



HAL
open science

Convergence Rates of the Regularized Optimal Transport : Disentangling Suboptimality and Entropy

Hugo Malamut, Maxime Sylvestre

► **To cite this version:**

Hugo Malamut, Maxime Sylvestre. Convergence Rates of the Regularized Optimal Transport : Disentangling Suboptimality and Entropy. 2023. hal-04114127v1

HAL Id: hal-04114127

<https://hal.science/hal-04114127v1>

Preprint submitted on 8 Jun 2023 (v1), last revised 9 Jun 2023 (v2)

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.

Convergence Rates of the Regularized Optimal Transport : Disentangling Suboptimality and Entropy

Hugo Malamut, Maxime Sylvestre

June 8, 2023

Abstract

We study the convergence of the transport plans γ_ε towards γ_0 as well as the cost of the entropy-regularized optimal transport (c, γ_ε) towards (c, γ_0) as the regularization parameter ε vanishes in the setting of finite entropy marginals. We show that under the assumption of infinitesimally twisted cost and compactly supported marginals the distance $W_2(\gamma_\varepsilon, \gamma_0)$ is asymptotically greater than $C\sqrt{\varepsilon}$ and the suboptimality $(c, \gamma_\varepsilon) - (c, \gamma_0)$ is of order ε . In the quadratic cost case the compactness assumption is relaxed into a moment of order $2 + \delta$ assumption. Moreover, in the case of a Lipschitz transport map for the non-regularized problem, the distance $W_2(\gamma_\varepsilon, \gamma_0)$ converges to 0 at rate $\sqrt{\varepsilon}$. Finally, if in addition the marginals have finite Fisher information, we prove $(c, \gamma_\varepsilon) - (c, \gamma_0) \sim d\varepsilon/2$ and we provide a companion expansion of $H(\gamma_\varepsilon)$. These results are achieved by disentangling the role of the cost and the entropy in the regularized problem.

Contents

1	Introduction	2
2	Exact asymptotics from the Schrödinger problem	5
3	Rates for the quadratic cost	8
4	Rates for infinitesimally twisted costs	14

Notations

Ω, X	Ω is an open subset of \mathbb{R}^d and X a compact subset of Ω
$\mathcal{P}(\Omega)$	Probability measures on Ω , absolutely continuous ($\mathcal{P}_{ac}(\Omega)$), with finite variance ($\mathcal{P}_2(\Omega)$)
\mathcal{H}^k	Hausdorff measure of dimension k
$W_2(\mu, \nu)$	2-Wasserstein distance between μ and ν
$H(\mu \nu)$	relative entropy of μ with respect to ν (see equation (9)). Sometimes $H(\mu) := H(\mu \mathcal{H}^k)$ for some $k \in \mathbb{N}$
$I(\mu)$	Fisher information of μ , see equation (9)
$\ \Phi\ _{op}$	Operator norm of the affine transformation Φ
$\Pi(\mu_0, \mu_1)$	Transport plans between μ_0 and μ_1

We would like to thank G. Carlier and P. Pegon for fruitful discussions and helpful advices.

c	Cost function, $c \in \mathcal{C}^2(\Omega)$. In section 2 and 3, $c(x, y) := \frac{1}{2}\ x - y\ ^2$ (quadratic cost)
$T = \nabla f$	Brenier's optimal map from μ_0 to μ_1 , f is a convex function
φ, ψ	a pair of Kantorovich potentials (see equation (10))
E	a duality gap function defined by $E := c - (\varphi \oplus \psi)$
Σ	a set of optimal pairings $\Sigma := \{(x, y) / E(x, y) = 0\}$
OT_ε	ε -entropic transport cost from μ_0 to μ_1 (see equation (ε EOT))
γ_ε	Optimal transport plan for ε -Entropic Optimal Transport (ε EOT)
$v_t^\varepsilon, \rho_t^\varepsilon$	Solutions of ε -entropic Benamou Brenier formula (ε BB)
$N_k(\mu)$	Entropy power function of μ : if $H(\mu \mathcal{H}^k) < +\infty$, then $N_k(\mu) = \frac{1}{2\pi e} e^{-\frac{2}{k}H(\mu \mathcal{H}^k)}$
$m_k(\mu)$	Moment of order k of the probability μ
H_m	Mean entropy $H_m := \frac{1}{2}(H(\mu_0 \mathcal{H}^d) + H(\mu_1 \mathcal{H}^d))$
$C(a, b, c)$	A constant that only depends on the terms a, b and c

1 Introduction

We study the regularized optimal transport problem for a cost $c \in \mathcal{C}^2(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R})$

$$\inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int cd\gamma + \varepsilon H(\gamma | \mu_0 \otimes \mu_1) \quad (1)$$

where the infimum is taken over all measures $\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with marginals μ_0, μ_1 . Here $H(\cdot | \cdot)$ is the relative entropy also known as Kullback-Leibler divergence. In this paper we will focus on the case where μ_0, μ_1 have finite entropy with respect to \mathcal{H}^d , the Lebesgue measure, $H(\mu_i | \mathcal{H}^d) < \infty$ and finite order two moments. In that case the minimizer γ_ε is the same as if $\mu_0 \otimes \mu_1$ was replaced by the Lebesgue measure \mathcal{H}^{2d} in the entropy term, and hence we consider

$$OT_\varepsilon := \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int cd\gamma + \varepsilon H(\gamma | \mathcal{H}^{2d}) \quad (\varepsilon\text{EOT})$$

Note that $\varepsilon = 0$ yields the classical optimal transport problem. In this case the minimizer need not to be unique and γ_0 will sometimes denote any of them. We are interested in deriving rates for the cost term $\int cd\gamma_\varepsilon = (c, \gamma_\varepsilon)$ as well as the 2-Wasserstein distance between γ_ε and γ_0 , when it is unique.

In the last decade this problem has witnessed a rapid increase in interest. It has proved to be an efficient way to approximate OT problems, especially from a computational viewpoint. The celebrated Sinkhorn's algorithm [34] was applied in this framework in the pioneering works [14, 2]. The good convergence guarantees [19, 27] cemented the success of EOT and its applications.

Clearly EOT is a perturbation of classical OT thus it is natural to study the behaviour of this problem as ε vanishes. In this direction several aspects deserved to be studied such as the convergence of optimal values, potentials (optimizers of the dual problem) and optimal plans, possibly with quantitative rates. In the direction of convergence of optimal values recent contributions have thoroughly treated the issue: in the quadratic case, under regularity assumptions, the link between EOT and the Schrödinger problem [24] has allowed to find a second order [17] and more recently a third order [12, 9] expansion of the value OT_ε in ε . The second order expansion has been generalized to other cost functions [31], and the first order term has been obtained under very mild assumptions on the

cost function and the marginals [7, 15]. Those articles focus on the value problem of (εEOT) .

Our main objective is to disentangle the role of the cost $\int cd\gamma_\varepsilon$ and the entropy $H(\gamma_\varepsilon | \mathcal{L})$ in order to derive rate of convergence for both. The cost term is of interest itself because it is a faster converging approximation of OT_0 . The entropy term also allows to lower bound $W_2(\gamma_\varepsilon, \gamma_0)$, the Wasserstein distance between the entropic optimal transport plan γ_ε and the optimal transport plan γ_0 . The study of the convergence of the cost term in EOT has also been done recently in [1] for the semidiscrete case which grants an ε^2 rate. They also derive convergence rates for the Kantorovich potentials. In the discrete case the rate of convergence is exponential [10, 36]. In the continuous case, to our knowledge, no asymptotic rate for the suboptimality $\int cd\gamma_\varepsilon - \int cd\gamma_0$ was known but a rate of order ε was suspected based on simple examples such as Gaussian measures. We tackle the problem of sizing the suboptimality in this article.

1.1 Main results

The aim of this paper is to provide tight rates of convergence on the Wasserstein distance $W_2(\gamma_\varepsilon, \gamma_0)$, the cost term (c, γ_ε) and the entropy term $H(\gamma_\varepsilon | \mathcal{H}^{2d})$. Let $\mu_0, \mu_1 \in \mathcal{P}_{ac}(\mathbb{R}^d)$ such that $H(\mu_i | \mathcal{H}^d) < \infty$. It is known [7, 15] that under mild assumptions (see lemma 3.6) on μ_0, μ_1 ,

$$OT_\varepsilon = (c, \gamma_\varepsilon) + \varepsilon H(\gamma_\varepsilon | \mathcal{H}^{2d}) \leq OT_0 - \frac{d}{2}\varepsilon \ln(\varepsilon) + O(\varepsilon) \quad (2)$$

For the quadratic cost and under the stronger assumption of finite Fisher information for the marginals, we have (see [17, claim 4.1], see also [31] for other costs and different hypothesis)

$$OT_\varepsilon = OT_0 - \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) + \varepsilon H_m + o(\varepsilon) \quad (3)$$

where $H_m := \frac{1}{2}[H(\mu_0 | \mathcal{H}^d) + H(\mu_1 | \mathcal{H}^d)]$. Our goal are to disentangle the role of the cost term (c, γ_ε) and of the entropy term $H(\gamma_\varepsilon | \mathcal{H}^{2d})$ in those rates rates of convergence.

For clarity we present here the three main results present in the article. The first gives a second order expansion of the entropy and the cost of the entropic optimal transport. It relies on known expansions of the Benamou-Brenier (εBB) formulation of the entropic optimal transport problem. The proof of this result can be found in section 2

Theorem (Theorem 2.3). *Suppose that the cost is quadratic, that is $c(x, y) = \frac{1}{2}\|x - y\|^2$. Further assume that $I(\mu_i) < \infty$ and $\text{Supp}(\mu_i)$ compact. Then*

$$H(\gamma_\varepsilon | \mathcal{H}^{2d}) = -\frac{d}{2}\ln(2\pi\varepsilon) + H_m - \frac{d}{2} + o(1) \quad (4)$$

and

$$(c, \gamma_\varepsilon) = OT_0 + \frac{d}{2}\varepsilon + o(\varepsilon) \quad (5)$$

If we relax the finite Fisher information hypothesis the orders of magnitude still hold. The method, presented in section 3, used to prove that result relies on the Minty reparametrization trick (see definition 3.3) which implies a quadratic detachment for the duality gap function $E = c - (\varphi \oplus \psi)$, where φ, ψ are Kantorovich potentials.

Theorem (Theorem 3.7). *Suppose that the cost is quadratic, that is $c(x, y) = \frac{1}{2}\|x - y\|^2$. Further assume that μ_i have finite moment of order $2 + \delta$ then*

$$(c, \gamma_\varepsilon) = OT_0 + \Theta(\varepsilon), \quad H(\gamma_\varepsilon | \mathcal{H}^{2d}) = -\frac{d}{2}\ln(\varepsilon) + O(1), \quad \sqrt{\varepsilon} = O(W_2(\gamma_\varepsilon, \gamma_0)) \quad (6)$$

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

$$W_2(\gamma_\varepsilon, \gamma_0) = \Theta(\sqrt{\varepsilon}) \quad (7)$$

Finally for infinitesimally twisted costs we have a similar result but under the stronger assumption of compactly supported marginals. The concept of local quadratic detachment 4.1 is introduced and is key to the proof of the result which can be found in section 4.

Theorem (Theorem 4.8). *Suppose that the cost is \mathcal{C}^2 and infinitesimally twisted (see definition 4.3). Further assume that μ_i is compactly supported then*

$$(c, \gamma_\varepsilon) = OT_0 + \Theta(\varepsilon), \quad H(\gamma_\varepsilon | \mathcal{H}^{2d}) = -\frac{d}{2} \ln(\varepsilon) + O(1), \quad \sqrt{\varepsilon} = O(W_2(\gamma_\varepsilon, \gamma_0)) \quad (8)$$

1.2 Definitions and assumptions

First, let recall the definition of the *relative entropy* $H(\mu|\nu)$ and the *Fisher information* $I(\mu)$ that provide quantitative estimate of the smoothness of a measure : if $\mu, \nu \in \mathcal{P}_{ac}(\mathbb{R}^d)$, we note

$$H(\mu|\nu) := \begin{cases} \int \ln\left(\frac{d\mu}{d\nu}\right) d\mu & \text{if } \mu \ll \nu \\ +\infty & \text{otherwise} \end{cases} \quad \text{and} \quad I(\mu) = \begin{cases} \int \frac{\|\nabla \mu(x)\|^2}{\mu(x)^2} d\mu(x) & \text{if } \mu \ll \mathcal{H}^d \\ +\infty & \text{otherwise} \end{cases} \quad (9)$$

Where $\frac{d\mu}{d\nu}$ denotes the Radon-Nikodym derivative of μ with respect to ν and $\mu(x)$ the density of μ with respect to the Lebesgue measure \mathcal{H}^d by a slight abuse of notation. Note that a finite Fisher information $I(\mu) < +\infty$ implies a finite differential entropy $H(\mu|\mathcal{H}^d)$ [35, Chapter 9], that we will often denote simply by $H(\mu)$.

We now introduce the *Kantorovich potentials* which are the solutions to the dual of the classical optimal transport problem.

$$\sup_{\varphi, \psi} \int \varphi d\mu_0 + \int \psi d\mu_1 \quad (10)$$

Where the sup is taken over all functions $\varphi \in L^1(\mu_0), \psi \in L^1(\mu_1)$ such that for all $x, y \in \mathbb{R}^d \times \mathbb{R}^d$, we have $\varphi(x) + \psi(y) \leq c(x, y)$. It is well known that in our current setting this problem has maximizers which are continuous [35, Chapter 1], that is the Kantorovich potentials are continuous. Moreover it is also possible to assume that the constraint $\varphi(x) + \psi(y) \leq c(x, y)$ is binding, that is that φ and ψ are *c-transform*¹ one of each other. In the following, we will use the name *Kantorovich potentials* to denote a pair of continuous and mutual c-transform solutions to the dual problem (10).

For a pair (φ, ψ) of Kantorovich potentials, we will call *duality gap* the quantity $E(x, y) := c(x, y) - (\varphi(x) + \psi(y))$ and we will note $E = c - (\varphi \oplus \psi)$. Remark that in the quadratic case, when $\mu_i \in \mathcal{P}_{ac}(\mathbb{R}^d)$, the Kantorovich potentials are unique up to a constant and so the duality gap function E is unique. In the general case, we will be interested in the gap between $\int cd\gamma_\varepsilon$ and $\int cd\gamma_0$ which can be restated in terms of any duality gap function $E = c - (\varphi \oplus \psi)$ with φ, ψ Kantorovich potentials. Indeed we have by duality theorem that $\int cd\gamma_0 = \int \varphi d\mu_0 + \int \psi d\mu_1$ and so

$$\int cd\gamma_\varepsilon - \int cd\gamma_0 = \int cd\gamma_\varepsilon - \int (\varphi \oplus \psi) d\gamma_\varepsilon = \int Ed\gamma_\varepsilon \quad (11)$$

As explained in the main results, we will explore three different sets of assumptions, that we present, from the most specific to the most generic. Each set of assumption correspond to one of the theorems presented in the main results.

- **(H1)** The cost is quadratic : $c(x, y) := \frac{1}{2}\|x - y\|^2$.
The marginal have finite fisher information $I(\mu_i) < +\infty$ and compact support
- **(H2)** The cost is quadratic : $c(x, y) := \frac{1}{2}\|x - y\|^2$.
The marginals have finite differential entropy $H(\mu_i|\mathcal{H}^d) < +\infty$ and finite moments of order $2 + \delta$ for some $\delta > 0$

¹ $\varphi(x) = \inf_y c(x, y) - \psi(y)$ and $\psi(y) = \inf_x c(x, y) - \varphi(x)$

- **(H3)** The cost c belongs to \mathcal{C}^2 and is infinitesimally twisted (see definition 4.3).
The marginals (μ_i) have compact support and finite differential entropy.

Since a finite Fisher information implies a finite differential entropy, hypothesis **(H2)** is an important relaxation of **(H1)** on the regularity and concentration of (μ_i) . Similarly, hypothesis **(H3)** is an important relaxation of **(H1)** since the class of cost is broader and the marginals are assumed less regular. A last remark : while the finite entropy or finite information hypothesis are deeply linked with the nature of the problem and essential to our results, the concentration assumptions such as compact support or finite moment must be seen as technical and the results are likely to hold in any case.

2 Exact asymptotics from the Schrödinger problem

In this section, we assume validity of hypothesis **(H1)**, that is that the cost is quadratic and that the marginals μ_0 and μ_1 have finite Fisher information and compact support. Under finite Fisher information of the marginals, it is known that the value of (εEOT) has a second order expansion in ε (see [17, claim 4.1], see also [31] for other costs and different hypothesis)

$$OT_\varepsilon = \frac{1}{2}W_2^2 - \frac{d}{2}\varepsilon\ln(2\pi\varepsilon) + \varepsilon H_m + o(\varepsilon) \quad (\varepsilon \rightarrow 0) \quad (12)$$

In this section, we will disentangle the roles of (c, γ_ε) and $H(\gamma_\varepsilon | \mathcal{H}^{2d})$ in this asymptotic formula, and hence give a Taylor expansion of both. We will use the dynamic formulation of the entropic optimal transport.

2.1 Schrödinger problem and Benamou-Brenier formulation

The quadratic EOT problem can be reformulated as a *Schrödinger problem* (see [24])

$$\mathcal{C}_\varepsilon = \min_{\gamma \in \Pi(\mu_0, \mu_1)} H(\gamma | m_\varepsilon) \quad (13)$$

where m_ε be the measure of density $\frac{1}{(2\pi\varepsilon)^{d/2}}e^{-c/\varepsilon}$ with respect to the Lebesgue measure \mathcal{H}^{2d} . Since for $\gamma \in \Pi(\mu_0, \mu_1)$, we can compute explicitly $\varepsilon H(\gamma | m_\varepsilon) = (c, \gamma) + \varepsilon H(\gamma | \mathcal{H}^{2d}) + \frac{d}{2}\varepsilon \ln(2\pi\varepsilon)$, we have that the solutions of (13) and (εEOT) are the same and

$$OT_\varepsilon = \varepsilon \mathcal{C}_\varepsilon - \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) \quad (14)$$

Moreover there exists an analogous to Benamou-Brenier formula for \mathcal{C}_ε (see [22]) :

$$\mathcal{C}_\varepsilon = H_m + \min_{\rho, v} \frac{1}{\varepsilon} \int_0^1 \int \frac{1}{2} |v_t|^2 d\rho_t(x) dt + \frac{\varepsilon}{8} \int_0^1 \int \frac{\|\nabla \rho_t\|^2}{\rho_t} dx dt \quad (\varepsilon\text{BB})$$

Where the min is taken as in the classical Benamou-Brenier formula on paths from μ_0 to μ_1 that in the weak sense solve the continuity equation $\partial_t \rho + \text{div}(\rho v) = 0$. We will denote by $v^\varepsilon, \rho^\varepsilon$ the solution. This dynamic formulation and other variations are in fact the core of the Schrödinger problem, that intend to solve for the most probable trajectory of a free process at positive temperature with fixed initial and final marginals μ_0 and μ_1 (see [24] for a survey). The first term in (εBB) corresponds to the kinetic energy while the second term corresponds to the diffusion (ε is the temperature parameter).

2.2 A precise expansion of $H(\gamma_\varepsilon)$ and (c, γ_ε)

Minimization problems (ε EOT) and (ε BB) give the natural idea of using the envelope theorem to express the ε -derivative of OT_ε and \mathcal{C}_ε . For a minimization problem with a real parameter ε , this theorem ensures that at any differentiability point ε_0 of the value function and for any minimizer, the derivative of the value function coincide with the ε -partial derivative of the objective function with fixed minimizer (see [29]). This idea has been followed in [7] for (ε EOT) and in [12] for (ε BB). We will use the more general setting of [7] that only requires compact support for the marginals μ_i and obtain that $\varepsilon \mapsto OT_\varepsilon$ is C^∞ for $\varepsilon > 0$. From equation (14), it is direct that in that case the function $\varepsilon\mathcal{C}_\varepsilon$ is also smooth, and so we can apply envelope theorem both to $OT_\varepsilon = \min_\gamma \int \frac{1}{2} \|x-y\|^2 d\gamma + \varepsilon H(\gamma | \mathcal{H}^{2d})$ and to $\varepsilon\mathcal{C}_\varepsilon = \varepsilon H_m + \min_{\rho, v} \int_0^1 \int \frac{1}{2} |v_t|^2 d\rho_t(x) dt + \frac{\varepsilon^2}{8} \int_0^1 I(\rho_t) dt$ to get

$$\frac{d}{d\varepsilon} OT_\varepsilon = H(\gamma_\varepsilon) \quad \text{and} \quad \frac{d}{d\varepsilon} [\varepsilon\mathcal{C}_\varepsilon] = H_m + \frac{\varepsilon}{4} \int_0^1 I(\rho_t^\varepsilon) dt \quad (15)$$

Where $H(\gamma_\varepsilon)$ stands for $H(\gamma_\varepsilon | \mathcal{H}^{2d})$. But from equation (14) $\frac{d}{d\varepsilon} OT_\varepsilon = \frac{d}{d\varepsilon} [\varepsilon\mathcal{C}_\varepsilon - \frac{d}{2}\varepsilon \ln(2\pi\varepsilon)]$, so we have

$$H(\gamma_\varepsilon) = -\frac{d}{2} \ln(2\pi\varepsilon) + H_m - \frac{d}{2} + \frac{\varepsilon}{4} \int_0^1 I(\rho_t^\varepsilon) dt \quad (16)$$

This gives a quantitative relation between the entropy $H(\gamma_\varepsilon)$ of the static problem (ε EOT) and the average Fisher Information $\int_0^1 I(\rho_t^\varepsilon) dt$ of the dynamic problem (ε BB). It allows to state a complete coupling system between the different terms of (ε EOT) and (ε BB) as explained in the following proposition :

Proposition 2.1. *For the quadratic cost, suppose that the support of μ_i are compact, and that $H(\mu_i | \mathcal{H}^d) < +\infty$. Then*

$$\begin{cases} (c, \gamma_\varepsilon) &= \int_0^1 \int \frac{1}{2} |v_t^\varepsilon|^2 d\rho_t(x) dt - \frac{\varepsilon^2}{8} \int_0^1 I(\rho^\varepsilon) dt + \frac{d}{2}\varepsilon \\ H(\gamma_\varepsilon) &= \frac{\varepsilon}{4} \int_0^1 I(\rho^\varepsilon) dt - \frac{d}{2} \ln(2\pi\varepsilon) + H_m - \frac{d}{2} \end{cases} \quad (17)$$

Proof. The second line of the system (17) is nothing but equation (16). For the first line, equation (14) implies that

$$(c, \gamma_\varepsilon) = \varepsilon\mathcal{C}_\varepsilon - \frac{d}{2}\varepsilon \ln(2\pi\varepsilon) - \varepsilon H(\gamma_\varepsilon) \quad (18)$$

We can inject the expression of $H(\gamma_\varepsilon)$ given by (16) and the expression of \mathcal{C}_ε with the minimal continuous path $(\rho^\varepsilon, v^\varepsilon)$ of equation (ε BB) to obtain the first line of the system (17). \square

Thanks to the system (17), it is enough to study the asymptotics of $\varepsilon \int_0^1 I(\rho^\varepsilon) dt$ and $\int_0^1 \int |v_t^\varepsilon|^2 d\rho_t(x) dt$ to get results on (c, γ_ε) and $H(\gamma_\varepsilon)$. When the Fisher information of the marginal is finite, it turns out that the behaviour of $\int_0^1 I(\rho^\varepsilon) dt$ and $\int_0^1 \int |v_t^\varepsilon|^2 d\rho_t(x) dt$ can be described quite precisely :

Proposition 2.2. *Suppose that the cost is quadratic. If μ_0 and μ_1 have finite Fisher information, when ε tends to 0 :*

$$\varepsilon \int_0^1 I(\rho^\varepsilon) dt \rightarrow 0 \quad \text{and} \quad \int_0^1 \int |v_t^\varepsilon|^2 d\rho_t(x) dt = W_2^2 + o(\varepsilon) \quad (19)$$

Proof. Since the Fisher information of the marginal is finite, from [17, Claim 4.1] the Taylor expansion (12) of OT_ε holds. But since $OT_\varepsilon = \varepsilon\mathcal{C}_\varepsilon - \frac{d}{2}\varepsilon \ln(2\pi\varepsilon)$, expansion (12) is equivalent to the following :

$$\mathcal{C}_\varepsilon - \frac{1}{2\varepsilon} W_2^2 - H_m \rightarrow 0$$

However from the expression of C_ε with the minimal continuous path $(\rho^\varepsilon, v^\varepsilon)$ of equation (εBB) ,

$$C_\varepsilon - \frac{1}{2\varepsilon}W_2^2 - H_m = \frac{1}{\varepsilon} \left(\int_0^1 \int \frac{1}{2}|v_t^\varepsilon|^2 d\rho_t(x) dt - \frac{1}{2}W_2^2 \right) + \frac{\varepsilon}{8} \int_0^1 \int \frac{\|\nabla \rho_t^\varepsilon\|^2}{\rho_t^\varepsilon} dx dt$$

Both terms $\frac{1}{\varepsilon} \left[\int_0^1 \int \frac{1}{2}|v_t^\varepsilon|^2 d\rho_t(x) dt - \frac{1}{2}W_2^2(\mu_0, \mu_1) \right]$ and $\frac{\varepsilon}{8} \int_0^1 \int \frac{\|\nabla \rho_t^\varepsilon\|^2}{\rho_t^\varepsilon} dx dt$ are positive, the first one because $(\rho^\varepsilon, v^\varepsilon)$ is a path from μ_0 to μ_1 solving the continuity equation, and so the $(\varepsilon = 0)$ Benamou-Brenier formula holds. Since the sum tends to zero, we know that both terms tend to zero which is what we needed to prove. \square

Proposition 2.2 can be seen as a refinement of [20, Lemma 3.3] where, among other asymptotics, it is shown that $\varepsilon^2 \int_0^1 I(\rho^\varepsilon) dt \rightarrow 0$ and $\int_0^1 \int |v_t^\varepsilon|^2 d\rho_t(x) dt \rightarrow W_2^2$. We can now combine propositions 2.1 and 2.2 to obtain a Taylor expansion of (c, γ_ε) and $H(\gamma_\varepsilon)$:

Theorem 2.3. *Suppose that the cost is quadratic, let $\mu_0, \mu_1 \in \mathcal{P}_{ac}(\mathbb{R}^d)$ and suppose that for $i = 1, 2$, we have $I(\mu_i) < +\infty$ and the support of μ_i is compact. Then*

$$H(\gamma_\varepsilon) = -\frac{d}{2} \ln(2\pi\varepsilon) + H_m - \frac{d}{2} + o(1) \quad (20)$$

and

$$(c, \gamma_\varepsilon) = \frac{1}{2}W_2^2 + \frac{d}{2}\varepsilon + o(\varepsilon) \quad (21)$$

Proof. Since a finite Fisher information implies a finite differential entropy, the hypothesis of propositions 2.1 and 2.2 hold and we can inject the asymptotics on $\int_0^1 I(\rho^\varepsilon) dt$ and $\int_0^1 \int |v_t^\varepsilon|^2 d\rho_t(x) dt$ (19) obtained in the latter in the system (17) of the former. \square

Remark 2.4.

The rate in $\frac{d}{2}\varepsilon$ for the suboptimality $(c, \gamma_\varepsilon) - (c, \gamma_0)$ had been already observed in the case of gaussian measures in 1D (see introductions of [1],[4]). It is astonishing to remark that this first order term does not depend on μ_0 and μ_1 . Hence, the estimator $(c, \gamma_\varepsilon) - \frac{d}{2}\varepsilon$ could provide an interesting estimate of $W_2^2(\mu_0, \mu_1)$ with a lack of precision comparable to the one of the Sinkhorn divergences $OT_\varepsilon(\mu_0, \mu_1) - \frac{1}{2}[OT_\varepsilon(\mu_0, \mu_0) + OT_\varepsilon(\mu_1, \mu_1)]$ (see [18])

2.3 Rephrasing the results with different functionals

The asymptotic formula for the entropy (20) and the formula for (c, γ_ε) in the system (17) suggest to introduce two quantities that are linked to our problem. First, the *Entropy power function* $N_d(\mu) := \frac{e^{-\frac{2}{d}H(\mu|\mathcal{H}^d)}}{2\pi e}$. This functional has various links with our problem : Costa's theorem states that it is concave along the heat flow (see [13]), $\sqrt{N_{2d}}$ is concave along Wasserstein geodesics (see [16]) and along Schrödinger bridges (see [32]). And it turns out that the asymptotic (20) can be rewritten in the simple way²

$$N_{2d}(\gamma_\varepsilon) \sim N_d(\mu_0)^{1/4} N_d(\mu_1)^{1/4} \varepsilon^{1/2}$$

The other quantity that we need to present is the *energy*³

$$\mathcal{E}_\varepsilon := \frac{1}{\varepsilon^2} \int \frac{1}{2}|v_t^\varepsilon|^2 d\rho_t(x) - \frac{1}{8}I(\rho_t^\varepsilon)$$

²Remark that since the functional N is multiplicative, the right hand side is the $4d$ entropy power of $\mu_0 \otimes \mu_1 \otimes \mathcal{N}(0, \frac{\varepsilon}{2d}I_{2d})$

³see the introduction of [12] for a discussion on the name energy

As the notation let think and as it is proved in [11, Corollary 1.1] and [20, lemma 3.3], this energy does not depend on the time t . With this quantity, the expression of (c, γ_ε) in the system (17) becomes $(c, \gamma_\varepsilon) = \varepsilon^2 \mathcal{E}_\varepsilon + \frac{d}{2} \varepsilon$, and so the Taylor expansion (21) expresses

$$\varepsilon^2 \mathcal{E}_\varepsilon = \frac{1}{2} W_2^2 + o(\varepsilon)$$

Note that it was already known that $\varepsilon^2 \mathcal{E}_\varepsilon \rightarrow \frac{1}{2} W_2^2$ (see the introduction of [12] or the proof of [20, lemma 3.3]). The convergence in $o(\varepsilon)$ seems new to us. Given the expression for $\frac{d}{d\varepsilon}[\varepsilon \mathcal{C}_\varepsilon]$ of equation (15), the other Taylor expansion (20) corresponding to $\varepsilon \int_0^1 I(\rho^\varepsilon) dt \rightarrow 0$ is also equivalent to the fact that $\varepsilon \mathcal{C}_\varepsilon$ is C^1 in 0 with derivative H_m . It can be seen as an extension in 0 of [11, Theorem 1.1].

3 Rates for the quadratic cost

In the last section, under the strong assumptions **(H1)** of smoothness ($I(\mu_i) < +\infty$) and concentration ($Supp(\mu_i)$ compact), the Taylor expansions of (εEOT) and (εBB) allowed us to give precise rates for the entropy as well as the cost term in (εEOT) . The main idea was to disentangle the role of the entropy and the cost. In this section we choose the weaker set of assumptions **(H2)**. The cost is still the quadratic one : $c(x, y) = \frac{1}{2} \|x - y\|^2$, but the Fisher information need not to be finite and so the second order Taylor expansion (12) of (εEOT) need not to hold. However, the entropy $H(\mu_i)$ is still supposed finite. Thanks to the technical assumption that the moments of order $2 + \delta$ of the marginals are finite for some $\delta > 0$, we still have (see lemma 3.6) some rates of convergence for the value of (εEOT) of the form

$$0 \leq OT_\varepsilon - OT_0 \leq -\frac{d}{2} \varepsilon \ln(\varepsilon) + O(\varepsilon) \quad (22)$$

From that, it is still possible to disentangle the cost and the entropy (theorem 3.7) through a careful study of the behaviour of $H(\gamma_\varepsilon)$ as $\varepsilon \rightarrow 0$. It will grant rates for $(c, \gamma_\varepsilon) - (c, \gamma_0)$ of the same order ε as in last section. We start by providing a sketch of the proof :

Let E be a duality gap function $E = c - (\varphi \oplus \psi)$. Recall that since γ_ε has μ_i as marginals, we have that $\int E d\gamma_\varepsilon = (c, \gamma_\varepsilon) - (c, \gamma_0)$, and so we will denote by (E, γ_ε) this suboptimality. The goal is to prove $(E, \gamma_\varepsilon) \simeq \varepsilon$ and $H(\gamma_\varepsilon) \simeq -\frac{d}{2} \ln(\varepsilon)$, so it is natural to try to prove first $H(\gamma_\varepsilon) \simeq -\frac{d}{2} \ln(E, \gamma_\varepsilon)$, or in terms of entropy power, $N_{2d}(\gamma_\varepsilon) \simeq \sqrt{(E, \gamma_\varepsilon)}$. But thanks to the information encoded in inequality (22), it is enough to show an inequality of the type $N_{2d}(\gamma_\varepsilon) \leq C \sqrt{(E, \gamma_\varepsilon)}$.

In fact, as explained in detail in section 3.1, this kind of inequality is not specific to γ_ε and it comes from the following fact : the contact set $\Sigma := \{x, y / E(x, y) = 0\}$ has d dimensions less than the ambient space \mathbb{R}^{2d} and (E, γ) quantifies the average distance of γ to Σ because E has a *quadratic detachment* in Minty's coordinates (see definitions 3.1 and 3.3). As a matter of fact, $W_2^2(\gamma, \gamma_0)$ also quantifies this distance, because γ_0 is supported on Σ . In proposition 3.5 we obtain results of the form

$$N_{2d}(\gamma) \leq C \sqrt{(E, \gamma)} \quad \text{and} \quad N_{2d}(\gamma) \leq C W_2(\gamma, \gamma_0) \quad (23)$$

or equivalently in term of entropy,

$$H(\gamma) \geq -\frac{d}{2} \ln(E, \gamma) + C \quad \text{and} \quad H(\gamma) \geq -\frac{d}{2} \ln(W_2^2(\gamma, \gamma_0)) + C \quad (24)$$

These inequalities are the key for all our rates of convergence of (E, γ_ε) , $W_2^2(\gamma_\varepsilon, \gamma_0)$ and $H(\gamma_\varepsilon)$ detailed in section 3.2.

3.1 Lower bounds on the entropy

The results of this section are completely independent of the optimality of γ_ε , and even independent of the fact that γ_ε transports μ_0 on μ_1 . Hence in the following, we will replace γ_ε by a generic

$\gamma \in \mathcal{P}_{ac}(\mathbb{R}^{2d})$. The results presented in this section are lower bounds on its entropy $H(\gamma)$. These inequalities rely on the comparison of γ with a Gaussian along directions transverse to the support of a generic optimal plan γ_0 for $\varepsilon = 0$. Indeed we know that any optimal plan γ_0 is supported on $\Sigma = \{E = 0\}$ which is approximatively a submanifold of (co)dimension d . We will then show that $\int Ed\gamma$ as well as $W_2^2(\gamma, \gamma_0)$ control the variance of γ along transverse directions to Σ . Finally since under variance constraint the Gaussian has the smallest entropy we will conclude with the wanted lower bounds of equations (23) and (24). The same strategy will work in the case of twisted costs. However in the quadratic case the proof requires only a global change of variable (Minty's trick) which makes the arguments clearer. We start with the fundamental property of a function that allows for estimates like (23) or (24), the *quadratic detachment*

Definition 3.1 (Quadratic detachment). Let $G : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$. We say that G has a *quadratic detachment* if for any $u \in \mathbb{R}^d$ it exists $v_u \in \mathbb{R}^d$ such that

$$\forall (u, v) \in \mathbb{R}^{2d} \quad G(u, v) \geq \frac{1}{2} \|v - v_u\|^2 \quad (25)$$

This quadratic detachment property allows for a bound on the differential entropy $H(\gamma) := H(\gamma|\mathbb{R}^{2d})$ of any plan $\gamma \in \mathcal{P}_{ac}(\mathbb{R}^{2d})$:

Proposition 3.2. Let G be a function on $\mathbb{R}^d \times \mathbb{R}^d$ and $\gamma \in \mathcal{P}_{ac}(\mathbb{R}^{2d})$ and denote $C_d := -\frac{d}{2} \ln(\frac{4\pi e}{d})$. If G has a quadratic detachment

$$H(\gamma) \geq -\frac{d}{2} \ln \left(\int Gd\gamma \right) + H(\gamma_1) + C_d \quad (26)$$

Where γ_1 is the projection of γ on the first coordinate, ie the first marginal of γ , and where H is the differential entropy ($H(\gamma) = H(\gamma|\mathbb{R}^{2d})$ and $H(\gamma_1) = H(\gamma_1|\mathbb{R}^d)$)

Proof. If $H(\gamma) = +\infty$ or $\int Gd\gamma = +\infty$ there is nothing to prove. We assume that $H(\gamma) < +\infty$ and $\int Gd\gamma < +\infty$. Let $\gamma = \gamma_1 \otimes \gamma^u$ be the disintegration of γ with respect to the first coordinate. $\int Gd\gamma < +\infty$ implies that γ^u has a finite moment of order 2, γ_1 almost everywhere, because G has a quadratic detachment. By additivity of entropy we have

$$H(\gamma) = H(\gamma_1) + \int H(\gamma^u) d\gamma_1(u) \quad (27)$$

However under variance constraint the gaussian with independent coordinates has the smallest entropy thus $H(\gamma^u) \geq H(\mathcal{N}(0, \frac{\text{Var}(\gamma^u)}{d} I_d)) = -\frac{d}{2} \ln(\text{Var}(\gamma^u)) - \frac{d}{2} \ln(\frac{2\pi e}{d})$. Moreover, the average of a random variable minimizes the average square distance, so $\text{Var}(\gamma^u) \leq \int \|v - v_u\|^2 d\gamma_u(v) \leq 2 \int G(u, v) d\gamma_u(v)$ where the last inequality holds because G has a quadratic detachment. So we have

$$H(\gamma) \geq H(\gamma_1) - \int \frac{d}{2} \ln \left(2 \int G(u, v) d\gamma_u(v) \right) d\gamma_1(u) - \frac{d}{2} \ln(\frac{2\pi e}{d}) \quad (28)$$

Now by concavity of the logarithm, exchanging it with the integral against γ_1 gives an even lower quantity, hence we obtain the desired inequality. \square

This result shows a clear path to get lower bounds on the entropy of γ_ε : typically, it is enough to show that the duality gap has a quadratic detachment to show inequality (24). However, E has no quadratic detachment in the classic (x, y) coordinates, but in transverse coordinates, that we call *Minty's coordinates* ([30])

Definition 3.3 (Minty's coordinates). We will call *Minty's coordinates* the coordinates $(u, v) = \frac{1}{\sqrt{2}}(x + y, x - y)$. The change of variable $(x, y) \mapsto (u, v)$ is an isometry and for any convex function f , the graph of the subdifferential of f in the (x, y) -coordinates corresponds to the graph of a 1-Lipschitz function defined on the whole space in the (u, v) -coordinates (see [33, theorem 12.12, theorem 12.25]). Moreover, for $\gamma \in \mathcal{P}(\mathbb{R}^{2d})$, we will denote by $\hat{\gamma}$ the pushforward of γ through this change of variable and call *Minty's decomposition* of γ the disintegration of $\hat{\gamma}$ with respect to the orthogonal projection onto the first coordinate $\hat{\gamma} = \hat{\mu} \otimes \hat{\gamma}^u$.

The quadratic detachment of E in those coordinates follows from the slightly more general inequality (30), known as Minty's trick. It can be generalized for twisted costs (see lemma 4.4). In the quadratic case, it is somehow folklore and we recall the proof, but for more general cost, it was proved in [7], following arguments of [28]. We also prove the quadratic detachment of the function $d(\cdot, \Sigma)^2$ that we will use to bound the entropy with the Wasserstein distance to an optimizer.

Lemma 3.4. *Let $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ with $E := c - (\varphi \oplus \psi)$ the duality gap of the quadratic optimal transport problem from μ_0 to μ_1 . Denote by $d(z, \Sigma)$ the distance of a point $z \in \mathbb{R}^{2d}$ to the set $\Sigma = \{E = 0\}$. In Minty's coordinates, the functions $(u, v) \mapsto E(u, v)$ and $(u, v) \mapsto d((u, v), \Sigma)^2$ have quadratic detachment.⁴*

Proof. First note that in the quadratic case E rewrites as $E(x, y) = f(x) + f^*(y) - \langle x, y \rangle$ where f is a convex function and f^* is its Legendre-Fenchel conjugate. Thus for $(x, y), (x', y') \in \mathbb{R}^d \times \mathbb{R}^d$ we have

$$\begin{aligned} E(x, y) + E(x', y') &= f(x) + f^*(y) - \langle x, y \rangle + f(x') + f^*(y') - \langle x', y' \rangle \\ &\geq \langle x, y' \rangle - f^*(y') + f^*(y) - \langle x, y \rangle + \langle x', y \rangle - f(y) + f^*(y') - \langle x', y' \rangle \\ &\geq \langle x - x', y' - y \rangle \end{aligned} \quad (29)$$

In Minty's coordinates $(u, v) = \frac{1}{\sqrt{2}}(x + y, x - y)$ we can see E as a function of (u, v) . In these coordinates, inequality (29) becomes

$$E(u, v) + E(u', v') \geq \frac{1}{2}(\|v' - v\|^2 - \|u' - u\|^2) \quad (30)$$

The set $\Sigma = \{x, y / E(x, y) = 0\}$ corresponds to the subdifferential of the convex function f . Hence in the Minty's coordinates (u, v) , it corresponds to the graph in u of a 1-lipschitz function that is defined on the whole space \mathbb{R}^d (see [33, theorem 12.12, theorem 12.25]). So for any $u \in \mathbb{R}^d$, it exists a unique $v_u \in \mathbb{R}^d$ such that $E(u, v_u) = 0$. And thanks to Minty's trick (30) we have that $E(u, v) \geq \frac{1}{2}\|v - v_u\|^2$.

It remains to show that the distance to the graph of the 1-lipschitz function $u \mapsto v_u$ has a quadratic detachment. Let $u, u', v \in \mathbb{R}^d$. From the inequality $\|a\|^2 + \|b\|^2 \geq \frac{1}{2}\|a + b\|^2$ applied to $a = v_u - v_{u'}$ and $b = v_{u'} - v$, we get

$$\|v_u - v_{u'}\|^2 + \|v_{u'} - v\|^2 \geq \frac{1}{2}\|v_u - v\|^2$$

And we can bound $\|v_u - v_{u'}\|^2$ by $\|u - u'\|^2$ because $u \mapsto v_u$ is 1-lipschitz. Hence we obtain

$$\|u - u'\|^2 + \|v_{u'} - v\|^2 \geq \frac{1}{2}\|v_u - v\|^2$$

But the left hand side is the distance from (u, v) to $(u', v_{u'})$ and the inequality holds for any u' , so

$$d((u, v), \Sigma) \geq \frac{1}{2}\|v_u - v\|^2$$

□

Now that we know that E and $d(\cdot, \Sigma)^2$ have quadratic detachment, we can use proposition 3.2 to bound the entropy of γ from below with $\int E d\gamma$ and $W_2^2(\gamma, \gamma_0)$.

Proposition 3.5. *Suppose that the cost is quadratic. Let $(\mu_0, \mu_1) \in \mathcal{P}_2(\mathbb{R}^d)$, (φ, ψ) be associated Kantorovich potentials and $E := c - (\varphi \oplus \psi)$ be a duality gap function. Let γ_0 be an optimal transport plan from μ_0 to μ_1 . Then, for any $\gamma \in \mathcal{P}_{ac}(\mathbb{R}^{2d})$*

$$H(\gamma) \geq -\frac{d}{2} \ln \left(\int E d\gamma \right) + H(\hat{\mu}) + C_d \quad (31)$$

⁴here, by a slight abuse of notation, E denotes $E \circ \Phi^{-1}$ where $\Phi : (x, y) \mapsto (u, v)$ is Minty's change of variable (definition 3.3)

$$H(\gamma) \geq -\frac{d}{2} \ln W_2^2(\gamma, \gamma_0) + H(\hat{\mu}) + C_d \quad (32)$$

where $\hat{\gamma} = \hat{\mu} \otimes \hat{\gamma}^u$ is the Minty's decomposition of γ and $C_d := -\frac{d}{2} \ln(\frac{4\pi e}{d})$ (see definition 3.3). If we denote by $\sigma_\gamma(X+Y)$ the standard deviation of $X+Y$ when the law of (X, Y) is γ , we have

$$N_{2d}(\gamma) \leq \frac{\sigma_\gamma(X+Y)}{d} \sqrt{\int \int E d\gamma} \quad N_{2d}(\gamma) \leq \frac{\sigma_\gamma(X+Y)}{d} W_2(\gamma, \gamma_0) \quad (33)$$

Proof. By lemma 3.4, the duality gap E and the square distance $d(\cdot, \Sigma)^2$ have quadratic detachment. So we can apply proposition 3.2 on both to obtain

$$H(\gamma) \geq -\frac{d}{2} \ln \left(\int \int E d\gamma \right) + H(\hat{\mu}) + C_d \quad (34)$$

$$H(\gamma) \geq -\frac{d}{2} \ln \left(\int d(\mathbf{x}, \Sigma)^2 d\gamma(\mathbf{x}) \right) + H(\hat{\mu}) + C_d \quad (35)$$

The first inequality is exactly (31) and the second implies (32) because an optimal plan γ_0 is concentrated on a contact set $\Sigma = \{x, y/E(x, y) = 0\}$ by optimality and so $W_2^2(\gamma, \gamma_0) \geq \int d(\mathbf{x}, \Sigma)^2 d\gamma(\mathbf{x})$. The second part is a consequence and we provide only the proof for $\int \int E d\gamma$ because the proof for $W_2^2(\gamma_\varepsilon, \gamma_0)$ is stricly similar. Taking the exponential of (31), we get

$$N_{2d}(\gamma)^2 \leq \frac{2}{d} N_d(\hat{\mu}) \int \int E d\gamma$$

Since the gaussian minimizes the entropy at fixed variance, the entropy power of $\hat{\mu}$ is bigger than the one of a gaussian of same variance, that is $N_d(\hat{\mu}) \leq \frac{\text{Var}(\hat{\mu})}{d}$. Moreover $\text{Var}(\hat{\mu}) = \frac{1}{2} \text{Var}_\gamma(X+Y)$ because $\hat{\mu}$ is the law of the first marginal of γ in the coordinates $(u, v) = \frac{1}{\sqrt{2}}(x+y, x-y)$. This is enough to finish the proof of (33). \square

We have established the bounds (33) on a general plan γ . It is now time to apply these results to the entropic plan γ_ε under hypothesis **(H2)** in order to obtain rates of convergence of the suboptimality $(c, \gamma_\varepsilon) - (c, \gamma_0)$. We will also quantify the weak convergence of γ_ε towards γ_0 in W_2 distance.

3.2 Rates of convergence

In the sequel of the section, we use the set of assumptions **(H2)** that imply in particular that the optimizer γ_0 is unique. The main result of this section states that the suboptimality $(c, \gamma_\varepsilon) - (c, \gamma_0)$ converges to 0 at a speed of order ε . We also derive the same speed for $W_2^2(\gamma_\varepsilon, \gamma_0)$ and this matches in both cases the rate found for gaussians.

The proofs of the theorems require an upperbound on the rate of convergence of OT_ε towards OT_0 . The key point being that the convergence rate is of the form

$$\varepsilon H(\gamma_\varepsilon) \leq OT_\varepsilon - OT_0 \leq -\frac{d}{2} \varepsilon \ln(\varepsilon) + O(\varepsilon) \quad (36)$$

This result has been obtained recently for general costs under some regularity conditions [7] and in the quadratic case under moments conditions [15]. The following lemma explicits the dependence of the bounded term $O(\varepsilon)$ found in the latter.

Lemma 3.6. *Assume as in hypothesis **(H2)** that the cost is quadratic, that $\mu_0, \mu_1 \in \mathcal{P}_{2+\delta}(\mathbb{R}^d)$ for some $\delta > 0$ and that $H(\mu_i) < +\infty$. Then, as $\varepsilon \rightarrow 0$*

$$OT_\varepsilon - OT_0 \leq -\frac{d}{2} \varepsilon \ln(\varepsilon) + C\varepsilon \quad (37)$$

for a constant C depending on $d, m_{2+\delta}(\mu_i)$ and $H(\mu_i)$.

Proof. The constant C introduced in [15, corollary 3.12] is (up to a numerical constant) the quantization constant of the optimal transport γ_0 . It allows for the following upper bound

$$OT_\varepsilon - OT_0 \leq -\frac{d}{2}\varepsilon \ln(\varepsilon) + (H(\mu_0) + H(\mu_1))\varepsilon + C\varepsilon \quad (38)$$

However by Minty's trick [15, lemma 3.11] this quantization constant is tied to the quantization constant of a measure admitting a moment of order $2 + \delta$. In that case [23, corollary 6.7] gives an explicit bound

$$C \leq C' (m_{2+\delta}(\gamma_0) + d) \quad (39)$$

where C' is independent of dimension. We conclude by using triangular inequality to upper bound $m_{2+\delta}(\gamma_0)$ by $2^{2+\delta}(m_{2+\delta}(\mu_0) + m_{2+\delta}(\mu_1))$. \square

With the inequality (37) that gives an upper bound on $H(\gamma_\varepsilon)$ and the lower bound (31) found in the previous section, we are now able to show that $(c, \gamma_\varepsilon) - (c, \gamma_0) = \Theta(\varepsilon)$:

Theorem 3.7. *Suppose that the cost is quadratic. Let $\mu_0, \mu_1 \in \mathcal{P}_{2+\delta, ac}(\mathbb{R}^d)$ for some $\delta > 0$. And further assume that their entropy is finite. Then in the limit $\varepsilon \rightarrow 0$,*

$$(c, \gamma_\varepsilon) - (c, \gamma_0) = \Theta(\varepsilon), \quad H(\gamma_\varepsilon) = -\frac{d}{2}\ln(\varepsilon) + O(1) \quad \text{and} \quad OT_\varepsilon = OT_0 - \frac{d\varepsilon}{2}\ln(\varepsilon) + O(\varepsilon) \quad (40)$$

More precisely, there exist $c, C > 0$, depending on $m_{2+\delta}(\mu_i)$, $H(\mu_i)$ and the dimension d , such that

$$c\varepsilon \leq \int E d\gamma_\varepsilon = (c, \gamma_\varepsilon) - (c, \gamma_0) \leq C\varepsilon \quad (41)$$

Proof. Set $(E, \gamma_\varepsilon) := \int E d\gamma_\varepsilon$, and consider d , $H(\mu_i)$ and $m_{2+\delta}(\mu_i)$ as fixed. Lemma 3.6 grants the following upper bound on the regularized problem for ε small enough

$$(E, \gamma_\varepsilon) + \varepsilon H(\gamma_\varepsilon) \leq -\frac{d}{2}\varepsilon \ln(\varepsilon) + C\varepsilon \quad (42)$$

On the other hand, from proposition 3.5, we know that $N_{2d}(\gamma_\varepsilon) \leq \frac{\sigma_{\gamma_\varepsilon}(X+Y)}{d} \sqrt{(E, \gamma_\varepsilon)}$. But we have also $\sigma_{\gamma_\varepsilon}(X+Y) \leq \sigma_{\gamma_\varepsilon}(X) + \sigma_{\gamma_\varepsilon}(Y) = \sqrt{m_2(\mu_0)} + \sqrt{m_2(\mu_1)}$ because the marginals of γ_ε are μ_0 and μ_1 . And by Hölder inequality, $m_2(\mu) \leq m_{2+\delta}(\mu)^{\frac{2}{2+\delta}}$. So $N_{2d}(\gamma_\varepsilon) \leq C' \sqrt{(E, \gamma_\varepsilon)}$ or equivalently, by taking the logarithm,

$$H(\gamma_\varepsilon) \geq -\frac{d}{2}\ln(E, \gamma_\varepsilon) + C'' \quad (43)$$

Now by combining (43) with (42) we have

$$\frac{(E, \gamma_\varepsilon)}{\varepsilon} - \frac{d}{2}\ln\left(\frac{(E, \gamma_\varepsilon)}{\varepsilon}\right) \leq C''' \quad (44)$$

The function $x \mapsto x - d/2 \ln(x)$ goes to $+\infty$ near 0 and $+\infty$. Thus $\frac{(E, \gamma_\varepsilon)}{\varepsilon} = \Theta(1)$, or more precisely, there exist constants $c, C > 0$, depending on d , $m_{2+\delta}(\mu_i)$ and $H(\mu_i)$, such that $c\varepsilon \leq (E, \gamma_\varepsilon) \leq C\varepsilon$. Injecting $(E, \gamma_\varepsilon) = \Theta(\varepsilon)$ in inequalities (42) and (43) gives respectively $H(\gamma_\varepsilon) \leq -\frac{d}{2}\ln(\varepsilon) + O(1)$ and $H(\gamma_\varepsilon) \geq -\frac{d}{2}\ln(\varepsilon) + O(1)$ which concludes the proof. \square

Now that we have proven that the suboptimality $(c, \gamma_\varepsilon) - (c, \gamma_0)$ has speed ε , we want to do the same for the the Wasserstein distance $W_2^2(\gamma_\varepsilon, \gamma_0)$. Inequality (32) will give a lower bound on $W_2^2(\gamma_\varepsilon, \gamma_0)$, but we need also an upper bound. If the Brenier map is Lipschitz, the following lemma provides it.⁵

⁵This lemma and its proof was suggested to us by P. Pegon

Lemma 3.8. *For the quadratic cost, suppose that the Brenier map $T = \nabla f$ is L Lipschitz. Then*

$$W_2^2(\gamma_\varepsilon, \gamma_0) \leq \int \|y - T(x)\|^2 d\gamma_\varepsilon \leq 2L((c, \gamma_\varepsilon) - (c, \gamma_0)) \quad (45)$$

Proof. Let $S(x, y) = (x, T(x))$ with $T = \nabla f$ and f the Brenier convex function associated with γ_0 . Then S is a transport map from γ_ε to γ_0 thus $W_2^2(\gamma_\varepsilon, \gamma_0) \leq \int \|y - T(x)\|^2 d\gamma_\varepsilon$. We now use an inequality proved by Berman in [3] and Li and Nchetto in [25], who have built upon an earlier work of Gigli [21]. The inequality is that whenever the Brenier map is Lipschitz we have $\|y - T(x)\|^2 \leq 2LE$ with $E = c - (\varphi \oplus \psi)$ the duality gap. Combining these two inequalities grants the result, since $\int Ed\gamma_\varepsilon = (c, \gamma_\varepsilon) - (c, \gamma_0)$. \square

The hypothesis that the Brenier map is Lipschitz might seem abstract. However, note that Caffarelli's regularity theory ([5],[6]) ensures it holds under some simple regularity conditions on the marginals. For example as soon as the marginals have Hölder densities bounded away from zero on their supports and the latter are smooth and uniformly convex. In a different direction, recent works [8] ensure regularity of the transport map between log-concave measures under a variance constraint.

Theorem 3.9. *Suppose that the cost is quadratic. Let $\mu_0, \mu_1 \in \mathcal{P}_{2+\delta, ac}(\mathbb{R}^d)$ for some $\delta > 0$. And further assume that their entropy is finite and that the Brenier map $T = \nabla f$ is Lipschitz. Then $W_2^2(\gamma_\varepsilon, \gamma_0) = \Theta(\varepsilon)$. More precisely :*

1. *if $\mu_0, \mu_1 \in \mathcal{P}_{2+\delta, ac}(\mathbb{R}^d)$ with finite entropy, then there exist $c > 0$, depending on the moments $m_{2+\delta}(\mu_i)$, the entropies $H(\mu_i)$ and the dimension d , such that, as $\varepsilon \rightarrow 0$*

$$c\varepsilon \leq W_2^2(\gamma_\varepsilon, \gamma_0) \quad (46)$$

2. *if moreover the Brenier map is L -lipschitz, then there exist $C > 0$ depending on $d, m_{2+\delta}(\mu_i), H(\mu_i)$ and L , such that, as $\varepsilon \rightarrow 0$*

$$c\varepsilon \leq W_2^2(\gamma_\varepsilon, \gamma_0) \leq C\varepsilon \quad (47)$$

Proof. For the first part, let's recall that from proposition 3.5, we have $N_{2d}(\gamma_\varepsilon) \leq \frac{\sigma_{\gamma_\varepsilon}(X+Y)}{d} \sqrt{(E, \gamma_\varepsilon)}$. As in the proof of theorem 3.7, we bound can bound $\sigma_{\gamma_\varepsilon}(X+Y)$ with the moments $m_{2+\delta}(\mu_i)$ and so we obtain

$$N_{2d}(\gamma_\varepsilon) \leq C(m_{2+\delta}(\mu_i), d)W_2(\gamma_\varepsilon, \gamma_0) \quad (48)$$

But on the other hand, from lemma 3.6, we now that $H(\gamma_\varepsilon) \leq -\frac{d}{2} \ln(\varepsilon) + C(m_{2+\delta}(\mu_i), H(\mu_i), d)$. Or equivalently,

$$N_{2d}(\gamma_\varepsilon) \geq C(m_{2+\delta}(\mu_i), H(\mu_i), d)\sqrt{\varepsilon} \quad (49)$$

So, combining inequality (48) and (49), we get $W_2^2(\gamma_\varepsilon, \gamma_0) \geq c\varepsilon$ for some constant c depending on $d, H(\mu_i)$ and $m_{2+\delta}(\mu_i)$.

The second part is a direct combination of lemma 3.8 from which $W_2^2(\gamma_\varepsilon, \gamma_0) \leq 2L((c, \gamma_\varepsilon) - (c, \gamma_0))$ and theorem 3.7 which states that $(c, \gamma_\varepsilon) - (c, \gamma_0) \leq C(m_{2+\delta}(\mu_i), H(\mu_i), d)\varepsilon$. \square

Remark 3.10. *Note that, as explained in [7, Proposition 4.5], the method of lemma 3.8 allows to control also the L^2 -gap between T_ε , the barycentric projection of γ_ε , and the Brenier map $T = \nabla f$.*

$$\|T_\varepsilon - T\|_{L^2(\mu)}^2 \leq \int |y - T(x)|^2 d\gamma_\varepsilon(x, y) \leq 2L \int Ed\gamma_\varepsilon \leq C\varepsilon$$

In a similar fashion we get a control over the distance between the support of γ_ε and the graph of the optimal transport

$$\int d(y, \text{supp}(\gamma_0))^2 d\gamma_\varepsilon \leq \int |y - T(x)|^2 d\gamma_\varepsilon(x, y) = O(\varepsilon) \quad (50)$$

Hence, for the quadratic case, under technical assumptions, we have obtain generic rates of convergence when the marginals μ_i have finite differential entropy :

$$(c, \gamma_\varepsilon) = (c, \gamma_0) + \Theta(\varepsilon) \quad H(\gamma_\varepsilon) = -\frac{d}{2} \ln(\varepsilon) + O(1) \quad W_2(\gamma_\varepsilon, \gamma_0) = \Theta(\sqrt{\varepsilon}) \quad (51)$$

Can we generalize it to a broader class of costs ? This is the objective of next section.

4 Rates for infinitesimally twisted costs

In this section, we use the set of assumption **(H3)** that is adapted to twisted cost functions. As in the quadratic case, we begin by estimates on the entropy based on quadratic detachment.

4.1 Lower bound on the entropy

The key ingredient used to lower bound the entropy was the lower bound involving the variance $H(\gamma) \geq H(\mathcal{N}(0, \text{Var}(\gamma)/d))$. In fact we used a finer result using a disintegration of γ in two components, one of which had bounded variance. In order to extend the preceding results to more general costs our goal is to have a *local* version of the proposition 3.5 that states a precise lower bound for $H(\gamma)$. Indeed, imagine that $E \geq 0$ is a function such that there is a direction along which E grows quadratically locally around $\Sigma = \{E = 0\}$. For any measure γ , it is then possible to bound the variance of γ along that same direction which in turns give a bound on the entropy of γ in a similar fashion to lemma 3.4. In order to state this idea formally we introduce the concept of *local* quadratic detachment for a positive function.

Definition 4.1 (Local quadratic detachment). Let $X \subset \mathbb{R}^d$ a compact set and $E : X \times X \rightarrow \mathbb{R}_+$ continuous. Assume that $\Sigma = \{E = 0\} \neq \emptyset$. We says that E has a *local quadratic detachment* of parameters $((U_i)_i, (\Phi_i)_i, \kappa)$ where $\kappa > 0$, $(U_i)_i$ is a finite open covering of Σ , and $(\Phi_i)_i$ a family of volume preserving affine functions $\Phi_i : \mathbf{x} \mapsto (u, v) \in \mathbb{R}^d \times \mathbb{R}^d$ such that for all $(u, v), (u, v') \in \Phi_i(U_i)$

$$E(u, v) + E(u, v') \geq \kappa \|v - v'\|^2 \quad (52)$$

Where E denotes $E \circ \Phi_i^{-1}$ by a slight abuse of notation.

This local quadratic detachment property is a direct generalisation of the notion of quadratic detachment (definition 3.1). And it happens that the same result holds: the entropy of γ is lower bounded by the log of the integral of a function having a local quadratic detachment. The idea of proof is the same as in the quadratic case with some added technical issues associated to the locality of the quadratic detachment.

Proposition 4.2. *Let $X \subset \mathbb{R}^d$ a compact set and E a continuous function on $X \times X$ that has a local quadratic detachment with parameters $((U_i)_{i=1}^N, (\Phi_i)_{i=1}^N, \kappa)$. Set $R = X \times X \setminus \bigcup_i U_i$. There exists a constant C depending on $X, \|\Phi_i\|_{op}$ and N such that for any $\gamma \in \mathcal{P}(X \times X)$ we have*

$$H(\gamma) \geq -\frac{d}{2} \ln \left(\int E(\mathbf{x}) d\gamma \right) - \frac{d}{2} \ln \left(\frac{4\pi e}{\kappa d} \right) + \frac{d}{2} \ln \left(1 \wedge \inf_{\mathbf{x} \in R} E(\mathbf{x}) \right) + C \quad (53)$$

Proof. First if $H(\gamma) = +\infty$ or $\int E d\gamma = +\infty$, there is nothing to prove. We assume that $H(\gamma) < +\infty$ thus γ has a density that we denote ρ and $\int E d\gamma < +\infty$. Let $\kappa > 0$, $(U_i)_{i=1}^N$ and Φ_i as in the definition 4.1 of the local quadratic detachment. Let ζ_i a partition of unity subordinate to the family $(U_i)_i$. Denote by $U = \bigcup_i U_i$. If $p_i = (\zeta_i \rho)(U) > 0$ let $\rho_i = \frac{1}{p_i} \zeta_i \rho$, else $\rho_i = 0$. Then we have the following

decomposition $\rho = 1_U \sum_i p_i \rho_i + 1_{U^c} \rho$ which we can inject in the definition of the entropy

$$\begin{aligned}
H(\gamma) &= \int \rho \ln(\rho) \\
&= \int_U \sum_i p_i \rho_i \ln(\sum_j p_j \rho_j) + \int_{U^c} \rho \ln(\rho) \\
&= \sum_i p_i \int \rho_i \ln(\sum_j p_j \rho_j) + \int_{U^c} \rho \ln(\rho) \\
&\geq \sum_i p_i \int \rho_i \ln(p_i \rho_i) + \int_{U^c} \rho \ln(\rho) \\
&\geq \sum_i p_i \ln(p_i) + \sum_i p_i \int \rho_i \ln(\rho_i) + \int_{U^c} \rho \ln(\rho) \\
&\geq \sum_i p_i H(\rho_i) - \frac{1}{e} (N + \mathcal{H}^{2d}(X \times X))
\end{aligned} \tag{54}$$

Where the first inequality holds because $p_i \rho_i \leq \sum_j p_j \rho_j$ and the last inequality holds because under support constraint the uniform distribution has the smallest entropy. Now for i such that $\rho_i \neq 0$ set $\tau_i = (\Phi_i)_\# \rho_i$, since Φ_i is volume preserving we have $H(\tau_i) = H(\rho_i)$. We now disintegrate τ_i with respect to the projection onto the first coordinate u and denote the disintegration $\tau_i = \mu_i \otimes \tau_i^u$. By additivity of the entropy we have

$$H(\tau_i) = H(\mu_i) + \int H(\tau_i^u) d\mu_i \tag{55}$$

However let u in the support of μ_i and let v within the support of τ_i^u . Take $v' \in \pi_2 \Phi_i(X \times X)$ such that $E(u, v') = \min_{\pi_1 \Phi_i(X \times X)} E(u, \cdot)$, which is possible by compactness of $\pi_2 \Phi_i(X \times X)$ and continuity of E . Then by the local quadratic detachment of E we have

$$E(u, v) \geq \frac{1}{2}(E(u, v') + E(u, v)) \geq \frac{\kappa}{2} \|v' - v\|^2 \tag{56}$$

Moreover since the variance is the distance to the constants in L^2 space we have

$$\int E(u, v) d\tau_i^u \geq \frac{\kappa}{2} \int \|v' - v\|^2 d\tau_i^u \geq \frac{\kappa}{2} \text{Var}(\tau_i^u) \tag{57}$$

Using this control of the variance in conjunction with the fact that the gaussian has lowest entropy under variance constraint we have

$$\begin{aligned}
H(\tau_i^u) &\geq -\frac{d}{2} \ln(\text{Var}(\tau_i^u)) - \frac{d}{2} \ln\left(\frac{2\pi e}{d}\right) \\
&\geq -\frac{d}{2} \ln\left(\int E(u, v) d\tau_i^u\right) - \frac{d}{2} \ln\left(\frac{4\pi e}{\kappa d}\right)
\end{aligned} \tag{58}$$

On the other hand we know that μ_i is supported in $\pi_1 \Phi_i(U_i)$ the diameter of which is smaller than a constant time that of X , thanks to the continuity of Φ_i . Thus since on a bounded support the uniform distribution has the smallest entropy we have $H(\mu_i) \geq -\ln(\mathcal{H}^d(\pi_1 \Phi_i(U_i))) = C(\|\Phi_i\|_{op}, X)$.

Now coming back to equation (55) we have

$$\begin{aligned}
H(\rho_i) &= H(\tau_i) = H(\mu_i) + \int H(\tau_i^u) d\mu_i \\
&\geq -\frac{d}{2} \int \ln \left(\int E(u, v) d\tau_i^u \right) d\mu_i - \frac{d}{2} \ln \left(\frac{4\pi e}{\kappa d} \right) + C(\|\Phi_i\|_{op}, X) \\
&\geq -\frac{d}{2} \ln \left(\int E(u, v) d\tau_i \right) - \frac{d}{2} \ln \left(\frac{4\pi e}{\kappa d} \right) + C(\|\Phi_i\|_{op}, X) \\
&\geq -\frac{d}{2} \ln \left(\int E(\mathbf{x}) d\rho_i \right) - \frac{d}{2} \ln \left(\frac{4\pi e}{\kappa d} \right) + C(\|\Phi_i\|_{op}, X)
\end{aligned} \tag{59}$$

where the second to last inequality holds by concavity of the logarithm and the last inequality holds because Φ_i is volume preserving. We now multiply equation (59) by p_i and sum over i in order to combine it with the lower bound term in equation (54)

$$\begin{aligned}
H(\gamma) &\geq \sum_i p_i H(\rho_i) - \frac{1}{e} (N + \mathcal{H}^{2d}(X \times X)) \\
&\geq \sum_i p_i \left(-\frac{d}{2} \ln \left(\int E(\mathbf{x}) d\rho_i \right) - \frac{d}{2} \ln \left(\frac{4\pi e}{\kappa d} \right) + C(\|\Phi_i\|_{op}, X) \right) - \frac{1}{e} (N + \mathcal{H}^{2d}(X \times X)) \\
&\geq -\frac{d}{2} \rho(U) \ln \left(\int_U E(\mathbf{x}) d\rho \right) - \frac{d}{2} \ln \left(\frac{4\pi e}{\kappa d} \right) + C(N, \|\Phi_i\|_{op}, X)
\end{aligned} \tag{60}$$

where we used the concavity of the logarithm one last time. It remains to prove that there is a constant such that $-\rho(U) \ln \left(\int_U E d\rho \right) \geq -\ln \left(\int E d\rho \right) + C$.

Let $E_0 = 1 \wedge \inf_{\mathbf{x} \in U^c} E(\mathbf{x}) > 0$ because $\Sigma \cap U^c = \emptyset$ is a compact set and E is continuous. Then $\int_{U^c} E d\rho \geq E_0(1 - \rho(U))$ which implies $-\ln \left(\int E d\rho \right) \leq -\ln \left(\int_U E d\rho + E_0(1 - \rho(U)) \right)$. For $\alpha \in [0, 1]$ and $x > 0$ set $f : x \mapsto -\alpha \ln(x) + \ln(x + E_0(1 - \alpha))$. We have $f(x) \geq f(\alpha E_0) = -\alpha \ln(\alpha) + (1 - \alpha) \ln(E_0) \geq \ln(E_0)$. This allows us to conclude that

$$H(\gamma) \geq -\frac{d}{2} \ln \left(\int E(\mathbf{x}) d\rho \right) - \frac{d}{2} \ln \left(\frac{4\pi e}{\kappa d} \right) + \frac{d}{2} \ln(E_0) + C(N, \|\Phi_i\|_{op}, X) \tag{61}$$

□

4.2 Rates of convergence of OT_ε

The quadratic case showed that the duality gap function $E = c - (\varphi \oplus \psi)$ has a global quadratic detachment, where φ, ψ are any Kantorovich potentials. In fact the duality gap still has a quadratic detachment, though local, in the more general case of infinitesimally twisted costs, which includes situations where the optimal plans are not necessarily given by a map. This is the point of [7, lemma 4.2], which relies on a generalized version of Minty's trick also stating that in charts the support of γ_0 is the graph of a Lipschitz function. They also provide a quadratic lower bound on the duality gap in these local charts. Our goal is to then apply proposition 4.2 in order to control the entropy of the optimal regularized transport plans as $\varepsilon \rightarrow 0$. We will then be able to conclude by finding a rate of order ε for $(c, \gamma_\varepsilon) - (c, \gamma_0)$ in a similar fashion to the quadratic case.

Throughout this section we will work on X a compact subset of \mathbb{R}^d . We denote by Ω an open set containing X . We will assume that the marginals μ_0, μ_1 are supported in X and have finite entropy. We start by recalling the results found in [7].

Definition 4.3 (Infinitesimal twist). Given $c \in \mathcal{C}^2(\Omega^2)$ we say that c is infinitesimally twisted if $\nabla_{xy}^2 c(x, y) = (\partial_{x_i y_j}^2 c(x, y))_{i,j} \in M_d(\mathbb{R})$ is invertible for every $(x, y) \in \Omega^2$.

We now recall the quadratic detachment lemma [7, lemma 4.2]. Their proof closely follow an earlier proof found in [28], however they are interested in points that do not belong to the support of the optimal transport plan.

Lemma 4.4. *Let $c \in \mathcal{C}^2(\Omega^2)$ be an infinitesimally twisted cost, and $(\varphi, \psi) \in \mathcal{C}(X)^2$ be a pair of c -conjugate functions. We denote by $E = c - \varphi \oplus \psi \geq 0$ the duality gap function defined on $X \times X$, by $\Sigma = \{E = 0\}$ the contact set and for every $r > 0$,*

$$\tau(r) = \sup_{\substack{\mathbf{x}, \mathbf{x}' \in X \times X \\ \|\mathbf{x}' - \mathbf{x}\| \leq r}} \|\nabla_{xy}^2 c(\mathbf{x}')^{-1} \nabla_{xy}^2 c(\mathbf{x}) - Id\| \in [0, \infty) \quad (62)$$

If $\bar{\mathbf{x}} \in X \times X$ and $\mathbf{x}, \mathbf{x}' \in B_r(\bar{\mathbf{x}}) \cap (X \times X)$, then

$$E(\mathbf{x}) + E(\mathbf{x}') \geq \|\Delta v\|^2 - \|\Delta u\|^2 - \tau(r)(\|\Delta v\|^2 + \|\Delta u\|^2) \quad (63)$$

where $\Delta u = u(\mathbf{x}') - u(\mathbf{x})$, $\Delta v = v(\mathbf{x}') - v(\mathbf{x})$, and

$$u(\mathbf{x}) = \frac{1}{2}(x - \nabla_{xy}^2 c(\bar{\mathbf{x}})y), \quad v(\mathbf{x}) = \frac{1}{2}(x + \nabla_{xy}^2 c(\bar{\mathbf{x}})y) \quad (64)$$

for every $\mathbf{x} = (x, y)$.

Remark 4.5. *The following remark was made in [7]. What directly follows from the lemma is that locally, up to a change of variable, the support of γ_0 lies within the graph of a Lipschitz function, with Lipschitz constant arbitrarily close to 1. Indeed for $(u, v), (u', v') \in \Phi(B(\bar{\mathbf{x}}, r)) \cap \Phi(\text{supp}\gamma_0)$ we have*

$$0 = E(u, v) + E(u', v') \geq (1 - \tau(r))\|\Delta v\|^2 - (1 + \tau(r))\|\Delta u\|^2 \quad (65)$$

Thus

$$\sqrt{\frac{1 + \tau(r)}{1 - \tau(r)}} \|\Delta u\| \geq \|\Delta v\| \quad (66)$$

This lemma essentially says that locally around the set $\Sigma = \{E = 0\}$ the function E grows at least quadratically along the direction $\text{Im}(Id + \nabla_{xy}^2 c(\bar{\mathbf{x}}))$. In particular when the functions (φ, ψ) are Kantorovich potentials of the optimal transport problem, this gives a quadratic lower bound of the duality gap close to Σ . The following lemma restates lemma 4.4 in the language of the quadratic detachment of the duality gap function.

Lemma 4.6. *Using the notations of lemma 4.4. Let $(\varphi, \psi) \in \mathcal{C}(X)^2$ be a pair of c -conjugate functions. Then $E = c - (\varphi \oplus \psi)$ has a local quadratic detachment of parameters $((B(\mathbf{x}_i, r))_i, (\Phi_i)_i, \kappa)$ where r is such that $\tau(r) \leq \frac{1}{2}$, and $\kappa = \frac{1}{4} \inf_{\mathbf{x} \in X^2} |\det(\nabla_{xy}^2 c(\mathbf{x}))|^{1/d}$. The family $(B(\mathbf{x}_i, r))_i$ is a finite covering of Σ , and $\Phi_i(\mathbf{x}) = \alpha_i(u_i(\mathbf{x}) - u_i(\mathbf{x}_i), v_i(\mathbf{x}) - v_i(\mathbf{x}_i))$ with α_i chosen such that $\mathcal{J}\Phi_i = 1$ and*

$$u_i(\mathbf{x}) = \frac{1}{2}(x - \nabla_{xy}^2 c(\mathbf{x}_i)y), \quad v_i(\mathbf{x}) = \frac{1}{2}(x + \nabla_{xy}^2 c(\mathbf{x}_i)y) \quad (67)$$

Proof. We use the notations of lemma 4.4. Let $r > 0$ such that $\tau(r) \leq 1/2$ which is possible since by continuity of $\nabla_{xy}^2 c$ over the compact set X^2 . We have $\tau(r) \rightarrow 0$ as $r \rightarrow 0$. Note that E is continuous because φ, ψ are c -conjugate functions and c is continuous. Thus Σ is closed, hence compact. We can choose $(\mathbf{x}_i)_{i=1}^N \in \Sigma^N$ such that $(B(\mathbf{x}_i, r))_i$ is a finite covering of Σ . For $i \in [1, N]$ and for every $\mathbf{x}, \mathbf{x}' \in B(\mathbf{x}_i, r)$ such that $u_i(\mathbf{x}) = u_i(\mathbf{x}')$, we have by lemma 4.4

$$E(\mathbf{x}) + E(\mathbf{x}') \geq (1 - \tau(r))\|\Delta v_i\|^2 \geq \frac{1}{2\alpha_i^2} \|\Delta(\alpha_i v_i)\|^2 \quad (68)$$

where $v_i(\mathbf{x}) = 1/2(x + \nabla_{xy}^2 c(\mathbf{x}_i)y)$. Note that Φ_i is an affine volume preserving map thus we have $1 = \frac{\alpha_i^{2d}}{2^d} |\det(\nabla_{xy}^2 c(\mathbf{x}_i))|$. The determinant and the cross derivative are continuous which implies $I = \inf_{\mathbf{x} \in X^2} |\det(\nabla_{xy}^2 c(\mathbf{x}))| > 0$ by the infinitesimal twist condition on c . We conclude that

$$E(\mathbf{x}) + E(\mathbf{x}') \geq \frac{1}{2\alpha_i^2} \|\Delta(\alpha_i v_i)\|^2 \geq \frac{I^{1/d}}{4} \|\Delta(\alpha_i v_i)\|^2 \quad (69)$$

□

We will use this lemma in the specific case of a duality gap function where φ, ψ are Kantorovich potentials for the optimal transport problem. Note that there is no dependence in ε and thus the quadratic detachment framework can be used to derive the rates found in the quadratic cost case. It is known (see [15, 7]) that for infinitesimally twisted cost and compactly supported marginals of finite entropy, the regularized problem ε EOT satisfies the following inequality for some real M

$$OT_\varepsilon - OT_0 \leq -\frac{d}{2}\varepsilon \ln(\varepsilon) + M\varepsilon \quad (70)$$

We formally recall that convergence result in the following lemma.

Lemma 4.7. [15, Theorem 3.8, Lemma 3.13] *Assume that μ_i are compactly supported and $c \in \mathcal{C}^2(\Omega^2)$ is infinitesimally twisted. Then there is $C > 0$ such that, as $\varepsilon \rightarrow 0$*

$$OT_\varepsilon - OT_0 \leq -\frac{d}{2}\varepsilon \ln(\varepsilon) + C\varepsilon \quad (71)$$

Proof. Let X compact such that $\text{supp}(\mu_i) \subset X$. Then by continuity, $\nabla^2 c$ is bounded. Moreover since μ_i are compactly supported [15, lemma 3.13] ensures that the optimal transport plan γ_0 satisfies the right quantization property for [15, theorem 3.8] to apply and grant a constant $C > 0$ such that

$$OT_\varepsilon - OT_0 \leq -\frac{d}{2}\varepsilon \ln(\varepsilon) + C\varepsilon \quad (72)$$

□

Note that $OT_\varepsilon - OT_0 = \int E d\gamma_\varepsilon + \varepsilon H(\gamma_\varepsilon)$. Thus in combination with lemma 4.7 and the lower bound on the entropy 4.2 we have the following result.

Theorem 4.8. *Let $c \in \mathcal{C}^2(\Omega^2)$ infinitesimally twisted. Let $\mu_i \in \mathcal{P}_{ac}(\Omega)$ two compactly supported measures satisfying $H(\mu_i) < \infty$. Then*

$$(c, \gamma_\varepsilon) - (c, \gamma_0) = \Theta(\varepsilon), \quad H(\gamma_\varepsilon) = -\frac{d}{2} \ln(\varepsilon) + O(1) \quad \text{and} \quad OT_\varepsilon = OT_0 - \frac{d\varepsilon}{2} \ln(\varepsilon) + O(\varepsilon) \quad (73)$$

Proof. We denote by X a compact subset of Ω such that $\mu_i(X) = 1$. It is known that the Kantorovich potentials are c -conjugate functions and thus are continuous on Ω . In particular the duality gap function $E(x, y) = c(x, y) - \varphi(x) - \psi(y)$ is continuous on $\Omega \times \Omega$. By lemma 4.6 we thus know that E has a local quadratic detachment. Thus proposition 4.2 implies that there is a constant C such that

$$H(\gamma_\varepsilon) \geq -\frac{d}{2} \ln \left(\int E d\gamma_\varepsilon \right) + C \quad (74)$$

Now combining this inequation with the upper bound on the rate of convergence (70) we have

$$-\frac{d}{2}\varepsilon \ln(\varepsilon) + C\varepsilon \geq \int E d\gamma_\varepsilon + \varepsilon H(\gamma_\varepsilon) \geq \int E d\gamma_\varepsilon - \frac{d}{2}\varepsilon \ln \left(\int E d\gamma_\varepsilon \right) + C\varepsilon \quad (75)$$

Dividing both sides by ε and adding $\ln(\varepsilon)$ we get

$$\frac{\int E d\gamma_\varepsilon}{\varepsilon} - \frac{d}{2} \ln \left(\frac{\int E d\gamma_\varepsilon}{\varepsilon} \right) \leq C \quad (76)$$

And we conclude as in theorem 3.7. □

4.3 Lower bound on $W_2(\gamma_\varepsilon, \gamma_0)$

We now focus on the lower bound of the Wasserstein distance between γ_ε and γ_0 . Unlike the quadratic case we won't be able to directly use a quadratic detachment for the Wasserstein distance. For $\varepsilon > 0$, by construction γ_ε is absolutely continuous with respect to the Lebesgue measure, thus there is an optimal transport map for the quadratic cost from γ_ε to γ_0 . In particular this map is the gradient of a convex function f . We are now able to write the Wasserstein distance as $W_2^2(\gamma_\varepsilon, \gamma_0) = \int \|\mathbf{x} - \nabla f(\mathbf{x})\|^2 d\gamma_\varepsilon(\mathbf{x})$. Inspired by the quadratic case and the last section we could say that $\mathbf{x} \mapsto \|\mathbf{x} - \nabla f(\mathbf{x})\|^2$ has a local quadratic detachment, but this function is not necessarily continuous which prevents us from applying the results on quadratic detachment. However the spirit of the proofs remains true and the results of this section are essentially an adaptation of the results presented before. First the Wasserstein distance between two measures satisfies a property close to a quadratic detachment whenever one of the measure is supported on a Lipschitz graph.

Lemma 4.9. *Let E a subspace of \mathbb{R}^n , $T : E \rightarrow E^\perp$ and $\mu = (Id \times T)_\# \mu_0$ with μ_0 a probability on E with finite moment of order 2. Suppose that T is L -Lipschitz. Then for any $\nu \in \mathcal{P}_{2,ac}(\mathbb{R}^d)$*

$$W_2^2(\mu, \nu) \geq (1 - L)W_2^2(\mu_0, \nu_0) + \frac{1}{L + 1} \int \text{Var}(\nu_x) d\nu_0(x) \quad (77)$$

where $\nu = \nu_0 \otimes \nu_x$ is the disintegration of ν with regard to the orthogonal projection on E . In particular

$$(1 + L)^2 W_2^2(\mu, \nu) \geq \int \text{Var}(\nu_x) d\nu_0(x) \quad (78)$$

Proof. Let π be the optimal coupling between μ, ν . Let $(X', Y') \sim \mu$ and $(X, Y) \sim \nu$ such that $(X', Y', X, Y) \sim \pi$. Thus

$$W_2^2(\mu, \nu) = \mathbb{E}_\pi [\|X - X'\|^2 + \|Y - T(X')\|^2] \quad (79)$$

. Now using that T is L -Lipschitz:

$$L\|X - X'\| + \|Y - T(X')\| \geq \|Y - T(X)\|. \quad (80)$$

and taking the square

$$L^2\|X - X'\|^2 + \|Y - T(X')\|^2 + 2L\|X - X'\|\|Y - T(X')\| \geq \|Y - T(X)\|^2. \quad (81)$$

Thus we have

$$\begin{aligned} W_2^2(\nu, \mu) &\geq \mathbb{E}_\pi [\|X - X'\|^2 + \|Y - T(X')\|^2] \\ &\geq (1 - L^2)\mathbb{E}_\pi [\|X - X'\|^2] + \mathbb{E}_\pi [\|Y - T(X)\|^2] \\ &\quad - 2L\mathbb{E}_\pi [\|X - X'\|\|Y - T(X')\|] \end{aligned} \quad (82)$$

but the last term satisfies by Young inequality $ab \leq \frac{1}{2}(a^2 + b^2)$

$$\mathbb{E}_\pi [\|X - X'\|\|Y - T(X')\|] \leq \frac{1}{2}(\mathbb{E}_\pi [\|X - X'\|^2] + \mathbb{E}_\pi [\|Y - T(X')\|^2]) \leq \frac{W_2^2(\mu, \nu)}{2} \quad (83)$$

Inequality (82) becomes

$$W_2^2(\mu, \nu)(1 + L) \geq (1 - L^2)\mathbb{E}_\pi [\|X - X'\|^2] + \mathbb{E}_\pi [\|Y - T(X)\|^2] \quad (84)$$

and finally we have

$$\begin{aligned} W_2^2(\mu, \nu) &\geq (1 - L)\mathbb{E}_\pi [\|X - X'\|^2] + \frac{1}{1 + L}\mathbb{E}_\pi [\|Y - T(X)\|^2] \\ &\geq (1 - L)W_2^2(\mu_0, \nu_0) + \frac{1}{1 + L} \int \mathbb{E}_{\nu_x} [\|Y - \mathbb{E}[Y | X = x]\|^2] d\nu_0 \\ &\geq (1 - L)W_2^2(\mu_0, \nu_0) + \frac{1}{1 + L} \int \text{Var}(\nu_x) d\nu_0 \end{aligned} \quad (85)$$

Where the second inequality is true since the mean is the L^2 orthogonal projection of the random variable on the space of constants. Now since $W_2^2(\mu, \nu) \geq W_2^2(\mu_0, \nu_0)$ we have

$$(1 + L)W_2^2(\mu, \nu) \geq W_2^2(\mu, \nu) + LW_2^2(\mu_0, \nu_0) \geq W_2^2(\mu_0, \nu_0) + \frac{1}{1 + L} \int \text{Var}(\nu_x) d\nu_0 \quad (86)$$

Which finally grants $(1 + L)^2 W_2^2(\mu, \nu) \geq \int \text{Var}(\nu_x) d\nu_0$. \square

The proof can be slightly modified to show that the function $\mathbf{x} \mapsto d(\mathbf{x}, \Gamma)^2$ where Γ is the graph of a Lipschitz function has a quadratic detachment. As stated before, to use the local quadratic detachment apparatus one could show that $\mathbf{x} \mapsto \|\mathbf{x} - \nabla f(\mathbf{x})\|^2$ has a local quadratic detachment where ∇f is the optimal transport from γ_ε to γ_0 . And thanks to remark 4.5, γ_0 is supported on a Lipschitz graph on charts. In particular if on an open set U , γ_0 is the graph of a Lipschitz function, then $\mathbf{x} \mapsto \|\mathbf{x} - \nabla f(\mathbf{x})\|^2$ satisfies a quadratic detachment on $(\nabla f)^{-1}(U)$. However since we have no information on the regularity of the transport map we cannot conclude that $(\nabla f)^{-1}(U)$ is open and observe a local quadratic detachment. Thus we have to adapt the proof of proposition 4.2 in order to manage this issue and derive a lower bound for the entropy of γ_ε using $W_2^2(\gamma_\varepsilon, \gamma_0)$.

Proposition 4.10. *Let $\mu_i \in \mathcal{P}(\Omega)$ two compactly supported measures of finite entropy. Then there exists $C > 0$ such that as $\varepsilon \rightarrow 0$*

$$H(\gamma_\varepsilon) \geq -\frac{d}{2} \ln(W_2^2(\gamma_\varepsilon, \gamma_0)) + C \quad (87)$$

Proof. Throughout we use the notations of lemma 4.4. Since γ_ε has a density, there is a transport map T for the optimal transport problem with quadratic cost from γ_ε to γ_0 . Let $r > 0$ such that $\tau(r) \leq \frac{1}{2}$. Let $(U_i = B(\mathbf{x}_i, r))_i$ an open covering of the support of γ_0 with $\mathbf{x}_i \in \text{supp} \gamma_0$. Let $(\zeta_i)_i$ a partition of unity subordinate to the covering $(U_i)_i$. For i set $\gamma_\varepsilon^i = \frac{1}{p_i} \zeta_i \circ T \gamma_\varepsilon$ and $\gamma_0^i = \frac{1}{p_i} \zeta_i \gamma_0$, where p_i is the normalization constant which is independent of ε . Note that T transports γ_ε^i to γ_0^i . Thus we have $W_2^2(\gamma_\varepsilon, \gamma_0) \geq \sum_i p_i W_2^2(\gamma_\varepsilon^i, \gamma_0^i)$. We now introduce $\Phi_i : \Phi_i(x, y) = \alpha_i(u_i(x, y) - u_i(\mathbf{x}_i), v_i(x, y) - v_i(\mathbf{x}_i))$ where α_i is such that Φ_i is volume preserving. Let $\tau_\varepsilon^i, \tau_0^i$ the pushforwards of γ_ε^i and γ_0^i with respect to this map. We disintegrate τ_ε^i with respect to the projection on the variable u and denote it $\mu_\varepsilon^i \otimes \tau_\varepsilon^{i,u}$. Note that Φ_i is not an isometry, however it is an affine map thus $W_2^2(\gamma_\varepsilon^i, \gamma_0^i) \geq C_i W_2^2(\tau_\varepsilon^i, \tau_0^i)$, where $C_i = 1/\|\Phi_i\|_{op}$. Note that C_i only depends on $\nabla_{xy}^2 c(\mathbf{x}_i)$, thus it is independent of ε . It was pointed out in remark 4.5 that for every i , τ_0^i is supported on the graph of a $\sqrt{3}$ -Lipschitz function. Thus thanks to lemma 4.9

$$\begin{aligned} H(\gamma_\varepsilon^i) &= H(\tau_\varepsilon^i) = H(\mu_\varepsilon^i) + \int H(\tau_\varepsilon^{i,u}) d\mu_\varepsilon^i \\ &\geq H(\mu_\varepsilon^i) - \frac{d}{2} \ln \left(\int \text{Var}(\tau_\varepsilon^{i,u}) d\mu_\varepsilon^i \right) + C \\ &\geq H(\mu_\varepsilon^i) - \frac{d}{2} \ln(W_2^2(\tau_\varepsilon^i, \tau_0^i)) + \frac{d}{2} \ln \left(\frac{C_i}{\sqrt{3} + 1} \right) + C \\ &\geq -\frac{d}{2} \ln(W_2^2(\gamma_\varepsilon^i, \gamma_0^i)) + C(X) \end{aligned} \quad (88)$$

Where the last inequality holds because μ_ε^i is supported on a compact set of diameter comparable to that of X , as seen in the proof of theorem 4.2. We proceed as in the proof of proposition 4.2. Thus by summing over i we get

$$\begin{aligned} H(\gamma_\varepsilon) &= \sum_i p_i \ln(p_i) + \sum_i p_i H(\gamma_\varepsilon^i) \\ &\geq -\frac{d}{2} \sum_i p_i \ln(W_2^2(\gamma_\varepsilon^i, \gamma_0^i)) + C \\ &\geq -\frac{d}{2} \ln(W_2^2(\gamma_\varepsilon, \gamma_0)) + C \end{aligned} \quad (89)$$

Where the last inequality holds by concavity of the logarithm. □

It remains to combine the last result with the rate of convergence of (ε EOT) in order to retrieve the negative result on the Wasserstein distance between γ_ε and γ_0 .

Theorem 4.11. *Let $c \in \mathcal{C}(\Omega^2)$ infinitesimally twisted. Let $\mu_i \in \mathcal{P}(\Omega)$ two compactly supported measures of finite entropy. Then there exists $c > 0$ such that as $\varepsilon \rightarrow 0$*

$$W_2^2(\gamma_\varepsilon, \gamma_0) \geq c\varepsilon \tag{90}$$

Proof. Lemma 4.7 ensures that

$$\varepsilon H(\gamma_\varepsilon) \leq \int E d\gamma_\varepsilon + \varepsilon H(\gamma_\varepsilon) \leq -\frac{d}{2}\varepsilon \ln(\varepsilon) + C\varepsilon \tag{91}$$

And proposition 4.10 grants the following lower bound for the entropy

$$H(\gamma_\varepsilon) \geq -\frac{d}{2}\ln(W_2^2(\gamma_\varepsilon, \gamma_0)) + C \tag{92}$$

Combining the two equations together we have

$$-\frac{d}{2}\ln(W_2^2(\gamma_\varepsilon, \gamma_0)) \leq -\frac{d}{2}\ln(\varepsilon) + C \tag{93}$$

Taking the exponential grants the result. □

Remark that in the general case, we obtain only the domination $\varepsilon = O(W_2^2(\gamma_\varepsilon, \gamma_0))$ whereas in the quadratic case, even if it is under strong assumptions, we could obtain $W_2^2(\gamma_\varepsilon, \gamma_0) = \Theta(\varepsilon)$ (theorem 3.9). The reason is not that the solution are not smooth enough. Indeed, there exist some precise assumptions, such as the Ma-Trudinger-Wang conditions (see [26]), that guarantee regularity of the solutions of the dual problem. The true difficulty is to replace the formula $W_2^2(\gamma_\varepsilon, \gamma_0) \leq L(E, \gamma_\varepsilon)$ of lemma 3.8 that transform this regularity (the lipschitz constant L of the transport map) into a bound of $W_2^2(\gamma_\varepsilon, \gamma_0)$.

References

- [1] Jason M. Altschuler, Jonathan Niles-Weed, and Austin J. Stromme. “Asymptotics for Semidiscrete Entropic Optimal Transport”. In: *SIAM Journal on Mathematical Analysis* 54.2 (Mar. 2022), pp. 1718–1741. DOI: 10.1137/21m1440165. URL: <https://doi.org/10.1137>.
- [2] Jean-David Benamou et al. “Iterative Bregman Projections for Regularized Transportation Problems”. In: *SIAM Journal on Scientific Computing* 37.2 (2015), A1111–A1138. DOI: 10.1137/141000439. eprint: <https://doi.org/10.1137/141000439>. URL: <https://doi.org/10.1137/141000439>.
- [3] Robert J. Berman. “Convergence Rates for Discretized Monge–Ampère Equations and Quantitative Stability of Optimal Transport”. In: *Foundations of Computational Mathematics* 21.4 (Dec. 2020), pp. 1099–1140. DOI: 10.1007/s10208-020-09480-x. URL: <https://doi.org/10.1007/s10208-020-09480-x>.
- [4] Espen Bernton, Promit Ghosal, and Marcel Nutz. “Entropic optimal transport: Geometry and large deviations”. In: *Duke Mathematical Journal* 171.16 (2022), pp. 3363–3400.
- [5] Luis Caffarelli. “The Regularity of Mappings with a Convex Potential”. In: *Journal of The American Mathematical Society - J AMER MATH SOC* 5 (Jan. 1992). DOI: 10.2307/2152752.
- [6] Luis A. Caffarelli. “Monotonicity Properties of Optimal Transportation and the FKG and Related Inequalities”. In: *Communications in Mathematical Physics* 214.3 (Nov. 2000), pp. 547–563. DOI: 10.1007/s002200000257. URL: <https://doi.org/10.1007/s002200000257>.

- [7] Guillaume Carlier, Paul Pegon, and Luca Tamanini. “Convergence rate of general entropic optimal transport costs”. In: *arXiv preprint arXiv:2206.03347* (2022).
- [8] Sinho Chewi and Aram-Alexandre Pooladian. *An entropic generalization of Caffarelli’s contraction theorem via covariance inequalities*. 2022. DOI: 10.48550/ARXIV.2203.04954. URL: <https://arxiv.org/abs/2203.04954>.
- [9] Lenaïc Chizat et al. “Faster Wasserstein distance estimation with the Sinkhorn divergence”. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 2257–2269.
- [10] R. Cominetti and J. San Martín. “Asymptotic analysis of the exponential penalty trajectory in linear programming”. In: *Mathematical Programming* 67.1-3 (Oct. 1994), pp. 169–187. DOI: 10.1007/bf01582220. URL: <https://doi.org/10.1007/bf01582220>.
- [11] Giovanni Conforti. “A second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost”. In: *Probability Theory and Related Fields* 174.1-2 (2019), pp. 1–47.
- [12] Giovanni Conforti and Luca Tamanini. *A formula for the time derivative of the entropic cost and applications*. 2019. arXiv: 1912.10555 [math.PR].
- [13] Max Costa. “A new entropy power inequality”. In: *IEEE Transactions on Information Theory* 31.6 (1985), pp. 751–760.
- [14] Marco Cuturi. “Sinkhorn Distances: Lightspeed Computation of Optimal Transportation Distances”. In: (2013). DOI: 10.48550/ARXIV.1306.0895. URL: <https://arxiv.org/abs/1306.0895>.
- [15] Stephan Eckstein and Marcel Nutz. *Convergence Rates for Regularized Optimal Transport via Quantization*. 2022. DOI: 10.48550/ARXIV.2208.14391. URL: <https://arxiv.org/abs/2208.14391>.
- [16] Matthias Erbar, Kazumasa Kuwada, and Karl-Theodor Sturm. “On the equivalence of the entropic curvature-dimension condition and Bochner’s inequality on metric measure spaces”. In: *Inventiones mathematicae* 201.3 (2015), pp. 993–1071.
- [17] Matthias Erbar, Jan Maas, and Michiel Renger. “From large deviations to Wasserstein gradient flows in multiple dimensions”. In: *Electronic Communications in Probability* 20 (2015), pp. 1–12.
- [18] Jean Feydy et al. “Interpolating between Optimal Transport and MMD using Sinkhorn Divergences”. working paper or preprint. Oct. 2018. URL: <https://hal.archives-ouvertes.fr/hal-01898858>.
- [19] Joel Franklin and Jens Lorenz. “On the scaling of multidimensional matrices”. In: *Linear Algebra and its Applications* 114-115 (1989). Special Issue Dedicated to Alan J. Hoffman, pp. 717–735. ISSN: 0024-3795. DOI: [https://doi.org/10.1016/0024-3795\(89\)90490-4](https://doi.org/10.1016/0024-3795(89)90490-4). URL: <https://www.sciencedirect.com/science/article/pii/0024379589904904>.
- [20] Ivan Gentil et al. “An entropic interpolation proof of the HWI inequality”. In: *Stochastic Processes and their Applications* 130.2 (2020), pp. 907–923.
- [21] Nicola 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 (Mar. 2011), pp. 401–409. DOI: 10.1017/S001309150800117x. URL: <https://doi.org/10.1017/S001309150800117x>.
- [22] Nicola Gigli and Luca Tamanini. “Benamou–Brenier and duality formulas for the entropic cost on $RCD^*(K, N)$ spaces”. In: *Probability Theory and Related Fields* 176.1 (2020), pp. 1–34.
- [23] Siegfried Graf and Harald Luschgy. *Foundations of Quantization for Probability Distributions*. Berlin, Heidelberg: Springer-Verlag, 2000. ISBN: 3540673946.
- [24] Christian Léonard. *A survey of the Schrödinger problem and some of its connections with optimal transport*. 2013. arXiv: 1308.0215 [math.PR].

- [25] Wenbo Li and Ricardo H Nochetto. “Quantitative stability and error estimates for optimal transport plans”. In: *IMA Journal of Numerical Analysis* 41.3 (July 2020), pp. 1941–1965. DOI: 10.1093/imanum/draa045. URL: <https://doi.org/10.1093/imanum/draa045>.
- [26] Xi-Nan Ma, Neil S Trudinger, and Xu-Jia Wang. “Regularity of potential functions of the optimal transportation problem”. In: *Archive for rational mechanics and analysis* 177 (2005), pp. 151–183.
- [27] Simone Di Marino and Augusto Gerolin. “An Optimal Transport Approach for the Schrödinger Bridge Problem and Convergence of Sinkhorn Algorithm”. In: *Journal of Scientific Computing* 85.2 (Oct. 2020). DOI: 10.1007/s10915-020-01325-7. URL: <https://doi.org/10.1007/s10915-020-01325-7>.
- [28] Robert J McCann, Brendan Pass, and Micah Warren. “Rectifiability of optimal transportation plans”. In: *Canadian Journal of Mathematics* 64.4 (2012), pp. 924–934.
- [29] Paul Milgrom and Ilya Segal. “Envelope theorems for arbitrary choice sets”. In: *Econometrica* 70.2 (2002), pp. 583–601.
- [30] George J. Minty. “Monotone (nonlinear) operators in Hilbert space”. In: *Duke Mathematical Journal* 29.3 (1962), pp. 341–346. DOI: 10.1215/S0012-7094-62-02933-2. URL: <https://doi.org/10.1215/S0012-7094-62-02933-2>.
- [31] Soumik Pal. *On the difference between entropic cost and the optimal transport cost*. 2019. arXiv: 1905.12206 [math.PR].
- [32] Luigia Ripani. “Convexity and regularity properties for entropic interpolations”. In: *Journal of Functional Analysis* 277.2 (2019), pp. 368–391.
- [33] R. Tyrrell Rockafellar and Roger J.-B. Wets. *Variational Analysis*. Heidelberg, Berlin, New York: Springer Verlag, 1998.
- [34] Richard Sinkhorn. “A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices”. In: *The Annals of Mathematical Statistics* 35.2 (June 1964), pp. 876–879. DOI: 10.1214/aoms/1177703591. URL: <https://doi.org/10.1214/aoms/1177703591>.
- [35] C. Villani and American Mathematical Society. *Topics in Optimal Transportation*. Graduate studies in mathematics. American Mathematical Society, 2003.
- [36] Jonathan Weed. *An explicit analysis of the entropic penalty in linear programming*. 2018. DOI: 10.48550/ARXIV.1806.01879. URL: <https://arxiv.org/abs/1806.01879>.