HAL open science

Convergence rate of general entropic optimal transport costs

Guillaume Carlier, Paul Pegon, Luca Tamanini

▶ To cite this version:

Guillaume Carlier, Paul Pegon, Luca Tamanini. Convergence rate of general entropic optimal transport costs. Calculus of Variations and Partial Differential Equations, 2023, 62 (4), pp.116. 10.1007/s00526-023-02455-0 . hal-03689945

HAL Id: hal-03689945

https://hal.science/hal-03689945v1

Submitted on 7 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Convergence rate of general entropic optimal transport costs

Guillaume Carlier * Paul Pegon † Luca Tamanini ‡

We investigate the convergence rate of the optimal entropic cost v_ε to the optimal transport cost as the noise parameter $\varepsilon \downarrow 0$. We show that for a large class of cost functions c on $\mathbb{R}^d \times \mathbb{R}^d$ (for which optimal plans are not necessarily unique or induced by a transport map) and compactly supported and L^∞ marginals, one has $v_\varepsilon - v_0 = \frac{d}{2}\varepsilon \log(1/\varepsilon) + O(\varepsilon)$. Upper bounds are obtained by a block approximation strategy and an integral variant of Alexandrov's theorem. Under an infinitesimal twist condition on c, i.e. invertibility of $\nabla^2_{xy} c(x, y)$, we get the lower bound by establishing a quadratic detachment of the duality gap in d dimensions thanks to Minty's trick.

Keywords. optimal transport, entropic regularization, Schrödinger problem, convex analysis, entropy dimension.
2020 Mathematics Subject Classification. Primary: 49Q22 ; Secondary: 49N15, 94A17, 49K40.

Contents

1 Introduction **2**

2 Regularity of the EOT cost **5**

3 Upper bound **12**
3.1 Upper bound for $\mathscr{C}^{0,1}$ costs 12
3.2 Upper bound for $\mathscr{C}^{1,1}$ costs and L^∞ marginals 15

*CEREMADE, Université Paris-Dauphine, Université PSL, CNRS, Mokaplan, Inria Paris, 75016 Paris, France. email: carlier@ceremade.dauphine.fr
†CEREMADE, Université Paris-Dauphine, Université PSL, CNRS, Mokaplan, Inria Paris, 75016 Paris, France. email: pegon@ceremade.dauphine.fr
‡Department of Decision Sciences and BIDSA, Bocconi University, Via Roberto Sarfatti 25, 20100 Milano MI, Italy. email: luca.tamanini@unibocconi.it

4 Lower bound for infinitesimally twisted costs **20**
4.1 Proof of the lower bound 20
4.2 Quantitative stability of optimal plans for the quadratic cost 23

Notations

$x, \boldsymbol{x}$	generic points of $\mathbb{R}^d$ and of $\mathbb{R}^d \times \mathbb{R}^d$ respectively;
$\lvert\cdot\rvert$	Euclidean norm on $\mathbb{R}^d$;
$\lVert\cdot\rVert$	norm on $\mathbb{R}^d \times \mathbb{R}^d$ defined by $\lVert\boldsymbol{x}\rVert = \max\{\lvert x\rvert, \lvert y\rvert\}$ if $\boldsymbol{x} = (x, y)$;
$B_r(x), B_r(\boldsymbol{x})$	open ball of radius r centered at $x \in \mathbb{R}^d$ or $\boldsymbol{x} \in (\mathbb{R}^d)^2$ for the above norms;
$\mathscr{C}^{0,1}(X)$	space of real-valued Lipschitz functions on X which is a subset of $\mathbb{R}^d$ or $(\mathbb{R}^d)^2$;
$[f]_{\mathscr{C}^{0,1}(X)}$	Lipschitz constant of $f : X \to \mathbb{R}$ where X is a subset of $\mathbb{R}^d$ or $(\mathbb{R}^d)^2$ for the above norms;
$\mathscr{C}^{1,1}_{\mathrm{loc}}(\Omega)$	space of differentiable real-valued functions on Ω, an open subset of $\mathbb{R}^d$ or $(\mathbb{R}^d)^2$, with locally Lipschitz gradient;
$\mathscr{P}(\mathbb{R}^d)$	space of probability measures on $\mathbb{R}^d$;
S_μ	support of the measure μ;
$M_d(\mathbb{R})$	space of real matrices of size $d \times d$, endowed with the Frobenius norm induced by the scalar product $A \cdot B := \mathrm{Tr}(A^T B)$, for $A, B \in M_d(\mathbb{R})$;
$S_d(\mathbb{R})$	subspace of real symmetric matrices of size $d \times d$.

1 Introduction

We consider two probability measures $\mu^\pm$ compactly supported in $X^\pm \subseteq \mathbb{R}^d$ and a cost function $c : X^- \times X^+ \to \mathbb{R}$. The *Entropic Optimal Transport problem* (also called *entropy-regularized optimal transport problem*, EOT for short) reads as:

$$v_\varepsilon := \inf \left\{ \int_{X^- \times X^+} c(x, y)\ \mathrm{d}\gamma(x, y) + \varepsilon \mathrm{Ent}(\gamma \,|\, \mu^- \otimes \mu^+) \right\} \tag{1.1}$$

where the infimum is taken among all *couplings* γ between μ^- and μ^+, i.e. probability measures having μ^- and μ^+ as marginals. The classical optimal transport (OT) problem corresponds to $\varepsilon = 0$. The penalization term Ent is the Boltzmann-Shannon relative entropy (also called Kullback-Leibler divergence) and $\varepsilon > 0$ can be interpreted as a temperature parameter. Heuristically, this consists in moving μ^- onto μ^+ in the cheapest and (at the same time) most "diffuse" way, since the (unique) minimizer γ_ε of (1.1) is forced to be absolutely continuous with respect to $\mu^- \otimes \mu^+$ because of the entropy term and so its mass has to be "spread out", in contrast with solutions to the unperturbed transport problem.

In the last decade, this class of problems has witnessed a rapidly increasing interest and is now an extremely active research topic, because it has found numerous ap-

plications and proved to be an efficient way to approximate OT problems, especially from a computational viewpoint. Indeed, when it comes to solving EOT by alternating Kullback-Leibler projections on the two marginal constraints, by the algebraic properties of the entropy such iterative projections correspond to the celebrated Sinkhorn's algorithm [Sin64], applied in this framework in the pioneering works [Cut13; Ben+15]. The simplicity and the good convergence guarantees (see [FL89; MG20]) of this method compared to the algorithms used for the Monge-Kantorovich problem, then determined the success of EOT for applications in machine learning, statistics, image processing, language processing and other areas (see the monograph [PC19] and references therein).

As it appears clearly from (1.1), EOT is a perturbed transport problem and, as such, it is natural to investigate its behaviour as the parameter ε vanishes. In this direction, several aspects deserve to be studied, such as the convergence of optimal values, potentials and optimal plans, possibly with quantitative rates. Leveraging on a large deviations interpretation and on the notion of (c, ε)-cyclical monotonicity, the two last questions have been addressed very recently in [BGN22; NW21].

As concerns the convergence of the optimal values (also called *entropic costs*) denoted by v_ε, let us mention earlier contributions. In the pioneering works [Mik04; MT08; Léo12] (which tackled the question from the Schrödinger problem's viewpoint) and [Car+17], the Γ-convergence of the EOT problem towards the unregularized OT problem was proved in the quadratic case, i.e. for $c(x, y) = |x - y|^2$. As a consequence, the optimal value of the quadratic EOT problem converges to the optimal value of the quadratic OT problem, namely the squared Wasserstein distance.

Since then, this convergence result was generalized in two directions: at the level of the accuracy and at the level of the cost function. As for the former, in [Ada+11; EMR15; Pal19] the first-order asymptotic expansion for the cost is established, both in a pointwise and Γ-convergence sense. A further improvement has been obtained independently in [CT21; Chi+20], where also the second-order term in the expansion is determined (under a regularity assumption on the Wasserstein geodesic connecting the two marginals). A second-order expansion in the same spirit has been obtained for semi-discrete OT in [ANS22]. As regards the second direction (more general costs), the Γ-convergence result actually holds true for a very large class of continuous cost functions, but the first-order expansion is much more difficult to extend. In [Pal19], this is achieved for costs satisfying strong regularity assumptions (roughly speaking, a uniform ellipticity condition on the Hessian of the Kantorovich potential of the associated OT problem).

The aim of this paper is to continue in the second direction, namely to weaken even further the assumptions on the cost function c. First of all, it is folklore that v_ε is a $\mathscr{C}^\infty$ function of $\varepsilon > 0$ (and even analytic as we shall see in Theorem 2.1). However, the differentiability of v_ε at $\varepsilon = 0^+$ is equivalent to the existence of a finite entropy solution for v_0, which is generally false, apart from the discrete case. Instead, one expects $v_\varepsilon - v_0$ to be of the order of $\varepsilon \log(1/\varepsilon)$. We shall prove that it is indeed the case and that the remainder is $O(\varepsilon)$ under quite general assumptions, one of which, inspired by [MTW05; MPW12], is the *infinitesimal twist condition* which requires the cross-derivative $\nabla^2_{xy} c(x, y)$ to be invertible for every (x, y). It is worth stressing that for

costs satisfying this condition, the solutions to the associated OT problem need not be concentrated on a graph and may even fail to be unique, see [MPW12, Example 3.1]. Our main findings may be summarized as:

Theorem 1.1. *Let $\Omega^{\pm} \subseteq \mathbb{R}^d$ be open convex sets and $\mu^{\pm}$ be absolutely continuous probability measures compactly supported on $\Omega^{\pm}$ and with L^∞ densities. Assume that the cost c is $\mathscr{C}^2$ on $\Omega^- \times \Omega^+$ and infinitesimally twisted. Then*

$$v_\varepsilon = v_0 + \frac{d}{2}\varepsilon \log(1/\varepsilon) + O(\varepsilon),$$

v_ε being the value of (1.1).

Hence v_ε approximates the transport cost v_0 with accuracy $\frac{d}{2}\varepsilon\log(1/\varepsilon)$ and the next term in the expansion is at most of order ε ; it cannot be better in general, as shown in the cases handled in [Pal19]. As an application, if we debias the problem, namely if we consider the Sinkhorn divergence (see [RTC17])

$$\mathrm{OT}_\varepsilon(\mu^-, \mu^+) := v_\varepsilon(\mu^-, \mu^+) - \frac{1}{2}(v_\varepsilon(\mu^-, \mu^-) + v_\varepsilon(\mu^+, \mu^+))$$

instead of the entropic cost v_ε, it immediately follows from our main theorem that the term in $\varepsilon\log(1/\varepsilon)$ disappears and therefore

$$\mathrm{OT}_\varepsilon(\mu^-, \mu^+) = v_0(\mu^-, \mu^+) + O(\varepsilon).$$

For the proof of Theorem 1.1 we will show separately that for $\varepsilon \ll 1$,

$$v_\varepsilon \leq v_0 + \frac{d}{2}\varepsilon\log(1/\varepsilon) + M\varepsilon \qquad \text{and} \qquad v_\varepsilon \geq v_0 + \frac{d}{2}\varepsilon\log(1/\varepsilon) - m\varepsilon, \tag{1.2}$$

for some constants m, M.

For the **upper bound**, the proof relies on the block approximation introduced in [Car+17] and on an integral variant of Alexandrov's theorem tailored to our purpose (see Lemma 3.6). Let us emphasize that this upper bound only requires $\mathscr{C}^{1,1}$ regularity and no twist condition. We also treat more general situations where the cost function is only Lipschitz, in which case the constant $d/2$ has to be replaced by the *upper entropy dimension* (without the $1/2$ factor) of the marginals (see Proposition 3.1), and this is shown to be sharp in Example 3.3.

For the **lower bound**, from the dual formulation of (1.1) it follows:

$$v_\varepsilon \geq v_0 - \varepsilon \log \int_{\mathbb{R}^d\times\mathbb{R}^d} e^{-\frac{E(x,y)}{\varepsilon}}\, \mathrm{d}\mu^-(x)\, \mathrm{d}\mu^+(y),$$

where $E(x,y) := c(x,y) - \phi(x) - \psi(y)$ is the duality gap and (ϕ, ψ) are Kantorovich potentials for the OT problem. By using the so-called Minty's trick as presented in [MPW12] (here the infinitesimal twist condition on c is required), we are able to prove that E detaches quadratically from the set $\{E = 0\}$ and this allows us to estimate the previous integral in the desired way. We also explain in the quadratic case how to derive a quantitative stability result for optimal plans from the same trick.

Related literature

The discrete case with finitely supported marginals falls within the class of finite-dimensional linear programming problems. In this framework, the choice of the entropy as regularizing function has a long tradition and a detailed investigation of convergence rates for regularized primal/dual formulations and expansion of the cost was already addressed in the early work [CM94].

In the continuous setting, entropic regularization is a more recent technique [Car+17] and has been initially employed in the quadratic case $c(x, y) = |x - y|^2$, where EOT appears as a perturbation of the (squared) Wasserstein distance. For this choice of cost function, EOT is intimately linked (and actually equivalent in the Euclidean setting) to a much older problem, known as Schrödinger problem (SP for short, see the survey [Léo14]). The latter can be better described as a control problem for Brownian particles, hence a dynamical problem, unlike EOT which is static in nature. This dynamical aspect turns out to be very useful to overcome regularity problems affecting Wasserstein geodesics, as already shown in [GT21; Gen+20]. The fact that, for $c(x, y) = |x-y|^2$, EOT and SP coincide (see [GT20]) explains why the quadratic EOT problem has attracted a lot of attention coming from different research areas: not only the ones mentioned before, where EOT has proved to be a valuable tool, but also (stochastic) control, statistical mechanics and many others.

In the very last years, several generalizations of both EOT and SP have appeared. As concerns the latter, extensions to mean-field and interacting particle systems [Bac+20; CCT21], and to an abstract framework [MTV20] are worth mentioning. As regards EOT, regularizing functions other than the entropy (and their impact on algorithms) have been considered for instance in [DG20; LM22]. Moreover, adding entropy penalizations in the spirit of EOT has also been used in the study of multi-marginal OT [Car22; BCN19; MG20], incompressible fluids [Arn+20; BM20] and unbalanced OT [BL21], where the entropy term gives rise to a branching unbalanced OT problem.

For an account of latest developments let us finally mention a series of recent works [GNB21; EN21; NW22], where stability with respect to marginals of plans and potentials for EOT is addressed.

Structure of the paper

In Section 2, we recall some main features of the OT and EOT problems and prove the analyticity of v_ε (Theorem 2.1). The upper and lower bounds on v_ε are established respectively in Section 3 (Proposition 3.1 and Proposition 3.4) and Section 4 (Proposition 4.4).

2 Regularity of the EOT cost

Let us first collect all notations, definitions and relevant results concerning the optimal transport problem and its entropy-regularized counterpart. As already anticipated in the introduction, given a parameter $\varepsilon > 0$, two measures $\mu^\pm \in \mathscr{P}(\mathbb{R}^d)$ with compact

support $X^{\pm} := S_{\mu^{\pm}}$ and a continuous cost function $c \in \mathscr{C}(X^- \times X^+)$, the EOT problem is given by

$$v_\varepsilon := \inf_{\gamma \in \Pi(\mu^-,\, \mu^+)} \left\{ \int_{X^- \times X^+} c(x,y)\ \mathrm{d}\gamma(x,y) + \varepsilon \mathrm{Ent}(\gamma \,|\, \mu^- \otimes \mu^+) \right\}, \tag{EOT$_\varepsilon$}$$

where $\Pi(\mu^-,\ \mu^+)$ denotes the set of couplings between μ^- and μ^+ and $\mathrm{Ent}(\cdot \,|\, \mu^- \otimes \mu^+)$ the Boltzmann-Shannon relative entropy (or Kullback-Leibler divergence) w.r.t. the product measure $\mu^- \otimes \mu^+$, defined for general probability measures p, q as

$$\mathrm{Ent}(p \,|\, q) = \begin{cases} \displaystyle\int_{\mathbb{R}^d} \rho \log(\rho)\ \mathrm{d}q & \text{if } p = \rho q, \\ +\infty & \text{otherwise.} \end{cases}$$

The fact that q is a probability measure ensures that $\mathrm{Ent}(p \,|\, q) \geq 0$. The value of (EOT$_\varepsilon$), denoted by v_ε, is called *entropic cost*.

The dual problem of (EOT$_\varepsilon$) in the sense of convex analysis reads as

$$\sup_{\substack{\phi \in \mathscr{C}(X^-) \\ \psi \in \mathscr{C}(X^+)}} \left\{ \int_{X^-} \phi\ \mathrm{d}\mu^- + \int_{X^+} \psi\ \mathrm{d}\mu^+ - \varepsilon \int_{X^- \times X^+} e^{\frac{\phi(x)+\psi(y)-c(x,y)}{\varepsilon}}\ \mathrm{d}\mu^-(x)\, \mathrm{d}\mu^+(y) + \varepsilon \right\},$$

which is invariant by $(\phi, \psi) \mapsto (\phi + \lambda, \psi - \lambda)$ where $\lambda \in \mathbb{R}$. Another way to write the dual problem, which is also invariant by $\phi \mapsto \phi + \lambda$ and $\psi \mapsto \psi + \lambda$, reads as

$$\sup_{\substack{\phi \in \mathscr{C}(X^-) \\ \psi \in \mathscr{C}(X^+)}} \left\{ \int_{X^-} \phi\ \mathrm{d}\mu^- + \int_{X^+} \psi\ \mathrm{d}\mu^+ - \varepsilon \log \int_{X^- \times X^+} e^{\frac{\phi(x)+\psi(y)-c(x,y)}{\varepsilon}}\ \mathrm{d}\mu^-(x)\, \mathrm{d}\mu^+(y) \right\}, \tag{D$_\varepsilon$}$$

see [Léo01] or [Nut22] for a more recent presentation. From (EOT$_\varepsilon$) and (D$_\varepsilon$) we recover, as $\varepsilon \to 0$, the (unregularized) optimal transport problem associated with c and its dual, that we recall for the reader's sake:

$$v_0 := \inf_{\gamma \in \Pi(\mu^-,\, \mu^+)} \int_{X^- \times X^+} c(x,y)\ \mathrm{d}\gamma(x,y) \tag{OT}$$

and

$$\sup_{\substack{\phi \in \mathscr{C}(X^-) \\ \psi \in \mathscr{C}(X^+)}} \left\{ \int_{X^-} \phi\ \mathrm{d}\mu^- + \int_{X^+} \psi\ \mathrm{d}\mu^+ : \phi \oplus \psi \leq c \right\}, \tag{D}$$

respectively, where $\phi \oplus \psi(x,y) := \phi(x) + \psi(y)$. The set of optimal couplings between μ^- and μ^+ will be denoted by $\mathsf{Opt}(\mu^-, \mu^+)$. The value of (OT), denoted by v_0, is called *transport cost*. As for (D), it is well known that optimizers, called *Kantorovich potentials*, exist whenever $X^{\pm}$ are compact and they can be chosen c-conjugate (see [San15, Chapter 1]):

$$\phi = \inf_{y \in X^+} c(\cdot, y) - \psi(y), \quad \psi = \inf_{x \in X^-} c(x, \cdot) - \phi(x).$$

The link between EOT and OT is very strong, as already discussed in the introduction. For later use, recall that a consequence of the Γ-convergence of (EOT_ε) towards (OT) is that

$$\lim_{\varepsilon \downarrow 0} v_\varepsilon = v_0. \tag{2.1}$$

A proof tailored to our setting can be found in [Car+17, Section 2.3] (although formulated for the quadratic cost, it actually works for any continuous cost as long as $\mu^\pm$ have compact supports).

The fact that (EOT_ε) admits a unique solution γ_ε, called *optimal entropic plan*, is a consequence of the direct method in the calculus of variations and strict convexity of the entropy. The structure of γ_ε is then very rigid: indeed, there exist two real-valued Borel functions $\phi_\varepsilon, \psi_\varepsilon$ such that

$$\gamma_\varepsilon = \exp\left(\frac{\phi_\varepsilon \oplus \psi_\varepsilon - c}{\varepsilon}\right)\mu^- \otimes \mu^+, \tag{2.2}$$

which implies in particular that

$$v_\varepsilon = \int_{X^-} \phi_\varepsilon \, \mathrm{d}\mu^- + \int_{X^+} \psi_\varepsilon \, \mathrm{d}\mu^+ \tag{2.3}$$

and these functions, which have continuous representatives, are a.s. uniquely determined up to additive constants, in the sense that if (2.2) holds for (ϕ, ψ) and (ϕ', ψ'), then $\phi = \phi' + \lambda$ μ^--a.e. and $\psi = \psi' - \lambda$ μ^+-a.e. for some $\lambda \in \mathbb{R}$. Moreover, γ_ε is the unique coupling $\gamma \in \Pi(\mu^-, \mu^+)$ whose density w.r.t. $\mu^- \otimes \mu^+$ can be written as in (2.2). A self-contained proof of all these facts in the case of compactly supported measures $\mu^\pm$ can be found in [GT21, Proposition 2.1]. The reader is refered to the analysis of [MG20], to the notes of Nutz [Nut22] for a more general framework, and to [BL92; BLN94; Csi75; FG97; RT98] for earlier references.

The functions $\phi_\varepsilon, \psi_\varepsilon$ in (2.2) are called *Schrödinger potentials*, the terminology being motivated by the fact that they solve the dual problem (D_ε). Furthermore, $\phi_\varepsilon, \psi_\varepsilon$ are the (unique) solutions to the so-called *Schrödinger system*

$$\begin{cases} \phi(x) = -\varepsilon \log \displaystyle\int_{X^+} e^{\frac{\psi(y)-c(x,y)}{\varepsilon}} \, \mathrm{d}\mu^+(y) & \text{for } \mu^-\text{-a.e. } x, \\ \psi(y) = -\varepsilon \log \displaystyle\int_{X^-} e^{\frac{\phi(x)-c(x,y)}{\varepsilon}} \, \mathrm{d}\mu^-(x) & \text{for } \mu^+\text{-a.e. } y, \end{cases} \tag{2.4}$$

named after E. Schrödinger who introduced it in the seminal papers [Sch31; Sch32] (see also [CMS21] for a recent English translation of the former). Note that (2.4) is a softmin version of the classical c-conjugacy relation for Kantorovich potentials. From (2.4) and the continuity of c, we easily deduce that ϕ, ψ share a common modulus of continuity with c (see [MG20] for details). From an OT viewpoint, (2.4) simply means that $e^{(\phi \oplus \psi - c)/\varepsilon}\mu^- \otimes \mu^+ \in \Pi(\mu^-, \mu^+)$.

We are now ready to state and prove the first main result. Recall that v_ε denotes the value of (EOT_ε) and v_0 the one of (OT); the dependence on $\mu^+, \mu^- \in \mathscr{P}(\mathbb{R}^d)$ is omitted, as they will always be fixed.

Theorem 2.1. *Let $\mu^\pm \in \mathscr{P}(\mathbb{R}^d)$ have compact support $X^\pm \subseteq \mathbb{R}^d$ and $c \in \mathscr{C}(X^- \times X^+)$. Then the function $\varepsilon \mapsto v_\varepsilon$ is continuous, non-decreasing and concave on $[0, +\infty)$; it is analytic on $(0, +\infty)$.*

Proof. The convergence of v_ε to v_0 as $\varepsilon \downarrow 0$ is precisely (2.1). The fact that v is non-decreasing follows from the non-negativity of the entropy; concavity is also obvious since v_ε is an infimum of affine functions of ε.

Analyticity of v on $(0, +\infty)$ can be deduced from the analytic version of the implicit function theorem, adapting the arguments of [CL20] as follows. Let $\varepsilon > 0$ and, recalling that Schrödinger potentials are unique up to the trivial transformation $(\phi, \psi) \mapsto (\phi - \lambda, \psi + \lambda)$ with $\lambda \in \mathbb{R}$, impose the following normalization condition

$$\int_{X^-} \phi \, \mathrm{d}\mu^- = 0. \tag{2.5}$$

Denote then by $(\phi_\varepsilon, \psi_\varepsilon)$ the unique functions satisfying (2.4) and (2.5) and note that by (2.3) and (2.5) the entropic cost writes as

$$v_\varepsilon = \int_{X^+} \psi_\varepsilon \, \mathrm{d}\mu^+. \tag{2.6}$$

As already noticed, ϕ_ε and ψ_ε are continuous in X^- and X^+ respectively, whence the validity of both equations in (2.4) in a pointwise sense. This means that, if we define

$$\begin{aligned} F(\varepsilon, \phi, \psi)(x) &:= e^{\frac{\phi(x)}{\varepsilon}} \int_{X^+} e^{\frac{-c(x,y)+\psi(y)}{\varepsilon}} \, \mathrm{d}\mu^+(y), \qquad x \in X^- \\ G(\varepsilon, \phi, \psi)(y) &:= e^{\frac{\psi(y)}{\varepsilon}} \int_{X^-} e^{\frac{-c(x,y)+\phi(x)}{\varepsilon}} \, \mathrm{d}\mu^-(x), \qquad y \in X^+ \end{aligned}$$

the Schrödinger system (2.4) is equivalent to

$$F(\varepsilon, \phi_\varepsilon, \psi_\varepsilon)(x) = 1 \quad (\forall x \in X^-), \qquad \text{and} \qquad G(\varepsilon, \phi_\varepsilon, \psi_\varepsilon)(y) = 1 \quad (\forall y \in X^+).$$

Finally, define the spaces

$$A := \{(\phi, \psi) \in \mathscr{C}(X^-) \times \mathscr{C}(X^+) \ : \ (2.5) \text{ holds}\}$$

and

$$B := \left\{ (f, g) \in \mathscr{C}(X^-) \times \mathscr{C}(X^+) \ : \int_{X^-} f \, \mathrm{d}\mu^- = \int_{X^+} g \, \mathrm{d}\mu^+ \right\},$$

which are Banach spaces when equipped with the uniform norm, and the map $S : (0, +\infty) \times \mathscr{C}(X^-) \times \mathscr{C}(X^+) \to \mathscr{C}(X^-) \times \mathscr{C}(X^+)$

$$S(\varepsilon, \phi, \psi) := (F(\varepsilon, \phi, \psi), G(\varepsilon, \phi, \psi)).$$

Let us observe that S is analytic, takes values in B and the pair of normalized Schrödinger potentials $(\phi_\varepsilon, \psi_\varepsilon)$ is implicitly defined by

$$(\phi_\varepsilon, \psi_\varepsilon) \in A \quad \text{and} \quad S(\varepsilon, \phi_\varepsilon, \psi_\varepsilon) = (1, 1).$$

We claim that for every $\varepsilon > 0$ and $(\phi, \psi) \in A$, the derivative of S w.r.t. (ϕ, ψ) at $(\varepsilon, \psi, \phi)$ is an isomorphism between A and B, analyticity of $\varepsilon \mapsto (\phi_\varepsilon, \psi_\varepsilon)$ will then follow from the implicit function theorem in the analytic case. To ease notations assume $\varepsilon = 1$ and let $(\phi, \psi) \in A$ be fixed. Defining

$$S = S(1, \cdot, \cdot) = (F(1, \cdot, \cdot), G(1, \cdot, \cdot)), \qquad \tilde{S} := (\log(F), \log(G)),$$

and

$$L := S'(\phi, \psi) = (L_1, L_2), \qquad \tilde{L} := \tilde{S}'(\phi, \psi) = (\tilde{L}_1, \tilde{L}_2),$$

for $(h, k) \in \mathscr{C}(X^-) \times \mathscr{C}(X^+)$ we then have

$$\tilde{L}_1(h, k)(x) = h(x) + \frac{\displaystyle\int_{X^+} e^{-c(x,y)+\psi(y)} k(y) \, \mathrm{d}\mu^+(y)}{\displaystyle\int_{X^+} e^{-c(x,y)+\psi(y)} \, \mathrm{d}\mu^+(y)}, \quad L_1(h, k) = F(\phi, \psi)\tilde{L}_1(h, k),$$

and

$$\tilde{L}_2(h, k)(y) = k(y) + \frac{\displaystyle\int_{X^-} e^{-c(x,y)+\phi(x)} h(x) \, \mathrm{d}\mu^-(x)}{\displaystyle\int_{X^-} e^{-c(x,y)+\phi(x)} \, \mathrm{d}\mu^-(x)}, \quad L_2(h, k) = G(\phi, \psi)\tilde{L}_2(h, k).$$

We can write in a more synthetic way $\tilde{L}$ as

$$\tilde{L}_1(h, k)(x) = h(x) + \int_{X^+} k(y) \, \mathrm{d}Q^x(y), \quad \tilde{L}_2(h, k)(y) = k(y) + \int_{X^-} h(x) \, \mathrm{d}Q_y(x), \tag{2.7}$$

where Q^x and Q_y are the disintegration measures of the probability measure

$$\mathrm{d}Q(x, y) := \frac{e^{\phi(x)+\psi(y)-c(x,y)} \, \mathrm{d}\mu^-(x) \, \mathrm{d}\mu^+(y)}{\displaystyle\int_{X^- \times X^+} e^{\phi(x')+\psi(y')-c(x',y')} \, \mathrm{d}\mu^-(x') \, \mathrm{d}\mu^+(y')}$$

with respect to its first marginal α and second marginal β, that is:

$$Q = \alpha \otimes Q^x = Q_y \otimes \beta \tag{2.8}$$

(in other words given (X, Y) with law Q, Q^x is the conditional law of Y given $X = x$ and Q_y is the conditional law of X given $Y = y$). Assume now that $(h, k) \in \mathscr{C}(X^-) \times \mathscr{C}(X^+)$ belongs to $\ker(L) = \ker(\tilde{L})$, i.e.

$$h(x) + \int_{X^+} k(y) \, \mathrm{d}Q^x(y) = 0, \qquad \forall x \in X^- = S_\alpha \tag{2.9}$$

and

$$k(y) + \int_{X^-} h(x) \, \mathrm{d}Q_y(x) = 0, \qquad \forall y \in X^+ = S_\beta \tag{2.10}$$

where $X^- = S_\alpha$ and $X^+ = S_\beta$ follow from the equivalence of Q with $\mu^- \otimes \mu^+$ (in the sense that they have the same negligible sets). Multiplying (2.9) by $h(x)$ and integrating w.r.t. α, multiplying (2.10) by $k(y)$ and integrating w.r.t. β, and summing the two, by (2.8) we get

$$\int_{X^- \times X^+} (h(x) + k(y))^2 \, \mathrm{d}Q(x, y) = 0$$

from which we deduce that $h(x) + k(y) = 0$ for $\mu^- \otimes \mu^+$-a.e. (x, y), that is h and k are two constants that sum to 0. We thus have $\ker(\tilde{L}) = \ker(L) = \{(\lambda, -\lambda) : \lambda \in \mathbb{R}\}$ and in particular $\ker(L) \cap A = \{0\}$, so that L is injective on A. Since c is uniformly continuous on $X^- \times X^+$, all functions in the image of $\tilde{L}$ are equi-continuous and equi-bounded, and it follows from Ascoli's Theorem that $\tilde{L}$ is a compact perturbation of the identity map on $\mathscr{C}(X^-) \times \mathscr{C}(X^+)$. This enables us to invoke the Fredholm alternative theorem to deduce from the fact that $\ker(\tilde{L})$ is a line that $\tilde{L}(\mathscr{C}(X^-) \times \mathscr{C}(X^+)) = \tilde{L}(A)$ is a hyperplane of $\mathscr{C}(X^-) \times \mathscr{C}(X^+)$. Since $\tilde{L}(A)$ is isomorphic to $L(A)$, and $L(A)$ is included in the hyperplane B (because the original non-linear map S takes values in B), we have $L(A) = B$.

Hence L is an isomorphism between A and B and, as anticipated, the implicit function theorem for analytic maps (see chapter 4 of [BT16]) therefore implies that $(\phi_\varepsilon, \psi_\varepsilon)$ depends analytically on $\varepsilon > 0$. In particular $(0, +\infty) \ni \varepsilon \mapsto \psi_\varepsilon \in \mathscr{C}(X^-)$ is analytic too, whence the analyticity of v on $(0, +\infty)$ thanks to (2.6). □

Remark 2.2. The entropic cost $\varepsilon \mapsto v_\varepsilon$ being analytic on $(0, +\infty)$, it can be extended in an analytic fashion to some complex values of ε. Note that the proof above shows that the entropic cost is also an analytic function of the transport cost c.

Remark 2.3. Even if Theorem 2.1 is stated in the Euclidean setting, it is worth stressing that it is in fact true in a much more general framework. Indeed, it is not difficult to see that the previous proof works verbatim over compact metric spaces and an even higher level of generality is possible. More precisely, not even a topological structure is needed. It is indeed sufficient that $c \in L^\infty(\mu^- \otimes \mu^+)$ with μ^-, μ^+ probability measures on arbitrary measurable spaces, as it is the case in [CL20]. However, with no topological structure the proof of the analyticity of $\varepsilon \mapsto v_\varepsilon$ would be more involved.

Knowing that all order derivatives of $\varepsilon \mapsto v_\varepsilon$ exist on $(0, +\infty)$, it is natural to look for an explicit expression for them. The case of the first derivative is particularly easy to handle, as shown below.

Corollary 2.4. *Under the same assumptions of Theorem 2.1, it holds*

$$v'_\varepsilon = \mathrm{Ent}(\gamma_\varepsilon \,|\, \mu^- \otimes \mu^+), \qquad \forall \varepsilon > 0. \tag{2.11}$$

Besides, if there exists $\gamma_0 \in \mathsf{Opt}(\mu^- \otimes \mu^+)$ *with* $\mathrm{Ent}(\gamma_0 \,|\, \mu^- \otimes \mu^+) < +\infty$, *then* $\varepsilon \mapsto v_\varepsilon$ *belongs to* $\mathscr{C}^1([0, +\infty))$ *with right derivative at* $\varepsilon = 0$ *given by*

$$v'_\varepsilon|_{\varepsilon=0} = \inf_{\gamma \in \mathsf{Opt}(\mu^-, \mu^+)} \mathrm{Ent}(\gamma \,|\, \mu^- \otimes \mu^+), \tag{2.12}$$

in other words:

$$v_\varepsilon - v_0 = \varepsilon \inf_{\gamma \in \mathsf{Opt}(\mu^-,\mu^+)} \mathrm{Ent}(\gamma \,|\, \mu^- \otimes \mu^+) + o(\varepsilon). \tag{2.13}$$

Moreover, in that case, γ_ε converges narrowly to the unique coupling γ^* of minimal entropy in $\mathsf{Opt}(\mu^-,\mu^+)$, namely

$$\gamma^* = \operatorname*{arg\,min}_{\gamma \in \mathrm{Opt}(\mu^-,\mu^+)} \mathrm{Ent}(\gamma \,|\, \mu^- \otimes \mu^+).$$

Proof. Knowing that $\varepsilon \mapsto v_\varepsilon$ is analytic ($\mathscr{C}^1$ would be enough), (2.11) is a consequence of the envelope theorem and of the uniqueness of the minimizer of (EOT_ε).

As concerns the second part of the statement, define

$$C_\varepsilon(\gamma) := \int_{X^- \times X^+} c \, \mathrm{d}\gamma + \varepsilon \mathrm{Ent}(\gamma \,|\, \mu^- \otimes \mu^+)$$

and observe that $v_\varepsilon - v_0 = C_\varepsilon(\gamma_\varepsilon) - C_0(\gamma) \le C_\varepsilon(\gamma) - C_0(\gamma) = \varepsilon \mathrm{Ent}(\gamma \,|\, \mu^- \otimes \mu^+)$ for any $\gamma \in \mathsf{Opt}(\mu^-,\mu^+)$, hence

$$\limsup_{\varepsilon \downarrow 0} \frac{v_\varepsilon - v_0}{\varepsilon} \le \inf_{\gamma \in \mathsf{Opt}(\mu^-,\mu^+)} \mathrm{Ent}(\gamma \,|\, \mu^- \otimes \mu^+), \tag{2.14}$$

the right-hand side being finite thanks to the existence of γ_0 as in the statement. For the liminf inequality, recall that (EOT_ε) Γ-converges to (OT) as $\varepsilon \to 0$, so that up to subsequences γ_ε converges narrowly to $\gamma^* \in \mathsf{Opt}(\mu^-,\mu^+)$. Combining this information with $v_\varepsilon - v_0 \ge C_\varepsilon(\gamma_\varepsilon) - C_0(\gamma_\varepsilon) = \varepsilon \mathrm{Ent}(\gamma_\varepsilon \,|\, \mu^- \otimes \mu^+)$ and the lower semicontinuity of the entropy (see [San15, §7.1.2]) yields

$$\liminf_{\varepsilon \downarrow 0} \frac{v_\varepsilon - v_0}{\varepsilon} \ge \mathrm{Ent}(\gamma^* \,|\, \mu^- \otimes \mu^+) \ge \inf_{\gamma \in \mathsf{Opt}(\mu^-,\mu^+)} \mathrm{Ent}(\gamma \,|\, \mu^- \otimes \mu^+),$$

and this inequality[1] together with (2.14) implies the right differentiability of v_ε at $\varepsilon = 0$ as well as (2.12).

Note that the argument above also proves the existence of a $\gamma^* \in \mathsf{Opt}(\mu^-,\mu^+)$ of minimal entropy towards which γ_ε converges narrowly (up to subsequences) as well as the lower semicontinuity of v'_ε at $\varepsilon = 0$. The uniqueness of such γ^* comes from the convexity of $\mathsf{Opt}(\mu^-,\mu^+)$ and the strict convexity of the entropy, hence the convergence of the whole sequence. The upper semicontinuity of v'_ε at $\varepsilon = 0$ is instead a consequence of concavity, so that in conclusion v'_ε is continuous up to $\varepsilon = 0$, namely $v_\varepsilon \in \mathscr{C}^1([0,+\infty))$. □

Let us stress that while (2.11) is an immediate consequence of Theorem 2.1, the proof of (2.12) (inspired by [CT21; MTV20]) requires the existence of an optimal coupling with finite entropy w.r.t. $\mu^- \otimes \mu^+$ and the reader may wonder whether this condition is reasonable or not. If μ^- and μ^+ are finitely-supported measures, the condition is

[1] By the way, this also shows that the differentiability of v_ε at $\varepsilon = 0$ is in fact equivalent to the existence of an optimal plan with finite entropy.

trivially satisfied; hence identities (2.12) and (2.13) may be of some interest in this case (in this direction, recall the pioneering work [CM94]).

However, as soon as μ^- and μ^+ have densities, the aforementioned condition is not satisfied in general. In particular, it cannot hold whenever optimal plans are concentrated on d-rectifiable sets. To obtain a non-trivial asymptotic expansion of the entropic cost around $\varepsilon = 0$ in more general situations, a more accurate study of $v_\varepsilon - v_0$ is thus required, and this will be carried out in Sections 3 and 4.

3 Upper bound

Our goal in this section is to establish upper bounds on $v_\varepsilon - v_0$ up to order $O(\varepsilon)$ as $\varepsilon \to 0$. We start with a general upper bound holding for $\mathscr{C}^{0,1}$ (i.e. Lipschitz) costs, then we prove a finer upper bound for $\mathscr{C}^{1,1}$ costs. The leading order terms, which are proportional to $\varepsilon \log(1/\varepsilon)$, will be shown to be sharp in this section and in Section 4.

3.1 Upper bound for $\mathscr{C}^{0,1}$ costs

The upper bound we are going to establish will depend on the dimension of μ^- and μ^+. The natural notion of dimension which arises in our context is the *entropy dimension*, also called *information dimension* or *Rényi dimension*, since it was originally introduced by Rényi in [Rén59]. The definition that we use corresponds to the one given in [You82]: if μ is a probability measure over $\mathbb{R}^d$, we set for every $\delta > 0$,

$$H_\delta(\mu) = \inf \left\{ \sum_{n\in\mathbb{N}} \mu(A_n) \log(1/\mu(A_n)) : \forall n, \operatorname{diam}(A_n) \le \delta, \text{ and } \mathbb{R}^d = \bigcup_{n\in\mathbb{N}} A_n \right\},$$

where the infimum is taken over countable coverings $(A_n)_{n\in\mathbb{N}}$ of $\mathbb{R}^d$ by Borel subsets, and we define the *lower and upper entropy dimension of* μ respectively by:

$$\underline{\dim}_E(\mu) := \liminf_{\delta\to 0^+} \frac{H_\delta(\mu)}{\log(1/\delta)}, \quad \overline{\dim}_E(\mu) := \limsup_{\delta\to 0^+} \frac{H_\delta(\mu)}{\log(1/\delta)}.$$

When μ has compact support, notice that its upper entropy dimension is always smaller than the upper box dimension of the support of μ, that is:

$$\overline{\dim}_E(\mu) \le \overline{\dim}_B(S_\mu) := \limsup_{\delta\to 0^+} \frac{\log(N_\delta(S_\mu))}{\log(1/\delta)},$$

where for every $A \subseteq \mathbb{R}^d$, $N_\delta(A)$ denotes the box-counting number of A at scale δ, i.e. the minimal number of sets of diameter $\delta > 0$ which cover A. Indeed if $(A_1, \dots, A_{N_\delta(S_\mu)})$ is such a covering, by concavity of $t \mapsto t\log(1/t)$, we have

$$\sum_{1\le n\le N_\delta(S_\mu)} \mu(A_n) \log(1/\mu(A_n)) \le \log N_\delta(S_\mu). \tag{3.1}$$

The reader interested in the different notions of dimension of sets and measures may consult [Fal97].

Proposition 3.1. *If $\mu^\pm \in \mathscr{P}(\mathbb{R}^d)$ and $c \in \mathscr{C}^{0,1}(S_{\mu^-} \times S_{\mu^+})$, then for every $\varepsilon > 0$,*

$$v_\varepsilon \leq v_0 + \varepsilon(H_\varepsilon(\mu^-) \wedge H_\varepsilon(\mu^+)) + [c]_{\mathscr{C}^{0,1}(S_{\mu^-} \times S_{\mu^+})}\varepsilon. \tag{3.2}$$

In particular, setting $d^\pm = \overline{\dim}_E(\mu^\pm)$,

$$v_\varepsilon \leq v_0 + (d^+ \wedge d^-)\varepsilon \log(1/\varepsilon) + o(\varepsilon \log(1/\varepsilon)) \quad \text{as } \varepsilon \to 0, \tag{3.3}$$

and if $\mu^\pm$ are concentrated on compact submanifolds of dimension $d^\pm$,

$$v_\varepsilon \leq v_0 + (d^+ \wedge d^-)\varepsilon \log(1/\varepsilon) + O(\varepsilon). \tag{3.4}$$

The proof is based on a generalization of the block approximation defined in [Car+17, Definition 2.9].

Proof. Let $\gamma_0 \in \mathsf{Opt}(\mu^-, \mu^+)$. First of all, let us extend the cost to a function $c \in \mathscr{C}^{0,1}(\mathbb{R}^d \times \mathbb{R}^d)$ with same Lipschitz constant $L := [c]_{\mathscr{C}^{0,1}(\mathbb{R}^d \times \mathbb{R}^d)} = [c]_{\mathscr{C}^{0,1}(S_{\mu^-} \times S_{\mu^+})}$. We consider a covering $\mathbb{R}^d = \bigcup_{n \in \mathbb{N}} A_n$ of Borel sets such that $\operatorname{diam}(A_n) \leq \delta$ for every $n \in \mathbb{N}$. We set for every $n \in \mathbb{N}$,

$$\mu_n^\pm := \begin{cases} \dfrac{\mu^\pm \llcorner A_n}{\mu^\pm(A_n)} & \text{if } \mu^\pm(A_n) > 0, \\ 0 & \text{otherwise,} \end{cases}$$

then for every pair $i, j \in \mathbb{N}$,

$$(\gamma_0)_{i,j} := \gamma_0(A_i \times A_j)\mu_i^- \otimes \mu_j^+,$$

and finally,

$$\gamma^\delta := \sum_{i,j \in \mathbb{N}} (\gamma_0)_{i,j}.$$

By definition, $\gamma^\delta \ll \mu^- \otimes \mu^+$ and we may check that its marginals are $\mu^\pm$, for instance:

$$(\pi_1)_\# \gamma^\delta = \sum_{i,j \in \mathbb{N}} (\pi_1)_\#(\gamma_0(A_i \times A_j)\mu_i^- \otimes \mu_j^+) = \sum_{i,j \in \mathbb{N}} \gamma_0(A_i \times A_j)\mu_i^- = \sum_{i \in \mathbb{N}} \mu^-(A_i)\mu_i^- = \mu^-.$$

Besides, for $\mu^- \otimes \mu^+$-almost every $(x, y) \in A_i \times A_j$,

$$\frac{\mathrm{d}\gamma^\delta}{\mathrm{d}\mu^- \otimes \mu^+}(x, y) = \frac{\gamma_0(A_i \times A_j)}{\mu^-(A_i)\mu^+(A_j)} \quad \text{if } \mu^-(A_i)\mu^+(A_j) > 0,$$

and it is 0 otherwise. Let us compute its entropy, summing only over indices i, j such

that $\mu^-(A_i) > 0$ and $\mu^+(A_j) > 0$.

$$\begin{aligned}
\mathrm{Ent}(\gamma^\delta \,|\, \mu^- \otimes \mu^+) &= \sum_{i,j\in\mathbb{N}} \int_{A_i\times A_j} \log\left(\frac{\gamma_0(A_i\times A_j)}{\mu^-(A_i)\mu^+(A_j)}\right) \mathrm{d}\gamma^\delta \\
&= \sum_{i,j\in\mathbb{N}} \gamma_0(A_i\times A_j) \log\left(\frac{\gamma_0(A_i\times A_j)}{\mu^-(A_i)\mu^+(A_j)}\right) \\
&= \sum_{i,j\in\mathbb{N}} \gamma_0(A_i\times A_j) \log\left(\frac{\gamma_0(A_i\times A_j)}{\mu^-(A_i)}\right) + \sum_{j\in\mathbb{N}} \mu^+(A_j)\log(1/\mu^+(A_j)) \\
&\le \sum_{j\in\mathbb{N}} \mu^+(A_j)\log(1/\mu^+(A_j)),
\end{aligned}$$

the last inequality coming from the inequality $\gamma_0(A_i\times A_j) \le \mu^-(A_i)$. Taking arbitrary coverings $(A_n)_{n\in\mathbb{N}}$ of diameter smaller than δ, we get by definition of H_δ,

$$\mathrm{Ent}(\gamma^\delta \,|\, \mu^- \otimes \mu^+) \le H_\delta(\mu^+).$$

Notice that $W_\infty(\gamma^\delta, \gamma_0) \le \delta$, where W_p denotes the p-Wasserstein distance[2] between compactly supported probability measures for $p \in [1,\infty]$. As a consequence, taking γ^δ as competitor in (EOT$_\varepsilon$), since $W_1 \le W_\infty$ we obtain

$$\begin{aligned}
v_\varepsilon \le \int_{\mathbb{R}^d\times\mathbb{R}^d} c\,\mathrm{d}\gamma^\delta + \varepsilon H_\delta(\mu^+) &= v_0 + \int_{\mathbb{R}^d\times\mathbb{R}^d} c\,\mathrm{d}(\gamma^\delta - \gamma_0) + \varepsilon H_\delta(\mu^+) \\
&\le v_0 + LW_1(\gamma^\delta,\gamma_0) + \varepsilon H_\delta(\mu^+) \\
&\le v_0 + L\delta + \varepsilon H_\delta(\mu^+).
\end{aligned} \tag{3.5}$$

Taking $\delta = \varepsilon$ and interchanging the roles of μ^- and μ^+ yields (3.2), and (3.3) follows from the definition of the upper entropy dimension. Finally, if $\mu^\pm$ are concentrated on compact $d^\pm$-dimensional manifolds $M^\pm$, knowing that their box counting number $N_\varepsilon(M^\pm)$ is bounded from above by $C/\varepsilon^{d^\pm}$ for some constant $C > 0$ and every $\varepsilon \in (0,1)$, applying (3.1) to $\mu^\pm$ yields $H_\varepsilon(\mu^\pm) \le d^\pm \log(1/\varepsilon) + \log C$, thus

$$v_\varepsilon \le v_0 + (d^- \wedge d^+)\varepsilon\log(1/\varepsilon) + (L + \log C)\varepsilon$$

and (3.4) is proved. □

Remark 3.2. If the cost c is only assumed to be uniformly continuous on $S_{\mu^-}\times S_{\mu^+}$, with modulus of continuity $\omega_c : \delta \mapsto \sup_{\|\boldsymbol{x}-\boldsymbol{x}'\|\le\delta}|c(\boldsymbol{x}) - c(\boldsymbol{x}')|$, a straightforward adaptation of the above proof shows that

$$v_\varepsilon \le v_0 + (d^+ \wedge d^- + o(1))\varepsilon\log(1/\omega_c^{-1}(\varepsilon)),$$

and if $\mu^\pm$ are concentrated on $d^\pm$-dimensional submanifolds,

$$v_\varepsilon \le v_0 + (d^+ \wedge d^-)\varepsilon\log(1/\omega_c^{-1}(\varepsilon)) + O(\varepsilon),$$

where ω_c^{-1} is the generalized inverse defined by $\omega_c^{-1}(t) := \inf\{s \ge 0 : \omega_c(s) \ge t\}$.

[2]For measures on $\mathbb{R}^d\times\mathbb{R}^d$, we consider the Wasserstein distance with respect to the norm that we introduced, namely $\|(x,y)\| = \max\{|x|,|y|\}$ for every $(x,y)\in\mathbb{R}^d\times\mathbb{R}^d$.

The following easy example shows that the leading order term is in general sharp in the class of $\mathscr{C}^{0,1}$ cost functions.

Example 3.3. Consider $\mu^- = \mu^+ = \mathscr{L} \llcorner [0,1]$ and $c(x,y) := |x-y|$ for $x, y \in [0,1]$. Then

$$v_\varepsilon \geq v_0 + \varepsilon \log(1/\varepsilon).$$

Indeed, the unique optimizer to (OT) is $\gamma_0 = (\mathrm{Id}, \mathrm{Id})_\# \mu^-$ with cost $v_0 = 0$, and the pair $(\phi_0, \psi_0) = (0,0)$ is a solution to the dual problem (D). We use (ϕ_0, ψ_0) as a competitor in (D_ε) to get

$$v_\varepsilon \geq -\varepsilon \log \left(\int_{[0,1]^2} e^{-\frac{|y-x|}{\varepsilon}} \, \mathrm{d}x \, \mathrm{d}y \right).$$

We have

$$\int_{[0,1]^2} e^{-\frac{|y-x|}{\varepsilon}} \, \mathrm{d}x \, \mathrm{d}y = 2 \int_0^1 \int_0^y e^{-\frac{t}{\varepsilon}} \, \mathrm{d}t \, \mathrm{d}y = 2 \int_0^1 \varepsilon (1 - e^{-\frac{y}{\varepsilon}}) \, \mathrm{d}y = 2\varepsilon(1 - \varepsilon(1 - e^{-\frac{1}{\varepsilon}})) \leq 2\varepsilon,$$

thus for every $\varepsilon > 0$,

$$v_\varepsilon \geq -\varepsilon \log 2\varepsilon = v_0 + \varepsilon \log(1/\varepsilon) + O(\varepsilon),$$

and together with (3.4), since $d^\pm = 1$, we get $v_\varepsilon - v_0 = \varepsilon \log(1/\varepsilon) + O(\varepsilon)$.

3.2 Upper bound for $\mathscr{C}^{1,1}$ costs and L^∞ marginals

Building upon the same block approximation used to establish Proposition 3.1, we provide a finer upper bound when the cost is of class $\mathscr{C}^{1,1}$ and $\mu^\pm$ have L^∞ densities.

Proposition 3.4. *Let $c \in \mathscr{C}^{1,1}_{\mathrm{loc}}(\Omega^- \times \Omega^+)$ where $\Omega^\pm \subseteq \mathbb{R}^d$ are open convex sets and $\mu^\pm \in L^\infty(\Omega^\pm)$ be two probability measures compactly supported in $\Omega^\pm$. Then there exists a constant $M \geq 0$ such that for every $\varepsilon \in (0,1)$,*

$$v_\varepsilon \leq v_0 + \frac{d}{2} \varepsilon \log(1/\varepsilon) + M\varepsilon. \tag{3.6}$$

Remark 3.5. Notice that Proposition 3.1 would only show that $v_\varepsilon \leq v_0 + d\varepsilon \log(1/\varepsilon) + O(\varepsilon)$. The term $d\varepsilon \log(1/\varepsilon)$ term is generally sharp as shown in Example 3.3, when c is only Lipschitz, but Proposition 3.4 shows that d may actually be replaced by $d/2$ when c is $\mathscr{C}^{1,1}$. This is in turn sharp in the class of $\mathscr{C}^{1,1}$ costs, as already exhibited in the quadratic case in [Ada+11] or in the setting of [Pal19], but it is also a consequence of our lower bound stated in Proposition 4.4 in the case of infinitesimally twisted cost functions.

To prove Proposition 3.4, we will need a quadratic bound on the average error between a λ-convex function f and its first-order Taylor expansion. We recall that a function $f : \Omega \to \mathbb{R}$ defined on a convex set $\Omega \subseteq \mathbb{R}^d$ is λ-convex, for $\lambda \in \mathbb{R}$, if $x \mapsto f(x) - \lambda \dfrac{|x|^2}{2}$ is convex, and f is λ-concave if $-f$ is $(-\lambda)$-convex. If Ω is open and f is λ-convex,

then it is differentiable almost everywhere on Ω by Rademacher's Theorem, and its distributional Hessian $D^2 f$ is a $S_d(\mathbb{R})$-valued Radon measure over Ω, which may be written (see [AFP00, §1.3]) as $D^2 f = \sigma_f |D^2 f|$ where $|D^2 f|$ is a positive Radon measure, and $\sigma_f : \Omega \to S_d(\mathbb{R})$ is a Borel unit matrix field for the Frobenius norm.

Lemma 3.6. *Let $f : \Omega \to \mathbb{R}$ be a λ-convex function on a convex open set for some $\lambda \leq 0$. There exists a constant $C \geq 0$ depending only on d such that:*

$$\int_\Omega \sup_{y \in B_r(x) \cap \Omega} |f(y) - (f(x) + \nabla f(x) \cdot (y - x))| \, \mathrm{d}x \leq C r^2 \mathscr{H}^{d-1}(\partial\Omega)([f]_{\mathscr{C}^{0,1}(\Omega)} + |\lambda| \operatorname{diam}(\Omega)). \tag{3.7}$$

The proof is based on that of Alexandrov's Theorem, asserting the existence of a second-order Taylor expansion at almost every point, which is given in [EG15, §6.4].

Proof. **Step 1.** Let g be a function which is convex on $B_r(x)$ and assume that x is a differentiability point of g and a Lebesgue point of both g and ∇g. Let us show that

$$\sup_{y \in B_{r/2}(x)} |g(y) - (g(x) + \nabla g(x) \cdot (y - x))| \leq C r^2 \frac{|D^2 g|(B_r(x))}{\mathscr{L}^d(B_r(x)))} \tag{3.8}$$

for a constant $C = C(d)$. For $\varepsilon \in (0, r)$, we set $r_\varepsilon = r - \varepsilon$ and we define $g_\varepsilon \coloneqq g \star \eta_\varepsilon$ on $B_{r_\varepsilon}(x)$, where η_ε is a standard mollifier supported on B_ε. Since $g_\varepsilon \in \mathscr{C}^2(B_{r_\varepsilon}(x))$, by the classical Taylor integral formula, for every $h \in B_{r_\varepsilon}$,

$$g_\varepsilon(x + h) = g_\varepsilon(x) + \nabla g_\varepsilon(x) \cdot h + \int_0^1 (1 - t) D^2 g_\varepsilon(x + th) \cdot h \otimes h \, \mathrm{d}t,$$

and integrating w.r.t. h over B_{r_ε}, we get the bound:

$$\begin{aligned}
\int_{B_{r_\varepsilon}} |g_\varepsilon(x + h) - (g_\varepsilon(x) + \nabla g_\varepsilon(x) \cdot h)| \, \mathrm{d}h &\leq r_\varepsilon^2 \int_0^1 (1 - t) |D^2 g_\varepsilon|(B_{r_\varepsilon t}(x)) \, \mathrm{d}t \\
&\leq r_\varepsilon^2 \int_0^1 (1 - t) |D^2 g|(B_{r_\varepsilon t + \varepsilon}(x)) \, \mathrm{d}t \\
&\leq \frac{r^2}{2} |D^2 g|(B_r(x)).
\end{aligned}$$

The second inequality is a simple consequence of the identity $|\mu|(O) = \sup\{\int \phi \cdot \mathrm{d}\mu : \phi \in \mathscr{C}_c(O, S_d(\mathbb{R})), \|\phi\| \leq 1\}$ when μ is a $S_d(\mathbb{R})$-valued measure and O is open. Besides, since x is assumed to be a differentiability point of g, and a Lebesgue point of g and ∇g, we may pass to the inferior limit on the left-hand side and use Fatou's lemma to get:

$$\int_{B_r} |g(x + h) - (g(x) + \nabla g(x) \cdot h)| \, \mathrm{d}h \leq \frac{r^2}{2} |D^2 g|(B_r(x)). \tag{3.9}$$

We set $T_x u(y) = u(x) + \nabla u(x) \cdot (y - x)$ whenever a function u is differentiable at x. By convexity of g on $B_r(x)$, [EG15, Theorem 6.7 (ii)] yields[3]

$$\sup_{B_{r/2}(x)} |g - T_x g| \leq C \fint_{B_r(x)} |g - T_x g|, \tag{3.10}$$

for a constant $C = C(d)$, which together with (3.9) gives (3.8).

Step 2. Given a convex function g over Ω, let us show that:

$$\int_\Omega \sup_{B_{r/2}(x)} |g - T_x g| \, \mathrm{d}x \leq C \mathscr{H}^{d-1}(\partial\Omega) r^2 [g]_{\mathscr{C}^{0,1}(\Omega)}, \tag{3.11}$$

for some constant $C = C(d)$. Without loss of generality, we may assume that Ω is bounded and $[g]_{\mathscr{C}^{0,1}(\Omega)} < \infty$, otherwise the right-hand side is trivial. Since g is convex, Lebesgue-almost every point $x \in \Omega$ is a differentiability point of g by Rademacher's Theorem, and a Lebesgue point of g and ∇g by Lebesgue's Theorem, thus integrating (3.8) over $\Omega_r := \{x : \mathrm{dist}(x, \mathbb{R}^d \setminus \Omega) > r\}$ and using Fubini-Tonelli's Theorem,

$$\begin{aligned} \int_{\Omega_r} \sup_{B_{r/2}(x)} |g - T_x g| \, \mathrm{d}x &\leq C r^2 \int_\Omega \int_{B_r(y) \cap \Omega_r} \frac{1}{\omega_d r^d} \, \mathrm{d}x \, \mathrm{d}|D^2 g|(y) \\ &\leq C r^2 |D^2 g|(\Omega). \end{aligned} \tag{3.12}$$

Again, we consider the regularization $g_\varepsilon = g \star \eta_\varepsilon$ over Ω_ε for small $\varepsilon > 0$. By convexity of g_ε, we know that $|D^2 g_\varepsilon| \leq C \Delta g_\varepsilon$ on Ω_ε for some constant $C = C(d)$. By Stokes' Theorem,

$$\int_{\Omega_\varepsilon} \Delta g_\varepsilon = -\int_{\partial\Omega_\varepsilon} \nabla g_\varepsilon \cdot \nu \leq \mathscr{H}^{d-1}(\partial\Omega_\varepsilon)[g_\varepsilon]_{\mathscr{C}^{0,1}(\Omega_\varepsilon)} \leq \mathscr{H}^{d-1}(\partial\Omega)[g]_{\mathscr{C}^{0,1}(\Omega)},$$

where we have used the inequality $\mathscr{H}^{d-1}(\partial\Omega_\varepsilon) \leq \mathscr{H}^{d-1}(\partial\Omega)$, which holds because $\partial\Omega_\varepsilon = p(\partial\Omega)$ where p is the projection onto the convex set $\overline{\Omega}_\varepsilon$. Taking the limit $\varepsilon \to 0$ yields, by weak convergence $D^2 g_\varepsilon \llcorner \Omega_\varepsilon \rightharpoonup D^2 g$ in $\mathscr{C}_c(\Omega, S_d(\mathbb{R}))'$,

$$|D^2 g|(\Omega) \leq \liminf_{\varepsilon \to 0} \int_{\Omega_\varepsilon} |D^2 g_\varepsilon| \leq \liminf_{\varepsilon \to 0} C \int_{\Omega_\varepsilon} \Delta g_\varepsilon \leq C \mathscr{H}^{d-1}(\partial\Omega)[g]_{\mathscr{C}^{0,1}(\Omega)}.$$

Then by monotone convergence, if we take the limit $\varepsilon \to 0$ and report this in (3.12), we obtain:

$$\int_{\Omega_r} \sup_{B_{r/2}(x)} |g - T_x g| \, \mathrm{d}x \leq C \mathscr{H}^{d-1}(\partial\Omega) r^2 [g]_{\mathscr{C}^{0,1}(\Omega)},$$

for some constant $C = C(d)$.

Now we need to take care of the integral over $\Omega \setminus \Omega_r$. But if g is Lipschitz on Ω with constant $L = [g]_{\mathscr{C}^{0,1}(\Omega)} < \infty$, then for almost every $x \in \Omega \setminus \Omega_r$ and every $y \in B_{r/2}(x)$, we have

$$|g(y) - (g(x) + \nabla g(x) \cdot (y - x))| \leq 2L|y - x| \leq Lr.$$

[3] Notice that the proof works as soon as g is defined on $B_r(x)$, not necessarily on the whole space $\mathbb{R}^d$.

Finally, notice that since Ω is an open convex set, by the co-area formula (see [EG15, § 3.4]), for every $r > 0$,

$$\mathscr{L}^d(\Omega \setminus \Omega_r) = \int_0^r \mathscr{H}^{d-1}(\{\mathrm{dist}(\cdot, \mathbb{R}^d \setminus \Omega) = t\})\,\mathrm{d}t = \int_0^r \mathscr{H}^{d-1}(\partial\Omega_t)\,\mathrm{d}t \le r\mathscr{H}^{d-1}(\partial\Omega),$$

and as a consequence,

$$\begin{aligned}\int_\Omega \sup_{B_{r/2}(x)} |g - T_x g| &= \int_{\Omega\setminus\Omega_r} |g - T_x g| + \int_{\Omega_r} |g - T_x g| \\ &\le (Lr)(r\mathscr{H}^{d-1}(\partial\Omega)) + C\mathscr{H}^{d-1}(\partial\Omega) r^2 [g]_{\mathscr{C}^{0,1}(\Omega)} \\ &= C\mathscr{H}^{d-1}(\partial\Omega)[g]_{\mathscr{C}^{0,1}(\Omega)} r^2,\end{aligned}$$

for some constant $C = C(d)$ and (3.11) is proved.

Step 3. Let us conclude by establishing (3.7). Since f is λ-convex for some $\lambda \le 0$, we apply (3.11) to the convex function $g : x \mapsto f(x) - \lambda|x|^2/2$ to get

$$\begin{aligned}\int_\Omega \sup_{B_{r/2}(x)\cap\Omega} |f - T_x f| \le &\int_\Omega \sup_{B_{r/2}(x)\cap\Omega} |g - T_x g|\,\mathrm{d}x \\ &+ \frac{|\lambda|}{2}\int_\Omega \sup_{y\in B_{r/2}(x)\cap\Omega} ||y|^2 - (|x|^2 + 2x\cdot(y-x))|\,\mathrm{d}x \\ &\le Cr^2\mathscr{H}^{d-1}(\partial\Omega)[g]_{\mathscr{C}^{0,1}(\Omega)} + \frac{|\lambda|}{8} r^2 \mathscr{L}^d(\Omega) \\ &\le Cr^2\mathscr{H}^{d-1}(\partial\Omega)([f]_{\mathscr{C}^{0,1}(\Omega)} + |\lambda|\,\mathrm{diam}(\Omega)) + \frac{|\lambda|}{8} r^2\mathscr{L}^d(\Omega).\end{aligned}$$

Knowing that $\mathscr{L}^d(\Omega) \le \mathrm{diam}(\Omega)\mathscr{H}^{d-1}(\partial\Omega)$ since Ω is convex, we obtain (3.8) by replacing r by $2r$, for some (different) constant $C = C(d)$. □

We are now ready to establish Proposition 3.4.

Proof of Proposition 3.4. The measures $\mu^\pm$ are supported in some open and bounded convex sets $\Omega_0^\pm$ such that their closure $X^\pm := \mathrm{cl}(\Omega_0^\pm)$ are included in $\Omega^\pm$. Take $\gamma_0 \in \mathsf{Opt}(\mu^-, \mu^+)$ and a pair of c-conjugate Kantorovich potentials $(\phi, \psi) \in \mathscr{C}(X^-) \times \mathscr{C}(X^+)$. Such potentials exist because c is continuous on the compact set $X^- \times X^+$ (see for example [San15, Proposition 1.11]). We set $\lambda := \sup_{\boldsymbol{x}\neq\boldsymbol{x}'\in X^-\times X^+} \frac{|\nabla c(\boldsymbol{x}') - \nabla c(\boldsymbol{x})|}{|\boldsymbol{x}' - \boldsymbol{x}|} < \infty$, so that c is λ-concave on $X^- \times X^+$, which implies that ϕ, ψ are λ-concave on X^-, X^+ respectively. As such, they are differentiable Lebesgue-a.e. on Ω_0^-, Ω_0^+, thus μ^--a.e. and μ^+-a.e. respectively, which in turn implies that

$$E := c - \phi \oplus \psi$$

is differentiable γ_0-a.e. on $\Omega_0^- \times \Omega_0^+$ because $\gamma_0 \in \Pi(\mu^-,\ \mu^+)$ and $c \in \mathscr{C}^1(\Omega_0^- \times \Omega_0^+)$. Moreover, by optimality of γ_0 and (ϕ, ψ) we have $E \ge 0$ everywhere on $X^- \times X^+$

with equality on S_{γ_0}. As a consequence, for γ_0-a.e. $\boldsymbol{x_0} \in \Omega_0^- \times \Omega_0^+$, $E(\boldsymbol{x_0}) = 0$, E is differentiable at $\boldsymbol{x_0}$ and $\nabla E(\boldsymbol{x_0}) = 0$.

Take $\delta > 0$ and consider the block approximation γ^δ of γ_0 built in the proof of Proposition 3.1. Since by construction $W_\infty(\gamma_0, \gamma^\delta) \leq \delta$, there exists $\kappa \in \Pi(\gamma_0, \gamma^\delta)$ such that $\|\boldsymbol{x_0} - \boldsymbol{x}\| \leq \delta$ for κ-a.e. $(\boldsymbol{x_0}, \boldsymbol{x})$, and we know in particular that E is differentiable at $\boldsymbol{x_0}$, and $T_{\boldsymbol{x_0}} E(\boldsymbol{x}) = 0$ for κ-a.e. $(\boldsymbol{x_0}, \boldsymbol{x})$. Thus we have:

$$\begin{aligned}
\int_{\Omega_0^- \times \Omega_0^+} (c - \phi \oplus \psi) \, \mathrm{d}\gamma^\delta &= \int_{(\Omega_0^- \times \Omega_0^+)^2} (E(\boldsymbol{x}) - T_{\boldsymbol{x_0}} E(\boldsymbol{x})) \, \mathrm{d}\kappa(\boldsymbol{x_0}, \boldsymbol{x}) \\
&= \int_{(\Omega_0^- \times \Omega_0^+)^2} (c(\boldsymbol{x}) - T_{\boldsymbol{x_0}} c(\boldsymbol{x})) \, \mathrm{d}\kappa(\boldsymbol{x_0}, \boldsymbol{x}) \\
&\quad - \int_{\Omega_0^- \times \Omega_0^-} (\phi(x) - T_{x_0} \phi(x)) \, \mathrm{d}\kappa^-(x_0, x) \\
&\quad - \int_{\Omega_0^+ \times \Omega_0^+} (\psi(y) - T_{y_0} \psi(y)) \, \mathrm{d}\kappa^+(y_0, y),
\end{aligned}$$

where $\kappa^- = (\pi_1, \pi_3)_\# \kappa, \kappa^+ = (\pi_2, \pi_4)_\# \kappa$ and π_i denotes the projection on the i-th component. Since c is λ-concave on $\Omega_0^- \times \Omega_0^+$,

$$c(\boldsymbol{x}) - T_{\boldsymbol{x_0}} c(\boldsymbol{x}) \leq \frac{\lambda}{2} |\boldsymbol{x} - \boldsymbol{x_0}|^2,$$

and since ϕ, ψ are λ-concave on Ω_0^-, Ω_0^+ respectively, using Lemma 3.6 and recalling that $(\pi_1)_\# \kappa^\pm = \mu^\pm$,

$$\begin{aligned}
\int_{\Omega_0^+ \times \Omega_0^+} (c - \phi \oplus \psi) \, \mathrm{d}\gamma^\delta &\leq \frac{\lambda}{2} |\boldsymbol{x} - \boldsymbol{x_0}|^2 + \int_{\Omega_0^-} \sup_{x \in B_\delta(x_0) \cap \Omega_0^-} |\phi(x) - T_{x_0} \phi(x)| \, \mathrm{d}\mu^-(x_0) \\
&\qquad + \int_{\Omega_0^+} \sup_{y \in B_\delta(y_0) \cap \Omega_0^+} |\psi(y) - T_{y_0} \psi(y)| \, \mathrm{d}\mu^+(y_0) \\
&\leq \lambda \delta^2 + C\delta^2 \|\mu^-\|_{L^\infty(\Omega_0^-)} \mathscr{H}^{d-1}(\partial \Omega_0^-)([\phi]_{\mathscr{C}^{0,1}(\Omega_0^-)} + \lambda \operatorname{diam}(\Omega_0^-)) \\
&\qquad + C\delta^2 \|\mu^+\|_{L^\infty(\Omega_0^+)} \mathscr{H}^{d-1}(\partial \Omega_0^+)([\psi]_{\mathscr{C}^{0,1}(\Omega_0^+)} + \lambda \operatorname{diam}(\Omega_0^+)) \\
&\leq C' \delta^2,
\end{aligned}$$

for some constant $C' > 0$ which does not depend on δ.

Now, we apply (3.1) to assert that $H_\delta(\mu^+) \leq \log(N_\delta(X^+)) \leq d \log(1/\delta) + \log C$ for some constant $C > 0$ depending only on X^+, and we use γ^δ as a competitor in (EOT$_\varepsilon$) to obtain:

$$\begin{aligned}
v_\varepsilon \leq \int_{\Omega^- \times \Omega^+} c \, \mathrm{d}\gamma^\delta + \varepsilon H_\delta(\mu^+) &= v_0 + \int_{\Omega^- \times \Omega^+} (c - \phi \oplus \psi) \, \mathrm{d}\gamma^\delta + \varepsilon H_\delta(\mu^+) \\
&\leq v_0 + C' \delta^2 + \varepsilon H_\delta(\mu^+) \\
&= v_0 + C' \delta^2 + d\varepsilon \log(1/\delta) + \varepsilon \log C.
\end{aligned}$$

Finally, taking $\delta = \sqrt{\varepsilon}$ yields:

$$v_\varepsilon \leq v_0 + \frac{d}{2} \varepsilon \log(1/\varepsilon) + \varepsilon(C' + \log C). \qquad \square$$

4 Lower bound for infinitesimally twisted costs

In this section, we establish a lower bound that is analogous to the fine upper bound given in Section 3.2, for a class of $\mathscr{C}^2$ costs satisfying an *infinitesimal twist condition*, which includes situations where optimal plans are not necessarily given by a map. Our proof is based on the so-called Minty's trick asserting that the graph of a monotone operator is a 1-Lipschitz graph in a rotated chart (see [Min62]). It has been used in particular by Alberti and Ambrosio [AA99] to study the fine properties of monotone functions on $\mathbb{R}^d$, and more recently in optimal transport by McCann, Pass and Warren [MPW12] to show the rectifiability of optimal transport plans. We will exploit this trick a little bit further to show a quadratic detachment of the duality gap, leading to the lower bound that we seek. As a side remark, we deduce from the same trick a quantitative stability result in the quadratic case $c(x,y) = |x-y|^2$.

4.1 Proof of the lower bound

We start with the definition of the infinitesimal twist condition, corresponding to condition (A2) in [MTW05], and the non-degeneracy condition in [MPW12].

Definition 4.1. Given $c \in \mathscr{C}^2(\Omega^- \times \Omega^+)$ where $\Omega^\pm \subseteq \mathbb{R}^d$ are open sets, we say that c is *infinitesimally twisted* if $\nabla^2_{xy} c(x,y) := (\partial^2_{x_i y_j} c(x,y))_{i,j} \in M_d(\mathbb{R})$ is invertible for every $(x,y) \in \Omega^- \times \Omega^+$.

McCann, Pass and Warren have proved that for such a cost, the support of any optimal transport plan is locally Lipschitz (see [MPW12, Theorem 1.2]).

We closely follow the computations of [MPW12] leading to the proof of their main theorem, but we consider points which do not necessarily belong to the support of an optimal plan.

Lemma 4.2. *Let $c \in \mathscr{C}^2(\Omega^- \times \Omega^+)$ be an infinitesimally twisted cost, and $(\phi,\psi) \in \mathscr{C}(X^-) \times \mathscr{C}(X^+)$ be a pair of c-conjugate functions on compact convex sets $X^\pm \subseteq \Omega^\pm$. We set $E := c - \phi \oplus \psi$ on $X^- \times X^+$, $\Sigma := \{E = 0\}$, and for every $r > 0$,*

$$\kappa(r) := \sup_{\substack{(\boldsymbol{x},\boldsymbol{x'}) \in X^- \times X^+ \\ \|\boldsymbol{x'} - \boldsymbol{x}\| \le r}} \|\nabla^2_{xy} c(\boldsymbol{x'})^{-1} \nabla^2_{xy} c(\boldsymbol{x}) - \mathrm{Id}\| \in [0,\infty).$$

If $\bar{\boldsymbol{x}} \in X^- \times X^+$ and $\boldsymbol{x}, \boldsymbol{x'} \in B_r(\bar{\boldsymbol{x}}) \cap (X^- \times X^+)$, then

$$E(\boldsymbol{x'}) + E(\boldsymbol{x}) \ge |\Delta u|^2 - |\Delta v|^2 - \kappa(r)(|\Delta u|^2 + |\Delta v|^2), \tag{4.1}$$

where we have set $\Delta u := u(\boldsymbol{x'}) - u(\boldsymbol{x})$, $\Delta v := v(\boldsymbol{x'}) - v(\boldsymbol{x})$, and

$$u(\boldsymbol{x}) := \frac{1}{2}(x + \nabla^2_{xy} c(\bar{\boldsymbol{x}}) y), \quad v(\boldsymbol{x}) := \frac{1}{2}(x - \nabla^2_{xy} c(\bar{\boldsymbol{x}}) y), \quad \text{for every } \boldsymbol{x} = (x,y).$$

Proof. Take $\bar{\boldsymbol{x}} \in X^- \times X^+$, $\boldsymbol{x} = (x, y)$ and $\boldsymbol{x'} = (x', y')$ in $B_r(\bar{\boldsymbol{x}}) \cap (X^- \times X^+)$. Knowing that (ϕ, ψ) is a pair of c-conjugate functions, and using Taylor's integral formula,

$$\begin{aligned}
E(\boldsymbol{x'}) &= c(x', y') - \phi(x') - \psi(y') \\
&\geq c(x', y') - (c(x', y) - \psi(y)) - (c(x, y') - \phi(x)) \\
&= c(x', y') - c(x', y) - c(x, y') + c(x, y) - E(\boldsymbol{x}) \\
&= -E(\boldsymbol{x}) + (x' - x) \cdot \left[\int_0^1 \int_0^1 \nabla^2_{xy} c(x + (1-s)x', y + (1-t)y') \, \mathrm{d}s \, \mathrm{d}t \right] (y' - y).
\end{aligned}$$

With the notations introduced in the statement of the lemma, $x = u(\boldsymbol{x}) + v(\boldsymbol{x})$ and $y = \nabla^2_{xy} c(\bar{\boldsymbol{x}})^{-1}(u(\boldsymbol{x}) - v(\boldsymbol{x}))$, and similarly for $\boldsymbol{x'}$, so that:

$$\begin{aligned}
&E(\boldsymbol{x'}) + E(\boldsymbol{x}) \\
&\quad \geq (\Delta u + \Delta v) \cdot \\
&\qquad \left[\int_0^1 \int_0^1 \nabla^2_{xy} c(x + (1-s)x', y + (1-t)y') \nabla^2_{xy} c(\bar{\boldsymbol{x}})^{-1} \, \mathrm{d}s \, \mathrm{d}t \right] (\Delta u - \Delta v) \\
&\quad \geq |\Delta u|^2 - |\Delta v|^2 - \kappa(r) |\Delta u + \Delta v| |\Delta u - \Delta v| \\
&\quad \geq |\Delta u|^2 - |\Delta v|^2 - \kappa(r)(|\Delta u|^2 + |\Delta v|^2),
\end{aligned}$$

the last inequality resulting from the fact that $|a+b||a-b| \leq |a|^2 + |b|^2$ for every $a, b \in \mathbb{R}^d$, as can be seen by expanding and comparing the squares of the two sides. □

Remark 4.3. The rectifiability theorem of McCann, Pass and Warren [MPW12, Theorem 1.2] is an immediate consequence of Lemma 4.2 (to no surprise, since we have essentially followed their computations). Indeed, for any optimal plan $\gamma \in \Pi(\mu^-, \mu^+)$ with $\mu^\pm$ supported on compact convex sets $X^\pm \subseteq \Omega^\pm$, by taking $\bar{\boldsymbol{x}}, \boldsymbol{x}, \boldsymbol{x'} \in S_\gamma \subseteq \Sigma$, (4.1) yields

$$|\Delta u| \leq \sqrt{\frac{1 + \kappa(r)}{1 - \kappa(r)}} |\Delta v|,$$

and since $\kappa(r) \xrightarrow{r \to 0} 0$, v is a Lipschitz function of u, so that $B_r(\bar{\boldsymbol{x}}) \cap S_\gamma$ is included in a Lipschitz d-dimensional graph.

Notice however that (4.1) gives extra information: if we take for example $\boldsymbol{x} = \bar{\boldsymbol{x}}$, and we fix $v' = v = \bar{v}$ but we let the u component free, we get

$$E(\boldsymbol{x'}) \geq (1 - \kappa(r)) |u' - u|^2.$$

Thus we have a *quadratic growth* of E away from $\boldsymbol{x}$ in the direction given by the u coordinate. This is what we shall use to get the lower bound.

Proposition 4.4. *Let $\Omega^\pm$ be open convex subsets of $\mathbb{R}^d$, $c \in \mathscr{C}^2(\Omega^- \times \Omega^+)$ be an infinitesimally twisted cost, and $\mu^\pm \in \mathscr{P}(\Omega^\pm) \cap L^\infty(\Omega^\pm)$ with compact support in $\Omega^\pm$. There exists a constant $m \in [0, \infty)$ such that for every $\varepsilon > 0$,*

$$v_\varepsilon \geq v_0 + \frac{d}{2} \varepsilon \log(1/\varepsilon) - m\varepsilon. \tag{4.2}$$

Proof. The measures $\mu^\pm$ being concentrated on some compact convex subsets $X^\pm \subseteq \Omega^\pm$, consider a pair $(\phi, \psi) \in \mathscr{C}(X^-) \times \mathscr{C}(X^+)$ of c-conjugate Kantorovich potentials. Taking (ϕ, ψ) as competitor in (D_ε), we get the lower bound:

$$\begin{aligned} v_\varepsilon &\geq \int_{X^-} \phi \, \mathrm{d}\mu^- + \int_{X^+} \psi \, \mathrm{d}\mu^+ - \varepsilon \log \left(\int_{X^- \times X^+} e^{-\frac{E}{\varepsilon}} \, \mathrm{d}\mu^- \otimes \mu^+ \right) \\ &= v_0 - \varepsilon \log \left(\int_{X^- \times X^+} e^{-\frac{E}{\varepsilon}} \, \mathrm{d}\mu^- \otimes \mu^+ \right), \end{aligned}$$

where $E := c - \phi \oplus \psi$ on $X^- \times X^+$ as in Lemma 4.2. We are going to show that for some constant $C > 0$ and for every $\varepsilon > 0$,

$$\int_{X^- \times X^+} e^{-E/\varepsilon} \, \mathrm{d}\mu^- \otimes \mu^+ \leq C \varepsilon^{d/2},$$

which yields (4.2) with $m = \log(C)$.

We follow the notation introduced in the statement of Lemma 4.2, and consider in particular the functions u, v defined for a fixed $\bar{\boldsymbol{x}} \in X^- \times X^+$. We define the affine map $\Phi_{\bar{\boldsymbol{x}}} : \boldsymbol{x} \mapsto (u(\boldsymbol{x}), v(\boldsymbol{x})) - (u(\bar{\boldsymbol{x}}), v(\bar{\boldsymbol{x}}))$ whose Jacobian determinant is easily computed: $J\Phi_{\bar{\boldsymbol{x}}} = (-1/2)^d \det(\nabla^2_{xy} c(\bar{\boldsymbol{x}}))$. Thus $\Phi_{\bar{\boldsymbol{x}}}$ is an affine isomorphism and $|J(\Phi_{\bar{x}})^{-1}| \leq C_1 := 2^d \sup_{\boldsymbol{x} \in X^- \times X^+} |\det \nabla^2_{xy} c(\boldsymbol{x})|^{-1} < \infty$ because c is infinitesimally twisted. We set $E_{\bar{\boldsymbol{x}}} := E \circ \Phi_{\bar{\boldsymbol{x}}}^{-1}$ over $D_{\bar{\boldsymbol{x}}} := \Phi_{\bar{\boldsymbol{x}}}(X^- \times X^+)$ and define the open neighborhood of $\bar{\boldsymbol{x}}$, $P_{\bar{\boldsymbol{x}}} := \Phi_{\bar{\boldsymbol{x}}}^{-1}(B_r \times B_r) \subseteq B_{Lr}(\bar{\boldsymbol{x}})$, where $L := 2(1 \wedge \sup_{\boldsymbol{x} \in X^- \times X^+} \|(\nabla^2_{xy} c(\boldsymbol{x}))^{-1}\|)$ and $r > 0$ is chosen such that $\kappa(Lr) \leq 1/2$. By the change of variable formula we get:

$$\begin{aligned} &\int_{P_{\bar{\boldsymbol{x}}} \cap X^- \times X^+} e^{-E/\varepsilon} \, \mathrm{d}\mu^- \otimes \mu^+ \\ &\quad = \int_{(B_r \times B_r) \cap D_{\bar{\boldsymbol{x}}}} e^{-E(\Phi_{\bar{\boldsymbol{x}}}^{-1}(u,v))/\varepsilon} (\mu^- \otimes \mu^+)(\Phi_{\bar{\boldsymbol{x}}}^{-1}(u,v)) |J(\Phi_{\bar{\boldsymbol{x}}})^{-1}(u,v)| \, \mathrm{d}u \, \mathrm{d}v \\ &\quad \leq C_1 \|\mu^-\|_{L^\infty(\Omega^-)} \|\mu^+\|_{L^\infty(\Omega^+)} \int_{\pi_2((B_r \times B_r) \cap D_{\bar{\boldsymbol{x}}})} \int_{\{u \in B_r : (u,v) \in D_{\bar{\boldsymbol{x}}}\}} e^{-E_{\bar{\boldsymbol{x}}}(u,v)/\varepsilon} \, \mathrm{d}u \, \mathrm{d}v. \end{aligned}$$

Now, for every $v \in \pi_2((B_r \times B_r) \cap D_{\bar{\boldsymbol{x}}})$, consider u_v a minimizer of $E_{\bar{\boldsymbol{x}}}(\cdot, v)$ over $\{u \in \bar{B}_r : (u,v) \in D_{\bar{\boldsymbol{x}}}\}$. By (4.1) of Lemma 4.2, for every $u \in B_r$ such that $(u,v) \in D_{\bar{\boldsymbol{x}}}$,

$$E_{\bar{\boldsymbol{x}}}(u,v) \geq \frac{1}{2}(E_{\bar{\boldsymbol{x}}}(u_v, v) + E_{\bar{\boldsymbol{x}}}(u,v))) \geq \frac{1}{2}(1 - \kappa(Lr))|u - u_v|^2 \geq \frac{1}{4}|u - u_v|^2.$$

As a consequence we obtain:

$$\begin{aligned} \int_{\pi_2((B_r \times B_r) \cap D_{\bar{\boldsymbol{x}}})} \int_{\{u \in B_r : (u,v) \in D_{\bar{\boldsymbol{x}}}\}} e^{-E_{\bar{\boldsymbol{x}}}(u,v)/\varepsilon} \, \mathrm{d}u \, \mathrm{d}v &\leq \int_{\pi_2(B_r \times B_r) \cap D_{\bar{\boldsymbol{x}}}} \int_{B_r} e^{-|u-u_v|^2/2\varepsilon} \, \mathrm{d}u \, \mathrm{d}v \\ &\leq \omega_d r^d \varepsilon^{d/2} \int_{\mathbb{R}^d} e^{-|u|^2/2} \, \mathrm{d}u \\ &\leq C_2 r^d \varepsilon^{d/2}, \end{aligned}$$

for some constant $C_2 > 0$. The sets $\{P_{\bar{\boldsymbol{x}}}\}_{\bar{\boldsymbol{x}}\in\Sigma}$ form an open covering of the compact set $\Sigma \subseteq X^- \times X^+$, hence we may extract a finite covering $P_{\bar{\boldsymbol{x}}_1}, \dots, P_{\bar{\boldsymbol{x}}_N}$, so that $\Sigma \subseteq \bigcup_{i=1}^N P_{\bar{\boldsymbol{x}}_i}$ and:

$$\int_{\bigcup_{i=1}^N P_{\bar{\boldsymbol{x}}_i}\cap X^-\times X^+} e^{-E/\varepsilon}\mu^- \otimes \mu^+ \leq NC_1C_2\|\mu^-\|_{L^\infty(\Omega^-)}\|\mu^+\|_{L^\infty(\Omega^+)}r^d\varepsilon^{d/2} \leq C_3\varepsilon^{d/2},$$

for some constant $C_3 > 0$. Finally, since E is continuous and does not vanish on the compact set $K = (X^- \times X^+)\setminus\bigcup_{i=1}^N P_{\bar{\boldsymbol{x}}_i}$, it is bounded from below on K by some constant $C_4 > 0$. Therefore, for every $\varepsilon > 0$,

$$\int_{X^-\times X^+} e^{-E/\varepsilon}\,\mathrm{d}\mu^- \otimes \mu^+ \leq C_3\varepsilon^{d/2} + e^{-C_4/\varepsilon} \leq C\varepsilon^{d/2},$$

for some constant $C > 0$. This concludes the proof. □

4.2 Quantitative stability of optimal plans for the quadratic cost

This final paragraph is devoted to the quadratic-cost case for which combining Minty's trick and the upper bound on the entropic cost, one can obtain a quantitative estimate between the optimal entropic plan and optimal transport plans. Given two probability measures $\mu^\pm \in \mathscr{P}(\mathbb{R}^d)$ with compact support, taking as cost function $c(x,y) := \frac{1}{2}|x-y|^2$, the transport cost is merely the square 2-Wasserstein distance:

$$v_0 = \frac{1}{2}W_2^2(\mu^-,\mu^+) := \frac{1}{2}\inf_{\gamma\in\Pi(\mu^-,\,\mu^+)}\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^2\,\mathrm{d}\gamma(x,y).$$

It is well known (see Brenier [Bre91], McCann [McC95]) that there exists a convex lsc function f on $\mathbb{R}^d$ such that $\gamma \in \Pi(\mu^-,\ \mu^+)$ is optimal in the above quadratic OT problem if and only if

$$S_\gamma \subseteq \Gamma_f := \{(x,y)\in\mathbb{R}^d\times\mathbb{R}^d\ :\ f(x)+f^*(y) = x\cdot y\} = \{(x,y)\in\mathbb{R}^d\times\mathbb{R}^d\ :\ y\in\partial f(x)\}$$

where f^* is the Legendre transform of f. Kantorovich potentials are then given by

$$\phi(x) := \frac{1}{2}|x|^2 - f(x), \quad \psi(y) := \frac{1}{2}|y|^2 - f^*(y), \qquad (\forall x,y\in\mathbb{R}^d).$$

If μ^- is absolutely continuous with respect to the Lebesgue measure, f is differentiable μ^--a.e. and there is a unique optimal plan $\gamma := (\mathrm{Id},T)_\#\mu^-$ where $T := \nabla f$ is Brenier's optimal transport map from μ^- to μ^+. For $\varepsilon > 0$, the entropic cost reads

$$v_\varepsilon = \inf_{\gamma\in\Pi(\mu^-,\,\mu^+)}\left\{\frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^2\,\mathrm{d}\gamma(x,y) + \varepsilon\mathrm{Ent}(\gamma\,|\,\mu^-\otimes\mu^+)\right\},$$

and we denote as before by γ_ε the optimal entropic plan. Our goal is to give an estimate on how γ_ε fails to be concentrated on Γ_f for small $\varepsilon > 0$ in a sense to be made precise. First we observe that by nonnegativity of the entropic term, we have

$$v_\varepsilon - v_0 \geq \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^2\,\mathrm{d}\gamma_\varepsilon(x,y) - \frac{1}{2}W_2^2(\mu^-,\mu^+).$$

Defining the duality gap as in the previous section by:

$$E(x,y) := \frac{1}{2}|x-y|^2 - \phi(x) - \psi(y) = f(x) + f^*(y) - x \cdot y \qquad (\forall x, y \in \mathbb{R}^d),$$

one has $E \geq 0$ and $E(x,y) = 0$ if and only if $(x,y) \in \Gamma_f$, i.e. $y \in \partial f(x)$. Denoting by γ_0 an optimal transport plan, since $E = 0$ on S_{γ_0} and γ_ε and γ share the same marginals, we have

$$v_\varepsilon - v_0 \geq \frac{1}{2}\int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \,\mathrm{d}\gamma_\varepsilon(x,y) - \frac{1}{2}W_2^2(\mu^-, \mu^+) = \int_{\mathbb{R}^d \times \mathbb{R}^d} E(x,y)\,\mathrm{d}\gamma_\varepsilon(x,y). \quad (4.3)$$

In the event where ∇f is M-Lipschitz, which, by Caffarelli's regularity theory [Caf92; Caf96] holds true if $\mu^\pm$ have Hölder densities bounded away from zero on their supports and the latter are smooth and uniformly convex, arguing as Berman in [Ber21] and as Li and Nochetto in [LN21], who build upon an earlier argument of Gigli [Gig11], one can use the inequality

$$E(x,y) = f^*(y) - f^*(\nabla f(x)) - x \cdot (y - \nabla f(x)) \geq \frac{1}{2M}|y - \nabla f(x)|^2 \qquad (4.4)$$

(using the fact that $x \in \partial f^*(\nabla f(x))$ and $f^* - \frac{1}{2M}|\cdot|^2$ is convex as soon as ∇f is L-Lipschitz) to arrive at:

Proposition 4.5. *If Brenier's optimal transport map $T = \nabla f$ between the compactly supported probability measures μ^- and μ^+ is M-Lipschitz, denoting by γ_ε the optimal entropic plan between μ^- and μ^+, one has*

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |y - T(x)|^2 \,\mathrm{d}\gamma_\varepsilon(x,y) \leq M(d\varepsilon \log(1/\varepsilon) + O(\varepsilon)). \qquad (4.5)$$

In particular, if T_ε denotes the barycentric projection of γ_ε (i.e. $T_\varepsilon(x)$ is the conditional expectation of Y given $X = x$ with (X,Y) distributed according to γ_ε), there holds

$$\|T_\varepsilon - T\|^2_{L^2(\mu^-)} \leq M(d\varepsilon \log(1/\varepsilon) + O(\varepsilon)). \qquad (4.6)$$

Proof. Inequality (4.5) follows directly from (4.3), (4.4) and the upper bound $v_\varepsilon - v_0 \leq d\varepsilon \log(1/\varepsilon) + O(\varepsilon)$ from Proposition 3.1 (note also that if $\mu^\pm$ have bounded densities, one can improve the factor d by $\frac{d}{2}$ thanks to Proposition 3.4). Finally (4.6) directly follows from (4.5) and Jensen's inequality:

$$\begin{aligned}
\int_{\mathbb{R}^d \times \mathbb{R}^d} |y - T(x)|^2 \,\mathrm{d}\gamma_\varepsilon(x,y) &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |y - T(x)|^2 \,\mathrm{d}\gamma_\varepsilon^x(y)\,\mathrm{d}\mu^-(x) \\
&\geq \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} (y - T(x))\,\mathrm{d}\gamma_\varepsilon^x(y) \right|^2 \mathrm{d}\mu^-(x) \\
&= \int_{\mathbb{R}^d} |T_\varepsilon(x) - T(x)|^2 \,\mathrm{d}\mu^-(x),
\end{aligned}$$

where γ_ε^x is the disintegration of γ_ε with respect to its first marginal. □

Of course the requirement that the convex potential f is smooth is quite demanding and cannot be taken for granted in general. Yet, if f is an arbitrary lsc convex function, one can take advantage of Minty's trick to have a quadratic detachment lower bound on E as we did in Lemma 4.2.

Proposition 4.6. *If $\mu^{\pm} \in \mathscr{P}(\mathbb{R}^d)$ are compactly supported probability measures, and if we denote by γ_ε the optimal entropic plan from μ^- to μ^+, the following stability bound holds:*

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - (\mathrm{Id} + \partial f)^{-1}(x+y)|^2 \, \mathrm{d}\gamma_\varepsilon(x,y) \leq (d\varepsilon \log(1/\varepsilon) + O(\varepsilon)),$$

where $(\mathrm{Id} + \partial f)^{-1}$ is the (single-valued) resolvent associated with f.

Proof. Let us first observe that when c is the quadratic cost (4.1) takes the form

$$E(x,y) \geq -E(x',y') - (x'-x)\cdot(y'-y) \qquad (\forall x,y,x',y' \in \mathbb{R}^d).$$

Now we observe that the resolvent $(\mathrm{Id} + \partial f)^{-1}$ is a single-valued 1-Lipschitz map and that $E(x',y')$ vanishes if and only if $y' + x' \in x' + \partial f(x')$, hence if and only if $x' = (\mathrm{Id} + \partial f)^{-1}(x' + y')$. As a consequence if we choose in the inequality above $x' = (\mathrm{Id} + \partial f)^{-1}(x+y)$ and $y' = x + y - x'$, we have $E(x',y') = 0$ and we get the quadratic detachment bound:

$$E(x,y) \geq |x - (\mathrm{Id} + \partial f)^{-1}(x+y)|^2, \qquad (\forall x,y \in \mathbb{R}^d). \tag{4.7}$$

Hence, from (4.7), (4.3) and Proposition 3.1, we obtain the desired inequality (and again one can improve the factor d by $\frac{d}{2}$ thanks to Proposition 3.4 if $\mu^{\pm}$ have bounded densities). □

Acknowledgments: G.C. acknowledges the support of the Lagrange Mathematics and Computing Research Center.

References

[AA99] G. Alberti and L. Ambrosio. "A geometrical approach to monotone functions in $\mathbb{R}^n$". In: *Mathematische Zeitschrift* 230.2 (1999), pp. 259–316. ISSN: 1432-1823. DOI: 10.1007/PL00004691.

[Ada+11] S. Adams, N. Dirr, M. A. Peletier, and J. Zimmer. "From a Large-Deviations Principle to the Wasserstein Gradient Flow: A New Micro-Macro Passage". In: *Communications in Mathematical Physics* 307.3 (2011), p. 791. ISSN: 1432-0916. DOI: 10.1007/s00220-011-1328-4.

[AFP00] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Mathematical Monographs. Oxford, New York: Oxford University Press, 2000. 452 pp. ISBN: 978-0-19-850245-6.

[ANS22] J. M. Altschuler, J. Niles-Weed, and A. J. Stromme. "Asymptotics for Semidiscrete Entropic Optimal Transport". In: *SIAM Journal on Mathematical Analysis* 54.2 (2022), pp. 1718–1741. ISSN: 0036-1410. DOI: 10.1137/21M1440165.

[Arn+20] M. Arnaudon, A. B. Cruzeiro, C. Léonard, and J.-C. Zambrini. "An Entropic Interpolation Problem for Incompressible Viscous Fluids". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 56.3 (2020), pp. 2211–2235. ISSN: 0246-0203. DOI: 10.1214/19-AIHP1036.

[Bac+20] J. Backhoff, G. Conforti, I. Gentil, and C. Léonard. "The Mean Field Schrödinger Problem: Ergodic Behavior, Entropy Estimates and Functional Inequalities". In: *Probability Theory and Related Fields* 178.1 (2020), pp. 475–530. ISSN: 1432-2064. DOI: 10.1007/s00440-020-00977-8.

[BCN19] J.-D. Benamou, G. Carlier, and L. Nenna. "Generalized Incompressible Flows, Multi-Marginal Transport and Sinkhorn Algorithm". In: *Numerische Mathematik* 142.1 (2019), pp. 33–54. ISSN: 0945-3245. DOI: 10.1007/s00211-018-0995-x.

[Ben+15] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. "Iterative Bregman Projections for Regularized Transportation Problems". In: *SIAM Journal on Scientific Computing* 37.2 (2015), A1111–A1138. ISSN: 1064-8275. DOI: 10.1137/141000439.

[Ber21] R. J. Berman. "Convergence Rates for Discretized Monge–Ampère Equations and Quantitative Stability of Optimal Transport". In: *Foundations of Computational Mathematics* 21.4 (2021), pp. 1099–1140. ISSN: 1615-3383. DOI: 10.1007/s10208-020-09480-x.

[BGN22] E. Bernton, P. Ghosal, and M. Nutz. *Entropic Optimal Transport: Geometry and Large Deviations*. Comment: Forthcoming in 'Duke Mathematical Journal'. 2022. URL: http://arxiv.org/abs/2102.04397.

[BL21] A. Baradat and H. Lavenant. *Regularized Unbalanced Optimal Transport as Entropy Minimization with Respect to Branching Brownian Motion*. 2021. URL: http://arxiv.org/abs/2111.01666.

[BL92] J. M. Borwein and A. S. Lewis. "Decomposition of Multivariate Functions". In: *Canadian Journal of Mathematics* 44.3 (1992), pp. 463–482. ISSN: 0008-414X, 1496-4279. DOI: 10.4153/CJM-1992-030-9.

[BLN94] J. M. Borwein, A. S. Lewis, and R. D. Nussbaum. "Entropy Minimization, DAD Problems, and Doubly Stochastic Kernels". In: *Journal of Functional Analysis* 123.2 (1994), pp. 264–307. ISSN: 0022-1236. DOI: 10.1006/jfan.1994.1089.

[BM20] A. Baradat and L. Monsaingeon. "Small Noise Limit and Convexity for Generalized Incompressible Flows, Schrödinger Problems, and Optimal Transport". In: *Archive for Rational Mechanics and Analysis* 235.2 (2020), pp. 1357–1403. ISSN: 1432-0673. DOI: 10.1007/s00205-019-01446-w.

[Bre91] Y. Brenier. "Polar Factorization and Monotone Rearrangement of Vector-Valued Functions". In: *Communications on Pure and Applied Mathematics* 44.4 (1991), pp. 375–417. ISSN: 1097-0312. DOI: 10.1002/cpa.3160440402.

[BT16] B. Buffoni and J. Toland. *Analytic Theory of Global Bifurcation: An Introduction*. Princeton University Press, 2016. ISBN: 978-1-4008-8433-9. DOI: 10.1515/9781400884339.

[Caf92] L. A. Caffarelli. "Boundary Regularity of Maps with Convex Potentials". In: *Communications on Pure and Applied Mathematics* 45.9 (1992), pp. 1141–1151. ISSN: 1097-0312. DOI: 10.1002/cpa.3160450905.

[Caf96] L. A. Caffarelli. "Boundary Regularity of Maps with Convex Potentials–II". In: *Annals of Mathematics* 144.3 (1996), pp. 453–496. ISSN: 0003-486X. DOI: 10.2307/2118564.

[Car+17] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer. "Convergence of Entropic Schemes for Optimal Transport and Gradient Flows". In: *SIAM Journal on Mathematical Analysis* 49.2 (2017), pp. 1385–1418. ISSN: 0036-1410. DOI: 10.1137/15M1050264.

[Car22] G. Carlier. "On the Linear Convergence of the Multimarginal Sinkhorn Algorithm". In: *SIAM Journal on Optimization* (2022), pp. 786–794. ISSN: 1052-6234. DOI: 10.1137/21M1410634.

[CCT21] A. Chiarini, G. Conforti, and L. Tamanini. "Schrödinger Problem for Lattice Gases: A Heuristic Point of View". In: *Geometric Science of Information*. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021, pp. 891–899. ISBN: 978-3-030-80209-7. DOI: 10.1007/978-3-030-80209-7_95.

[Chi+20] L. Chizat, P. Roussillon, F. Léger, F.-X. Vialard, and G. Peyré. *Faster Wasserstein Distance Estimation with the Sinkhorn Divergence*. 2020. URL: http://arxiv.org/abs/2006.08172.

[CL20] G. Carlier and M. Laborde. "A Differential Approach to the Multi-Marginal Schrödinger System". In: *SIAM Journal on Mathematical Analysis* 52.1 (2020), pp. 709–717. ISSN: 0036-1410. DOI: 10.1137/19M1253800.

[CM94] R. Cominetti and J. S. Martín. "Asymptotic Analysis of the Exponential Penalty Trajectory in Linear Programming". In: *Mathematical Programming* 67.1 (1994), pp. 169–187. ISSN: 1436-4646. DOI: 10.1007/BF01582220.

[CMS21] R. Chetrite, P. Muratore-Ginanneschi, and K. Schwieger. "E. Schrödinger's 1931 Paper "On the Reversal of the Laws of Nature" ["Über Die Umkehrung Der Naturgesetze", Sitzungsberichte Der Preussischen Akademie Der Wissenschaften, Physikalisch-Mathematische Klasse, 8 N9 144–153]". In: *The European Physical Journal H* 46.1 (2021), p. 28. ISSN: 2102-6467. DOI: 10.1140/epjh/s13129-021-00032-7.

[Csi75] I. Csiszar. "I-Divergence Geometry of Probability Distributions and Minimization Problems". In: *The Annals of Probability* 3.1 (1975), pp. 146–158. ISSN: 0091-1798, 2168-894X. DOI: 10.1214/aop/1176996454.

[CT21] G. Conforti and L. Tamanini. "A Formula for the Time Derivative of the Entropic Cost and Applications". In: *Journal of Functional Analysis* 280.11 (2021), p. 108964. ISSN: 0022-1236. DOI: 10.1016/j.jfa.2021.108964.

[Cut13] M. Cuturi. "Sinkhorn Distances: Lightspeed Computation of Optimal Transport". In: *Advances in Neural Information Processing Systems*. Vol. 26. Curran Associates, Inc., 2013.

[DG20] S. Di Marino and A. Gerolin. *Optimal Transport Losses and Sinkhorn Algorithm with General Convex Regularization*. 2020. URL: http://arxiv.org/abs/2007.00976.

[EG15] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions, Revised Edition*. New York: Chapman and Hall/CRC, 2015. 313 pp. ISBN: 978-0-429-16148-3. DOI: 10.1201/b18333.

[EMR15] M. Erbar, J. Maas, and M. Renger. "From Large Deviations to Wasserstein Gradient Flows in Multiple Dimensions". In: *Electronic Communications in Probability* 20 (none 2015), pp. 1–12. ISSN: 1083-589X, 1083-589X. DOI: 10.1214/ECP.v20-4315.

[EN21] S. Eckstein and M. Nutz. *Quantitative Stability of Regularized Optimal Transport and Convergence of Sinkhorn's Algorithm*. 2021. URL: http://arxiv.org/abs/2110.06798.

[Fal97] K. Falconer. *Techniques in Fractal Geometry*. Chichester: John Wiley & Sons, 1997. xvii + 256. ISBN: 978-0-471-95724-9.

[FG97] H. Föllmer and N. Gantert. "Entropy Minimization and Schrödinger Processes in Infinite Dimensions". In: *The Annals of Probability* 25.2 (1997), pp. 901–926. ISSN: 0091-1798, 2168-894X. DOI: 10.1214/aop/1024404423.

[FL89] J. Franklin and J. Lorenz. "On the Scaling of Multidimensional Matrices". In: *Linear Algebra and its Applications*. Special Issue Dedicated to Alan J. Hoffman 114–115 (1989), pp. 717–735. ISSN: 0024-3795. DOI: 10.1016/0024-3795(89)90490-4.

[Gen+20] I. Gentil, C. Léonard, L. Ripani, and L. Tamanini. "An Entropic Interpolation Proof of the HWI Inequality". In: *Stochastic Processes and their Applications* 130.2 (2020), pp. 907–923. ISSN: 0304-4149. DOI: 10.1016/j.spa.2019.04.002.

[Gig11] N. Gigli. "On Hölder Continuity-in-Time of the Optimal Transport Map towards Measures along a Curve". In: *Proceedings of the Edinburgh Mathematical Society* 54.2 (2011), pp. 401–409. ISSN: 1464-3839, 0013-0915. DOI: 10.1017/S001309150800117X.

[GNB21] P. Ghosal, M. Nutz, and E. Bernton. *Stability of Entropic Optimal Transport and Schrödinger Bridges*. 2021. URL: http://arxiv.org/abs/2106.03670.

[GT20] N. Gigli and L. Tamanini. "Benamou–Brenier and duality formulas for the entropic cost on $\mathsf{RCD}^*(K,N)$ spaces". In: *Probability Theory and Related Fields* 176.1 (2020), pp. 1–34. ISSN: 1432-2064. DOI: 10.1007/s00440-019-00909-1.

[GT21] N. Gigli and L. Tamanini. "Second order differentiation formula on $\mathsf{RCD}^*(K,N)$ spaces". In: *Journal of the European Mathematical Society* 23.5 (2021), pp. 1727–1795. ISSN: 1435-9855. DOI: 10.4171/JEMS/1042.

[Léo01] C. Léonard. "Minimization of Energy Functionals Applied to Some Inverse Problems". In: *Applied Mathematics and Optimization* 44.3 (2001), pp. 273–297. ISSN: 1432-0606. DOI: 10.1007/s00245-001-0019-5.

[Léo12] C. Léonard. "From the Schrödinger Problem to the Monge–Kantorovich Problem". In: *Journal of Functional Analysis* 262.4 (2012), pp. 1879–1920. ISSN: 0022-1236. DOI: 10.1016/j.jfa.2011.11.026.

[Léo14] C. Léonard. "A Survey of the Schrödinger Problem and Some of Its Connections with Optimal Transport". In: *Discrete and Continuous Dynamical Systems* 34.4 (2014), p. 1533. DOI: 10.3934/dcds.2014.34.1533.

[LM22] D. Lorenz and H. Mahler. "Orlicz Space Regularization of Continuous Optimal Transport Problems". In: *Applied Mathematics & Optimization* 85.2 (2022), p. 14. ISSN: 1432-0606. DOI: 10.1007/s00245-022-09826-7.

[LN21] W. Li and R. H. Nochetto. "Quantitative Stability and Error Estimates for Optimal Transport Plans". In: *IMA Journal of Numerical Analysis* 41.3 (2021), pp. 1941–1965. ISSN: 0272-4979. DOI: 10.1093/imanum/draa045.

[McC95] R. J. McCann. "Existence and Uniqueness of Monotone Measure-Preserving Maps". In: *Duke Mathematical Journal* 80.2 (1995), pp. 309–323. ISSN: 0012-7094, 1547-7398. DOI: 10.1215/S0012-7094-95-08013-2.

[MG20] S. D. Marino and A. Gerolin. "An Optimal Transport Approach for the Schrödinger Bridge Problem and Convergence of Sinkhorn Algorithm". In: *Journal of Scientific Computing* 85.2 (2020), p. 27. ISSN: 1573-7691. DOI: 10.1007/s10915-020-01325-7.

[Mik04] T. Mikami. "Monge's Problem with a Quadratic Cost by the Zero-Noise Limit of h-Path Processes". In: *Probability Theory and Related Fields* 129.2 (2004), pp. 245–260. ISSN: 1432-2064. DOI: 10.1007/s00440-004-0340-4.

[Min62] G. J. Minty. "Monotone (Nonlinear) Operators in Hilbert Space". In: *Duke Mathematical Journal* 29.3 (1962), pp. 341–346. ISSN: 0012-7094, 1547-7398. DOI: 10.1215/S0012-7094-62-02933-2.

[MPW12] R. J. McCann, B. Pass, and M. Warren. "Rectifiability of Optimal Transportation Plans". In: *Canadian Journal of Mathematics* 64.4 (2012), pp. 924–934. ISSN: 0008-414X, 1496-4279. DOI: 10.4153/CJM-2011-080-6.

[MT08] T. Mikami and M. Thieullen. "Optimal Transportation Problem by Stochastic Optimal Control". In: *SIAM Journal on Control and Optimization* 47.3 (2008), pp. 1127–1139. ISSN: 0363-0129. DOI: 10.1137/050631264.

[MTV20] L. Monsaingeon, L. Tamanini, and D. Vorotnikov. *The Dynamical Schrödinger Problem in Abstract Metric Spaces*. 2020. URL: http://arxiv.org/abs/2012.12005.

[MTW05] X.-N. Ma, N. S. Trudinger, and X.-J. Wang. "Regularity of Potential Functions of the Optimal Transportation Problem". In: *Archive for Rational Mechanics and Analysis* 177.2 (2005), pp. 151–183. ISSN: 1432-0673. DOI: 10.1007/s00205-005-0362-9.

[Nut22] M. Nutz. *Introduction to Entropic Optimal Transport*. 2022. URL: https://www.math.columbia.edu/~mnutz/docs/EOT_lecture_notes.pdf.

[NW21] M. Nutz and J. Wiesel. "Entropic Optimal Transport: Convergence of Potentials". In: *Probability Theory and Related Fields* (2021). ISSN: 1432-2064. DOI: 10.1007/s00440-021-01096-8.

[NW22] M. Nutz and J. Wiesel. *Stability of Schrödinger Potentials and Convergence of Sinkhorn's Algorithm*. 2022. URL: http://arxiv.org/abs/2201.10059.

[Pal19] S. Pal. *On the Difference between Entropic Cost and the Optimal Transport Cost*. 2019. URL: http://arxiv.org/abs/1905.12206.

[PC19] G. Peyré and M. Cuturi. "Computational Optimal Transport: With Applications to Data Science". In: *Foundations and Trends in Machine Learning* 11.5-6 (2019), pp. 355–607. ISSN: 1935-8237, 1935-8245. DOI: 10.1561/2200000073.

[Rén59] A. Rényi. "On the Dimension and Entropy of Probability Distributions". In: *Acta Mathematica Academiae Scientiarum Hungarica* 10.1 (1959), pp. 193–215. ISSN: 1588-2632. DOI: 10.1007/BF02063299.

[RT98] L. Rüschendorf and W. Thomsen. "Closedness of Sum Spaces and the Generalized Schrödinger Problem". In: *Theory of Probability & Its Applications* 42.3 (1998), pp. 483–494. ISSN: 0040-585X. DOI: 10.1137/S0040585X97976301.

[RTC17] A. Ramdas, N. G. Trillos, and M. Cuturi. "On Wasserstein Two-Sample Testing and Related Families of Nonparametric Tests". In: *Entropy* 19.2 (2 2017), p. 47. ISSN: 1099-4300. DOI: 10.3390/e19020047.

[San15] F. Santambrogio. *Optimal Transport for Applied Mathematicians*. Vol. 87. Progress in Nonlinear Differential Equations and Their Applications. Cham: Springer International Publishing, 2015. ISBN: 978-3-319-20827-5 978-3-319-20828-2. DOI: 10.1007/978-3-319-20828-2.

[Sch31] Schrödinger, Erwin. "Über Die Umkehrung Der Naturgesetze". In: *Angewandte Chemie*. W. de Gruyter, Berlin. Preis RM. 1 44.30 (1931), pp. 636–636. ISSN: 1521-3757. DOI: 10.1002/ange.19310443014.

[Sch32] E. Schrödinger. “Sur la théorie relativiste de l’électron et l’interprétation de la mécanique quantique”. In: *Annales de l’institut Henri Poincaré* 2.4 (1932), pp. 269–310.

[Sin64] R. Sinkhorn. “A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices”. In: *The Annals of Mathematical Statistics* 35.2 (1964), pp. 876–879. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177703591.

[You82] L.-S. Young. “Dimension, Entropy and Lyapunov Exponents”. In: *Ergodic Theory and Dynamical Systems* 2.1 (1982), pp. 109–124. ISSN: 1469-4417, 0143-3857. DOI: 10.1017/S0143385700009615.