alex Delalande, and Frederic Chazal. "Quantitative stability of optimal transport maps and linearization of the 2-Wasserstein space." International Conference on Artificial Intelligence and Statistics. PMLR, 2020.

[80] Gaspard Monge. "Mémoire sur la théorie des déblais et des remblais." In Histoire de l'Académie Royale des Sciences de Paris (1781), pp. 666–704.

[81] Octave Mischler and Dario Trevisan. "Quantitative stability in optimal transport for general power costs." arXiv preprint arXiv:2407.19337 (2024).

[82] Shahriar Mokhtari-Sharghi. "Cheeger inequality for infinite graphs." Geometriae Dedicata 100: 53–64 (2003).

[83] Jürgen Moser. “On the volume elements on a manifold.” Transactions of the American Mathematical Society 120.2 (1965): 286-294.

[84] Jonathan Niles-Weed, and Quentin Berthet. “Minimax estimation of smooth densities in Wasserstein distance.” The Annals of Statistics 50.3 (2022): 1519-1540.

[85] Marcel Nutz. “Introduction to entropic optimal transport.” Lecture notes, Columbia University (2021).

[86] Shin-ichi Ohta. “Gradient flows on Wasserstein spaces over compact Alexandrov spaces.” American journal of mathematics 131.2 (2009): 475-516.

[87] Felix Otto. “The geometry of dissipative evolution equations: the porous medium equation.” Comm. Partial Differential Equations 26 (2001): 101-174.

[88] Brendan Pass. “Multi-marginal optimal transport: theory and applications.” ESAIM: Mathematical Modelling and Numerical Analysis 49.6 (2015): 1771-1790.

[89] Gabriel Peyré, and Marco Cuturi. “Computational optimal transport: With applications to data science.” Foundations and Trends in Machine Learning 11.5-6 (2019): 355-607.

[90] Aram-Alexandre Pooladian, and Jonathan Niles-Weed. “Entropic estimation of optimal transport maps.” arXiv preprint arXiv:2109.12004 (2021).

[91] Aram-Alexandre Pooladian, Vincent Divol, and Jonathan Niles-Weed. “Minimax estimation of discontinuous optimal transport maps: The semi-discrete case.” International Conference on Machine Learning. PMLR, 2023.

[92] András Prékopa. “On logarithmic concave measures and functions.” Acta Scientiarum Mathematicarum 34 (1973): 335-343.

[93] Neil Pritchard, and Thomas Weighill. “Coarse embeddability of Wasserstein space and the space of persistence diagrams.” Discrete & Computational Geometry (2024): 1-16.

[94] Julien Rabin, Gabriel Peyré, Julie Delon and Marc Bernot. “Wasserstein barycenter and its application to texture mixing.” International conference on scale space and variational methods in computer vision. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[95] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. “The Earth mover’s distance as a metric for image retrieval.” International Journal of Computer Vision 40.2 (2000): 99-121.

[96] Filippo Santambrogio. Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling. Progress in Nonlinear Differential Equations and Their Applications, Vol. 87. Birkhäuser, 2015.

[97] Thibault Séjourné, Gabriel Peyré, and François-Xavier Vialard. “Unbalanced optimal transport, from theory to numerics.” Handbook of Numerical Analysis 24 (2023): 407-471.

[98] Richard Sinkhorn. “A relationship between arbitrary positive matrices and doubly stochastic matrices.” The Annals of mathematical statistics 35.2 (1964): 876-879.

[99] Daniel Spielman. Spectral and Algebraic Graph Theory. Ongoing draft available at `http://cs-www.cs.yale.edu/homes/spielman/sagt/`

[100] Sanvesh Srivastava, Cheng Li, and David B. Dunson. "Scalable Bayes via barycenter in Wasserstein space." Journal of Machine Learning Research 19.8 (2018): 1-35.

[101] Justin Solomon, Fernando de Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, Leonidas Guibas. "Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains." ACM Transactions on Graphics (ToG) 34.4 (2015): 1-11.

[102] Karl-Theodor Sturm. "On the geometry of metric measure spaces. I and II." (2006): 65-131 and 133-177.

[103] Luca Trevisan. Lecture notes on graph partitioning, expanders and spectral methods. Available at `https://lucatrevisan.github.io/books/expanders-2016.pdf`

[104] Adrien Vacher, Boris Muzellec, Alessandro Rudi, Francis Bach and François-Xavier Vialard. "A dimension-free computational upper-bound for smooth optimal transport estimation." Conference on Learning Theory. PMLR, 2021.

[105] Cédric Villani. Topics in optimal transportation. Graduate studies in mathematics, Vol. 58. American Mathematical Soc., 2021.

[106] Cédric Villani. Optimal transport: old and new. Grundlehren der mathematischen Wissenschaften, Vol. 338. Berlin: Springer, 2009.

[107] Wei Wang, Dejan Slepčev, Saurav Basu, John A. Ozolek, and Gustavo K. Rohde. "A linear optimal transportation framework for quantifying and visualizing variations in sets of images." International Journal of Computer Vision 101 (2013): 254-269.

[108] Hassler Whitney. "Analytic extensions of differentiable functions defined in closed sets." Transactions of the American Mathematical Society 36.1 (1934): 63-89.