Optimal transport on Wiener space with different norms

Vincent NOLOT

Institut de Mathématiques de Bourgogne,
Université de Bourgogne, 21078 Dijon, France.
vincent.nolot@u-bourgogne.fr

Abstract

In this paper we study two basic facts of optimal transportation on the classical Wiener space $W := C_0([0,1],\mathbb{R})$. Our first aim is to answer to the Monge Problem on the Wiener space endowed with the Sobolev type norm $\|\cdot\|_{k,\gamma}^p$ with $p \geq 1$ (cases $p = 1$ and $p > 1$ are considered apart). The second one is to prove $1$-convexity (resp. $1/C_{k,\gamma}^2$-convexity) along (constant speed) geodesics of the relative entropy in $(\mathcal{P}_2(W),W_2)$, where $W$ is endowed with the uniform norm (resp. with $\|\cdot\|_{k,\gamma}$), and $W_2$ is the $2$-distance of Wasserstein.

1 Introduction

We are interested in two problems in optimal transportation on Wiener space. We refer to a recent work of Ambrosio and Gigli [?] for a survey and basic tools of optimal transport theory.

Firstly we answer to the Monge Problem, that is

$$\inf_{G\#\rho_0=\rho_1} \int_W \|w - G(w)\|_{k,\gamma}^p d\rho_0(w),$$

on the Wiener space $W$ relatively to the cost $\|\cdot\|_{k,\gamma}^p$ defined as

$$\|x\|_{k,\gamma} := \left( \int_0^1 \int_0^1 \frac{\left( t_x(s) - x(s) \right)^{2k}}{|t-s|^{1+2k\gamma}} dt ds \right)^{1/2k},$$

for suitable parameters $k, \gamma$. This cost was introduced by Airault and Malliavin in [?] and has nice properties.

Monge Problem is largely considered in several settings since for years. Nowadays there are a lot of means to prove the existence of an optimal map resolving this Problem, which are summarized in [?].

Recently, there was considerable advances concerning Monge Problem in $\mathbb{R}^n$. First in 1996, Gangbo and McCann solved Monge Problem in [?] when the cost is strictly
In this paper, we turn our attention on the Wiener space, an infinite dimensional space. The Monge Problem in the Wiener space was solved by Feyel and Üstünel in \([9]\) for the cost \(|x|^2\) induced by Cameron-Martin norm \(|\cdot|_H\) which is Hilbertian. With this cost, we have unicity of optimal map. Strategy of these latter authors is to pass by finite dimensional approximations of Wiener space: the aim being to reduce the Monge Problem on Wiener space onto finite dimensional spaces and applying known results. Then they use a selection theorem to go back up on Wiener space. The Monge Problem was solved by Cavalletti in \([12]\) for the cost \(|\cdot|_H\), again passing by finite dimensional approximations.

We are interested to endow the Wiener space \((W,H,\mu)\) (where \(\mu\) is the Wiener measure) with two other natural norms: the uniform norm \(|\cdot|_{\infty}\) and the Sobolev type norm \(\|\cdot\|_{k,\gamma}\). For the first one, Monge Problem is still open, and we can expect not to have unicity of optimal map, providing it exists somehow. Here the norm considered \(\|\cdot\|_{k,\gamma}\) is not Hilbertian and is weaker than the Cameron-Martin norm in sense that for some \(C_{k,\gamma}>0\):

\[
\|x\|_{k,\gamma} \leq C_{k,\gamma}|x|_H \quad \text{for all } x \in W. \tag{1.2}
\]

Let us emphasize that the right hand side is infinite \(\mu\text{-}\)almost everywhere, because the Cameron-Martin space has measure 0. Nevertheless our norm \(\|\cdot\|_{k,\gamma}\) satisfies suitable conditions presented in section 2.1. The first aim of our paper is to solve the Monge Problem for the cost \(\|\cdot\|_{k,\gamma}^p\). In other words we will establish following

**Theorem 1.1** Let \(\rho_0\) and \(\rho_1\) be two measures on \(W\) satisfying condition (1.3) below,

1. If \(p>1\) and \(\rho_0\) is absolutely continuous with respect to \(\mu\), then there exists a map \(T\) unique up to a set of zero measure for \(\rho_0\), which minimizes (1.1). Moreover there is a unique optimal transference plan between \(\rho_0\) and \(\rho_1\) relatively to the cost \(\|\cdot\|_{k,\gamma}^p\), which is exactly \((Id \times T)_{\#}\rho_0\).
2. If \(p=1\) and both \(\rho_0\) and \(\rho_1\) are absolutely continuous with respect to \(\mu\), then there exists an optimal transference plan \(\Pi\) between \(\rho_0\) and \(\rho_1\) relatively to the cost \(\|\cdot\|_{k,\gamma}\), such that \(\Pi\) is concentrated on a graph of some map \(T\) minimizing (1.1).

The second purpose of this paper is to find a lower bound of Ricci curvature in the sense of Sturm in \([7]\), for Wiener space endowed with the uniform norm \(|\cdot|_{\infty}\) or the norm \(\|\cdot\|_{k,\gamma}\). It is relied to weak \(K\)-convexity along geodesic of relative entropy. More precisely we will prove:

**Theorem 1.2** If \(\rho_0\) and \(\rho_1\) are probability measures on \(W\) both absolutely continuous with respect to \(\mu\), then there exists some (constant speed) geodesic \(\rho_t\) induced by an
optimal transference between $\rho_0$ and $\rho_1$ such that:

$$\operatorname{Ent}_\mu(t) \leq (1-t)\operatorname{Ent}_\mu(\rho_0) + t\operatorname{Ent}_\mu(\rho_1) - \frac{Kt(1-t)}{2}W_{2,\infty}^2(\rho_0, \rho_1) \quad \forall t \in [0,1].$$

This theorem provides an example of $CD(K, \infty)$ space. For the uniform norm, we will see that $K$ equals to 1, and for the Sobolev type norm $\|\cdot\|_{k,\gamma}$ that $K$ equals to $1/C_k^2\gamma$. Precise that Lott-Villani introduced in [7] a stronger notion of weak $K$-convexity of relative entropy. Indeed in its definition, it is required that the property above holds for all (constant speed) geodesics. So in many cases weak $K$-convexity and $K$-convexity coincide, provided that there is unicity of geodesics between two given measures (this is the case for non branching spaces).

Now let us briefly summarize the following sections.

Throughout section 2, we solve the Monge Problem in the Wiener space with the cost $\|\cdot\|_k^p$, in different approaches according to parameter $p$. Among them, there is a direct method: in general settings, the support of an optimal transference plan $\Pi$ (between two probability measures and relatively to a cost function $c$) is included in $c$-subdifferential of a $c$-convex function (called potential of Kantorovich) $\phi$. It leads to the following system:

$$\begin{cases}
\phi^c(y) - \phi(x) = c(x,y) & \Pi - \text{almost everywhere} \\
\phi^c(y) - \phi(x) \leq c(x,y) & \text{everywhere}
\end{cases}$$

And this system can be solved directly when the cost $c$ and the potential $\phi$ are differentiable, so long as $\nabla c(x, \cdot)$ is injective, as it is explained in Villani’s book [7]. This is the case when $p > 1$. But this method fails when $p = 1$. In the latter case we head for a recent paper of Bianchi and Cavalletti [2] where the authors solve the Monge Problem in non branching geodesic metric spaces. It turns out that Wiener space endowed with the norm $\|\cdot\|_{k,\gamma}$ is a such space.

Section 3 is devoted to establish the $K$-convexity along geodesics of relative entropy (w.r.t. Wiener measure) on the Wiener space endowed with the uniform norm ($K = 1$), then with the Sobolev type norm $\|\cdot\|_{k,\gamma}$ ($K = 1/C_k^2\gamma$). This time we will process by finite dimensional approximation as Fang, Shao and Sturm in [2], who have treated the case of the norm $\|\cdot\|_H$. This part requires Wasserstein distance, which is defined below. Our main contribution consists in establishing results without applying powerful tools like Gromov-Hausdorff convergence (see [2]) or $\mathbb{D}$-convergence introduced by Sturm in [2]. In the language of latter authors, we can say that $(W,\|\cdot\|_\infty)$ is a $CD(1,\infty)$ space (satisfies such curvature-dimension condition) and $(W,\|\cdot\|_{k,\gamma})$ is a $CD(1/C_k^2\gamma, \infty)$ space. Such conditions imply a lot of important results over Wiener spaces. As consequences over spaces $(W,\|\cdot\|_\infty)$, $(W,\|\cdot\|_{k,\gamma})$, we can quote Brunn-Minkowski, Bishop-Gromov or also Log-Sobolev inequalities (see [2]).

1.1 Settings

Our ambient space will be the classical Wiener space $(W, H, \mu) = (C_0([0,1], \mathbb{R}), H(\mathbb{R}), \mu)$ where $H(\mathbb{R}) := \{h : [0,1] \rightarrow \mathbb{R}; \ h(t) = \int_0^t \dot{h}(s)ds \text{ and } \dot{h} \in L^2([0,1])\}$ and $\mu$ is the
1.1 Settings

Introduction

Wiener measure. $H(\mathbb{R})$ is a Hilbert space with the inner product:

$$(h,g)_H := \int_{[0,1]} \dot{h}(t)\dot{g}(t)dt.$$  

Important facts are that $H$ is dense in $W$ with respect to the uniform norm, and moreover $\mu(H) = 0$.

Given two Borel measures $\rho_0$ and $\rho_1$ on $W$, let us state a condition of preserving mass, necessary for all our discussion:

$$\int_W d\rho_0 = \int_W d\rho_1 < +\infty. \tag{1.3}$$

When (1.3) is satisfied, we will consider the following Monge-Kantorovich Problem

$$\inf_{\Pi \in \mathcal{P}(\rho_0,\rho_1)} \mathcal{I}_p(\Pi) = \inf_{\Pi \in \mathcal{P}(\rho_0,\rho_1)} \int_{W \times W} \|x-y\|^{2}_{k,\gamma}d\Pi(x,y), \tag{1.4}$$

which is a relaxed problem of Monge Problem (1.1) stated above. Provided that the cost $c$ is continuous and the ambient space is Polish, there is always an existing minimizer for the Monge-Kantorovich Problem (1.4) (see for example [?]). A such minimizer will be called an optimal coupling or optimal transference plan between $\rho_0$ and $\rho_1$.

We endow the space $W$ with the following norm:

$$\|x\|_{k,\gamma} := \left( \int_0^1 \int_0^1 \left( x(t) - x(s) \right)^{2k} \left| t - s \right|^{1+2k\gamma} dtds \right)^{1/2k}$$

such that $0 < \gamma < 1/2$, and $2 < 1 + 2k\gamma < k$ where $k$ is an integer. Actually this is a pseudo-norm over $W$ since it can take infinite value. For this reason, we consider $\hat{W} := \{ x \in W; \|x\|_{k,\gamma} < \infty \}$. It is well known that $\mu(\hat{W}) = 1$ hence for a sake of notation, we will write $W = \hat{W}$. $(\hat{W}, \|\cdot\|_{k,\gamma})$ is a separable Banach space, and all measures considered in the sequel will be Borel with respect to the topology induced by the norm $\|\cdot\|_{k,\gamma}$. $H$ is still dense in $(\hat{W}, \|\cdot\|_{k,\gamma})$. We can write

$$\|x\|^2_{k,\gamma} \leq \int_0^1 \int_0^1 \int_0^1 |\dot{x}(\xi)|^2 |s-t|^{1+2k\gamma} dtds$$

$$\leq \left( \int_0^1 |\dot{x}(\xi)|^2 d\xi \right) \int_0^1 \int_0^1 |t-s|^{1+2k\gamma} dtds$$

$$\leq C_{k,\gamma}^2 \|x\|^2_H,$$

where $C_{k,\gamma} := \left( \int_0^1 \int_0^1 |t-s|^{k-1-2k\gamma} dtds \right)^{1/2k}$.

Before to continue, let us explain how we can decompose the classical Wiener space in finite dimensional spaces. Consider the projections $\pi_n : W \rightarrow \hat{W}$ defined as

$$\pi_n(x)(t) := x \left( \frac{k}{2^n} \right)^n + 2^n \left( t - \frac{k}{2^n} \right) \left[ x \left( \frac{k+1}{2^n} \right) - x \left( \frac{k}{2^n} \right) \right], \text{ if } t \in \left[ \frac{k}{2^n}, \frac{k+1}{2^n} \right].$$
At a continuous path, $\pi_n$ associates its affine part. Denote by $V_n := \pi_n(W)$. We know that $V_n$ is spanned by the Haar functions and therefore has dimension $2^n$ and then will be identified with $\mathbb{R}^N$.

Denote by $V_n^\perp$ the subset of $W$ which is image of $W$ by the map $I_W - \pi_n$. Since $\pi_n$ is a projection, one can write $W = V_n \oplus V_n^\perp$.

An important fact is the following: the image measure of the Wiener measure $\mu$ by $\pi_n$ is the standard Gaussian measure on $V_n$. In other words $(\pi_n)_#\mu = \gamma_d$, where the indice $d$ is the dimension of $V_n$, i.e. $2^n = d$. For a sake of simplicity, we will denote $(\pi_n)_#\mu = \gamma_n$ instead of $\gamma_d$.

Recall that the Sobolev space over the Wiener space $D_{1,p}(W)$ is the set of all $\mu$-measurable function $F \in L^p(W,\mu)$ such that there exists $\nabla F \in L^p(W,H)$ where we have $\mu$–almost surely:

$$(\nabla F(w),h)_H = D_h F(w) := \lim_{\epsilon \to 0} \frac{F(w + \epsilon h) - F(w)}{\epsilon} \text{ in } L^p(W,\mu) \text{ } \forall h \in H.$$ 

We will denote

$$D_{1}^\infty(W) := \bigcap_{p>0} D_{1,p}(W) \quad \text{and} \quad \mathbb{E}(F) := \mathbb{E}_\mu(F).$$

In section 3 we will need 2–Wasserstein distance between two probability measures $\rho_0$ and $\rho_1$ on a measurable space $W$, defined as

$$W_{2,c}^2(\rho_0,\rho_1) := \inf_{\Pi \in \Gamma(\rho_0,\rho_1)} \int c(x,y)^2 d\Pi(x,y)$$

where $\Gamma(\rho_0,\rho_1)$ is the set of coupling measures between $\rho_0$ and $\rho_1$, i.e the measures on $W \times W$ with marginals $\rho_0$ and $\rho_1$. Moreover we will denote by $\Gamma_0(\rho_0,\rho_1)$ the set of optimal couplings for $c^2$ or equivalently the set of couplings which realize minimum on 2–Wasserstein distance induced by the cost $c$. A sufficient condition (but not natural) for that the Wasserstein distance is that $\rho_0$ and $\rho_1$ have second finite moments. To justify terminology, notice that $W_{2,c}$ is well a distance so long as $c$ is a distance on $W$ and in this case the space

$$\mathcal{P}_2(W) := \left\{ \rho \text{ probability measure on } W; \int_W c(x,x_0)^2 d\rho(x) < \infty \text{ for some } x_0 \in W \right\}$$

endowed with $W_{2,c}$ is a metric space.

In all this paper, $Pr_i$ (with $1 \leq i \leq N$) stand for the projections onto the $i$–th component:

$$Pr_i : X_1 \times X_2 \times \ldots X_N \to X_i,$$

where $N \in \mathbb{N}$ depends on the context.
2 Monge Problem on Wiener space with $\|\cdot\|^{p}_{k,\gamma}$

We can now focus on the Monge Problem. Notice that $(W, \|\cdot\|_{k,\gamma})$ is a Polish space and in addition the cost $c(x, y) = \|x - y\|^{p}_{k,\gamma}$ being continuous (for all $p \geq 1$), there is always an existing measure $\Pi \in \Gamma(\rho_0, \rho_1)$ which attains the minimum in the Monge-Kantorovich Problem (1.4).

For $\rho_0, \rho_1$ two measures on $W$ satisfying (1.3), recall that the Monge problem between $\rho_0$ and $\rho_1$ consists of finding an optimal map $T: W \to W$ which pushes $\rho_0$ forwards to $\rho_1$ and minimizes the quantity:

$$\int_{W} c(x, G(x))d\rho_0(x),$$

among all push-forward maps $G$ (i.e. $G_{\#}\rho_0(E) := \rho_0(G^{-1}(E)) = \rho_1(E)$ for all Borel $E$ subset of $W$).

The usual strategy is to use characterization of optimal coupling, with the help of Kantorovich Potentials. Hence we need the concept of $c$-convexity.

**Definition 2.1** Let $\varphi : W \to \mathbb{R}$. We say that $\varphi$ is $c$-convex if

$$\varphi(x) = \sup_{y \in W} (\varphi^{c}(y) - c(x, y)) \quad \forall x \in W,$$

where $\varphi^{c}$, called $c$-transform of $\varphi$, is defined as

$$\varphi^{c}(y) = \inf_{x \in W} (\varphi(x) + c(x, y)) \quad \forall y \in W.$$

Notice that our cost $c$ does not take an infinite value, so $c$-convex functions are real-valued well.

It is well known that any optimal coupling is $c$-cyclically monotone i.e. its support (defined as the smaller closed subset of $W \times W$ having full $\Pi$-measure) is $c$-cyclically monotone, namely: for all $N \in \mathbb{N}$ and $(x_1, y_1), \ldots, (x_N, y_N) \in \text{Supp}(\Pi)$ we have:

$$\sum_{i=1}^{N} c(x_i, y_i) \leq \sum_{i=1}^{N} c(x_i, y_{i+1}),$$

with $y_{N+1} := y_1$.

Rather there is equivalence (in our case) between optimality and $c$-cyclical monotonicity (see e.g. [?]). In particular if $\text{Supp}(\Pi)$ is $c$-cyclically monotone then any coupling $\Pi$ such that $\text{Supp}(\Pi) \subset \text{Supp}(\Pi)$ is also $c$-cyclically monotone.

Let us precise what are the difficulties we have met.

- When $p = 1$, $c(x, y) = \|x - y\|_{k,\gamma}$ is the considered norm on $W$. Hence if a map $\varphi$ is $c$-convex then it is also $1$-Lipschitz, hence $H$-Lipschitz. Indeed,

$$|\varphi(x + h) - \varphi(x)| \leq \|h\|_{k,\gamma} \leq C_{k,\gamma}\|h\|_{H} \quad \forall h \in H \quad \forall x \in W.$$

In this case, we have a version of Rademacher theorem (proved in [?]) and recalled in appendix for a sake of completeness) on Wiener space, for such $H$-Lipschitz
functions. With the remark above, it leads to \(c\)-convex functions on the Wiener space are almost surely differentiable. But the difficulty in this case is that the cost, being a norm, is not strictly convex, so we do not have injectivity of \(y \mapsto \nabla_x c(x, y)\), therefore our method is not available. Nevertheless we shall use an other method to solve Monge Problem. Indeed thanks to the part 2. of Lemma (2.1), \((W, \|\cdot\|_{k, \gamma})\) is a geodesic non branching space (i.e. geodesics cannot bifurcate). So we can apply method, detailed in [?].

- When \(p > 1\), \(c(x, y) = \|x - y\|_{k, \gamma}^p\) becomes strictly convex, we get this time the injectivity of \(\nabla_x c(x, \cdot)\). But we loose the \(H\)-Lipschitz property of \(c\)-convex functions. Indeed if \(\varphi\) is such function we can write:

\[
|\varphi(x) - \varphi(y)| \leq \|x - \xi\|_{k, \gamma}^{p} - \|y - \xi\|_{k, \gamma}^{p} \\
\leq \|x - y\|_{k, \gamma}M_{\xi},
\]

where the latter constant \(M_{\xi}\) depends on \(\xi\) and cannot be bounded. However we will see that in this case \(c\)-convex functions (hence potentials) are locally \(H\)-Lipschitz. Since differentiability is a local property, we should be able to apply Rademacher theorem again.

Before we continue, let us set properties of the considered norm, which will be useful for the sequel.

### 2.1 Properties of \(\|\cdot\|_{k, \gamma}\)

We give two ingredients that will be essential for the sequel.

**Lemma 2.1** If we denote by \(\tilde{F} : W \to \mathbb{R}_+\) the map \(\tilde{F}(x) = \|x\|_{k, \gamma}\), then we have the following properties:

1. \(\tilde{F}\) admits a gradient \(\nabla \tilde{F}(w)\) belonging to \(W^*\) for all \(x \in W \setminus \{0\}\), where \(W^*\) is the dual of \(W\). Moreover \(F^p\) is everywhere differentiable for all \(p > 1\).

2. \(\tilde{F}\) is a norm such that its unit ball is strictly convex.

**Proof:**

1. First we show the property for \(F := \tilde{F}^{2k}\). Take \(h \in W\), we can write for \(x \in W\) and \(\epsilon > 0\):

\[
F(x + \epsilon h) = \int_0^1 \int_0^1 \frac{((w(t) - w(s)) + \epsilon (h(t) - h(s)))^{2k}}{|t - s|^{1+2k}} \, dt \, ds.
\]

And taking the derivative at \(\epsilon = 0\), it is clear that \(\lim_{\epsilon \to 0} \frac{F(x + \epsilon h) - F(x)}{\epsilon}\) exists and moreover:

\[
|D_h F(x)| \leq 2k \int_0^1 \int_0^1 \frac{|w(t) - w(s)|^{2k-1}}{|t - s|^{1+2k}} |h(t) - h(s)| \, dt \, ds \\
\leq 2k \int_{[0,1]^2} |w(t) - w(s)|^{2k-1} \frac{|h(t) - h(s)|}{|t - s|^{1+2k}} \, dt \, ds
\]
and now applying H"older’s inequality, we get:

\[
|D_{h}F(x)| \leq 2k \left( \int_{[0,1]^{2}} \frac{|x(t) - x(s)\|^{2k}}{|t - s|^{1+2k\gamma}} dt ds \right)^{(2k-1)/(2k)} \left( \int_{[0,1]^{2}} \frac{|h(t) - h(s)\|^{2k}}{|t - s|^{1+2k\gamma}} dt ds \right)^{1/(2k)}
\]

\[= 2k \|w\|_{k,\gamma}^{2k-1} \cdot \|h\|_{k,\gamma}.
\]

Hence \(h \mapsto D_{h}F(x)\) is a bounded operator on \(W\) for all \(x \in W\). It leads to existence of a gradient \(\nabla F(x)\) which belong to the dual space \(W^{*} \subset H^{*} = H\) (by (1.2)). Since \(\tilde{F} = F^{1/(2k)}\), its gradient satisfies \(\nabla \tilde{F}(x) = F^{1/(2k) - 1}(x)\nabla F(x)\) and in particular \(x\) must not be equal to zero.

Since \(\tilde{F}\) is differentiable out of \(\{0\}\) it is just a remark to see that for any \(p > 1\), \(\tilde{F}^{p}\) is differentiable everywhere over \((W, \|\cdot\|_{k,\gamma})\).

2. This proof is the same as the proof of Minkowski’s inequality. Indeed if \(w_{1}, w_{2} \in W\) and \(\eta \in (0, 1)\) then we have:

\[
\|(1 - \eta)w_{1} + \eta w_{2}\|_{k,\gamma}^{2k} = \int_{[0,1]^{2}} \frac{|(1 - \eta)(w_{1}(t) - w_{1}(s)) + \eta(w_{2}(t) - w_{2}(s))|^{2k}}{|t - s|^{1+2k\gamma}} dt ds
\]

\[= \int_{[0,1]^{2}} \frac{|(1 - \eta)(w_{1}(t) - w_{1}(s)) + \eta(w_{2}(t) - w_{2}(s))|}{|t - s|^{1+2k\gamma}} \times (1 - \eta)(w_{1}(t) - w_{1}(s)) + \eta(w_{2}(t) - w_{2}(s))^{2k-1} dt ds
\]

\[\leq \int_{[0,1]^{2}} \frac{|(1 - \eta)(w_{1}(t) - w_{1}(s))|}{|t - s|^{(1+2k\gamma)/(2k)}} \frac{|(1 - \eta)(w_{1}(t) - w_{1}(s)) + \eta(w_{2}(t) - w_{2}(s))^{2k-1}}{|t - s|^{1+2k\gamma} - \frac{1}{2k} - \gamma} dt ds
\]

\[+ \int_{[0,1]^{2}} \frac{|\eta(w_{2}(t) - w_{2}(s))|}{|t - s|^{(1+2k\gamma)/(2k)}} \frac{|(1 - \eta)(w_{1}(t) - w_{1}(s)) + \eta(w_{2}(t) - w_{2}(s))^{2k-1}}{|t - s|^{1+2k\gamma} - \frac{1}{2k} - \gamma} dt ds
\]

\[\leq (1 - \eta)\|w_{1}\|_{k,\gamma} + \eta\|w_{2}\|_{k,\gamma} \left( (1 - \eta)\|w_{1}\|_{k,\gamma} + \eta\|w_{2}\|_{k,\gamma} \right)^{1-1/2k}
\]

The two inequalities above are respectively triangle inequality and Holder’s inequality, and are in fact equality if and only if \(w_{1}\) and \(w_{2}\) are almost everywhere collinear. This leads to the strict convexity of our norm.

Conditions on parameters \((p, \gamma)\) are sufficient to have \(F \in \mathbb{D}_{p}^{\infty}(W)\), as it is shown in [?]. Notice that the (Gateaux) differentiability of \(\tilde{F}\) exists in the direction of \(W\), hence in particular in the direction of \(H\).

2.2 The case \(p > 1\).

Throughout this subsection, the cost is \(c(x, y) = \|x - y\|_{k,\gamma}^{p}\), with \(p > 1\).

We follow Fathi and Figalli in [?] to get around the fact that \(c\)-convex functions are not \(1\)-Lipschitz with respect to \(\|\cdot\|_{k,\gamma}\), but nevertheless are locally Lipschitz with restriction to suitable subsets. The key argument is that the sup of a family of uniformly \(\|\cdot\|_{k,\gamma}\)-Lipschitz functions, is also \(\|\cdot\|_{k,\gamma}\)-Lipschitz.
2.2 The case $p > 1$.

**Theorem 2.1** Let $\rho_0$ and $\rho_1$ be two measures on $W$ satisfying (1.3) and such that the first one is absolutely continuous with respect to the Wiener measure $\mu$. Assume $I(\Pi)$ is finite for some $\Pi \in \Gamma(\rho_0, \rho_1)$.

Then there exists a unique optimal coupling between $\rho_0$ and $\rho_1$ relatively to the cost $c$. Moreover it is concentrated on a graph of some Borel map $T : W \rightarrow W$ unique up to a set of zero measure for $\mu$.

**Proof:**

Let $\Pi \in \Gamma(\rho_0, \rho_1)$ be an optimal coupling for $c$. We shall show that $\Pi$ is concentrated on a graph of some Borel map. It is well known (see e.g. [?]) that under condition $I(\Pi)$ is finite, since $\text{Supp}(\Pi)$ is $c$-cyclically monotone, there is a $c$-convex map $\varphi : W \rightarrow \mathbb{R}$ (called Kantorovich potential) such that

$$\varphi^c(y) - \varphi(x) = \|x - y\|_{k,\gamma}^p \quad \Pi - \text{a.s.}$$

Moreover from the definition of $c$-convexity, we also have

$$\varphi^c(y) - \varphi(x) \leq \|x - y\|_{k,\gamma}^p \quad \forall (x, y) \in W \times W. \quad (2.6)$$

Since $\varphi^c$ is finite everywhere, if we consider subsets $W_n := \{\varphi^c \leq n\}$ for $n \in \mathbb{N}$ then:

$$W_n \subset W_{n+1} \quad \text{and} \quad \bigcup_{n \in \mathbb{N}} W_n = W.$$  

Our cost $c(. , y) = \|.- y\|_{k,\gamma}^p$ is locally $\|\|_{k,\gamma}$-Lipschitz locally uniformly in $y$. Hence for each $y \in W$ there exists a neighborhood $E_y$ of $y$ such that $(\|.- z\|_{k,\gamma}^p)_{z \in E_y}$ is a uniform family of locally $\|\|_{k,\gamma}$-Lipschitz functions. Moreover $W$ being separable, we can find a sequence $\{y_l\}_{l \in \mathbb{N}}$ of elements of $W$ such that:

$$\bigcup_{l \in \mathbb{N}} E_{y_l} = W.$$

Now consider increasing subsets of $W$:

$$V_n := W_n \bigcap \left( \bigcup_{l=1}^n E_{y_l} \right).$$

We can define maps approximating $\varphi$ as follow:

$$\varphi_n : W \rightarrow W$$

$$x \mapsto \sup_{y \in V_n} \left( \varphi^c(y) - \|x - y\|_{k,\gamma}^p \right).$$

Notice that

$$\varphi_n(x) = \max_{l=1, \ldots, n} \sup_{y \in W_n \cap E_{y_l}} \left( \varphi^c(y) - \|x - y\|_{k,\gamma}^p \right).$$

But since $\varphi \leq n$ on $W_n$ and $-\|\|_{k,\gamma}^p$ is bounded from above, $\varphi_n$ is also bounded from above. Therefore the sequel $(\varphi^c(y) - \|.- y\|_{k,\gamma}^p)_{y \in W_n \cap E_{y_l}}$ is uniformly locally
2.2 The case $p > 1$.

Monge Problem on Wiener space with $\|\cdot\|_{k,\gamma}^p$

$\|\cdot\|_{k,\gamma}$–Lipschitz and bounded from above. Finally Proposition 4 in Appendices shows that $\varphi_n$ being a maximum of uniformly locally $\|\cdot\|_{k,\gamma}$–Lipschitz functions, is also locally $\|\cdot\|_{k,\gamma}$–Lipschitz. We can extend $\varphi_n$ to a $\|\cdot\|_{k,\gamma}$–Lipschitz function everywhere on $W$ still denoted by $\varphi_n$. By (1.2), we get:

$$|\varphi_n(w + h) - \varphi_n(w)| \leq C\|h\|_{k,\gamma} \leq 2C\|h\|_H \quad \forall w \in W, \forall h \in H.$$ 

Namely, $\varphi_n$ is a $H$–Lipschitz function. Thanks to Rademacher theorem, there exists a Borel subset $F_n$ of $W$ with plain $\mu$–(hence $\rho_0$–)measure such that for all $x \in F_n$, $\varphi_n$ is differentiable at $x$. Then for each $x \in F := \cap_n F_n$ (which has also plain $\rho_0$–measure), each $\varphi_n$ is differentiable at $x$.

By increasing of $(V_n)_n$, it is clear that $\varphi_n \leq \varphi_{n+1} \leq \varphi$ everywhere on $W$. Moreover with same argument as in [?], if $P_n := Pr_1(\text{Supp}(\Pi) \cap (W \times V_n))$ then $\varphi|P_n = \varphi|\rho_n$ for all $l \geq n$ and all $n \in \mathbb{N}$. Fix $x \in P_n \cap F$. By definition of $P_n$ it exists $y_x \in V_n$ such as:

$$\varphi^0(y_x) - \varphi_n(x) = \|x - y_x\|_{k,\gamma}^p,$$

i.e. $\varphi^0(y_x) - \varphi(x) = \|x - y_x\|_{k,\gamma}^p$.

Subtracting (2.6) with $(x', y_x)$ to the previous equality, we get for all $x' \in W$ and $h \in H$:

$$\varphi(x) - \varphi(x') \geq \|x - y_x\|_{k,\gamma}^p - \|x' - y_x\|_{k,\gamma}^p.$$ 

Taking $x' = x + \epsilon h$ with $\epsilon > 0$, $h \in H$, dividing by $\epsilon$ and taking the limit when $\epsilon$ tends to 0, we get (by linearity in $h$):

$$\nabla \varphi(x) + \nabla x c(x, y_x) = 0. \quad (2.7)$$

Indeed $c(\cdot, y_x)$ is differentiable at $x$ thanks to Proposition 2.1. The strict convexity of $c(x, y) = \|x - y\|_{k,\gamma}^p$ yields $\nabla x c(x, \cdot)$ is injective and (2.7) gives:

$$y_x = (\nabla x c(x, \cdot))^{-1}(-\nabla \varphi(x)) =: T(x),$$

where $(\nabla x c(x, \cdot))^{-1}$ is the inverse of the map $y \mapsto \nabla x c(x, y)$. Notice here that $T$ is uniquely determined. We deduce that $\text{Supp}(\Pi) \cap (W \times V_n)$ is the graph of the map $T$ over $P_n \cap F$ for all $n \in \mathbb{N}$. But $(P_n)_n$ and $(V_n)_n$ are increasing and such that $\bigcup_n V_n = W$. Therefore $\text{Supp}(\Pi)$ is a graph over $Pr_1(\text{Supp}(\Pi)) \cap F$ with $Pr_1(\text{Supp}(\Pi)) = \bigcup_n P_n$.

We can extend $T$ onto a measurable map over $W$ as it is explained in [?]. We obtain $\text{Supp}(\Pi)$ is included in the graph of a measurable map $T$, unique up to a set of $\rho_0$–measure. In other words $\Pi = (id \times T)\#\rho_0$.

We have proved that any optimal coupling is carried by a graph of some map. So if $\Pi_1, \Pi_2 \in \Gamma(\rho_0, \rho_1)$ are optimal for $\|\cdot\|_{k,\gamma}$ then any convex combination of $\Pi_1$ and $\Pi_2$ is also optimal. Take $\Pi := \frac{1}{2}(\Pi_1 + \Pi_2)$ be an optimal coupling between $\rho_0$ and $\rho_1$: there exists some measurable map $T$ such that $\Pi = (Id \times T)\#\rho_0$. Let $f$ be the density of $\Pi_1$. 

---

10
with respect to $\Pi$. Then for any continuous bounded functions $\varphi$ we have:
\[
\int_W \varphi(x) d\rho_0(x) = \int_{W \times W} \varphi(x) d\Pi_1(x, y) \\
= \int_{W \times W} \varphi(x) f(x, y) d\Pi(x, y) \\
= \int_W \varphi(x) f(x, T(x)) d\rho_0(x).
\]
This yields $f(x, T(x)) = 1 \rho_0$–a.e., hence $f = 1$ –a.e. It leads to $\Pi = \Pi_1$ and finally $\Pi_2 = \Pi_1 = (Id \times T)_\#\rho_0$. ■

\section{The case $p = 1$.}

When $p = 1$, $c(x, y) = \|x - y\|_{k, \gamma}$. Hence if a map $\varphi$ is $c$–convex then it is $1$–Lipschitz, hence $H$–Lipschitz. Indeed:
\[
|\varphi(x + h) - \varphi(x)| \leq \|h\|_{k, \gamma} \leq C_{k, \gamma} |h|_H, \quad \forall h \in H \quad \forall x \in X.
\]
Therefore we can use Rademacher theorem \cite{Rad} on the Wiener space, to differentiate any $H$–Lipschitz functions. But the difficulty in this case is that the cost, being a norm, is not strictly convex, so we do not have injectivity of $y \mapsto \nabla_x c(x, y)$ as in the above section. The method used in first section requires the differentiation theorem for the Wiener measure, which is not available.

We will follow the method of \cite{Bianchini} developed by Bianchini and Cavalletti in \cite{Bianchini}. The method uses a selection theorem. By strict convexity of our norm $\|\cdot\|_{k, \gamma}$ proved in Lemma II, $(X, \|\cdot\|_{k, \gamma})$ is a geodesic non branching space.

We will not develop fully the method but only briefly indicate the different steps:

1. reduce the initial Monge-Kantorovich Problem to the one-dimensional Monge-Kantorovich Problem along distinct geodesics: this is possible since the space is non-branching.

2. verify that the conditional measures provided by disintegration of both measures $\rho_0$ and $\rho_1$ on each geodesic have no atoms: this is possible thanks to properties of Gaussian measure in order to apply result of one-dimensional Monge Problem; on each geodesic there exists a transport map.

3. piece obtained maps together to get a transport map for the initial Monge Problem by a general selection theorem.

We refer to \cite{Bianchini} and \cite{Bianchini} for more details. In our case, the cost $\|\cdot\|_{k, \gamma}$ is smooth enough (continuity) to guarantee the existence of a Kantorovich potential $\varphi$ (Proposition II) such that there is a $\sigma$–compact subset $\Gamma$ on which any optimal coupling $\Pi$ is concentrated and
\[
\Gamma := \{(x, y) \in X \times X; \varphi(y) - \varphi(x) = \|x - y\|_{k, \gamma}\}.
\]
2.3 The case $p = 1$. Monge Problem on Wiener space with $||\cdot||^p_{k,\gamma}$

From now, let us consider an optimal (relative to the cost $c(x, y) = ||x - y||_{k,\gamma}$) couplage $\Pi_0$ between two probability measures $\rho_0$ and $\rho_1$ on $X$, both absolutely continuous with respect to the Wiener measure $\mu$. Let $\pi_n : XV_n$ be the finite dimensional approximation, where $V_n$ is a space of functions piecewisely linear, described in Chapter ???. Denote by $\rho_0^n := (\pi_n)_#\rho_0$ and $\rho_1^n := (\pi_n)_#\rho_1$, which are absolutely continuous with respect to the Gaussian measure $\gamma_n$ on $V_n$. Since the restriction of $||\cdot||_{k,\gamma}$ on $V_n$ is differentiable out of 0, by a result due to Caffarelli, M. Feldman, and R.J. McCann [??], there is an optimal map $T^n : V_nV_n$ such that $\Pi_0^n := (id \times T^n)_#\rho_0^n$ is the unique optimal couplage between $\rho_0^n$ and $\rho_1^n$. In other words, $\Pi_0^n$ is concentrated on some Borel set $\Gamma_n \subset Graph(T^n)$.

The following result shows that the method of [??] really works well.

**Proposition 2.1** Assume that there exists $M > 0$ such that densities $f_0$ and $f_1$ of respectively $\rho_0^n$ and $\rho_1^n$ are bounded by $M$. Then the following estimate holds true for all Borel subset $A \subset V_n$:

$$\gamma_n(T_{n,t}(A)) \geq \frac{1}{M} \rho_0^n(A) \quad \forall t \in [0,1],$$

where $T_{n,t} := (1 - t) Id + t T^n$.

We will follow the proof of [??]. The only difference is to consider Monge maps for the cost induced by $||\cdot||^p_{k,\gamma}$ with $(p > 1)$, instead of $||\cdot||^p_{k,\gamma}$. Indeed costs $||\cdot||^p_{k,\gamma}$ satisfy conditions of Proposition ??, so that the associated optimal maps $T_p^n$ are approximately differentiable.

**Proof:**

Fix $p > 1$. Since $\rho_0^n$ is absolutely continuous w.r.t. $\gamma_n := (\pi_n)_#\mu$, by Proposition ??, the Monge Problem

$$\int_{X} ||x - T(x)||^p_{k,\gamma} d\rho_0^n(x),$$

admits a unique solution $T_p$. Besides by Proposition ??, $T_p$ is approximately differentiable $\rho_0^n$-a.s, and by Lemma ??, 

$$f_0^n(x) \frac{1}{\sqrt{2\pi}} e^{-|x|^2/2} = f_1^n(T_p(x))|det(\nabla T_p(x))| e^{-\frac{1}{2} |T_p(x)|^2 }.$$

Besides $|det(\nabla T_p(x))| > 0$ and $f_1^n(T_p(x)) > 0$ for $\rho_0^n$-a.e. $x \in \mathbb{R}^n$. Hence we can write $\rho_0^n$-a.s.

$$|det(\nabla T_p(x))| = \frac{f_0^n(x)}{f_1^n(T_p(x))} \exp \left\{ -\frac{1}{2} (|x|^2 - |T_p(x)|^2) \right\}.$$

Now consider $T_{p,t} := (1 - t) Id + t T_p$. By the same arguments that in the proof of Proposition ?? and by the concavity of $t \mapsto det((1 - t) Id + t D)^{1/n}$, it holds

$$log \left( det(\nabla T_{p,t}(x))^{1/n} \right) \geq t log \left( det(\nabla T_p(x))^{1/n} \right).$$

Therefore:

$$det(\nabla T_{p,t}(x)) \geq |det(\nabla T_p(x))|^t \left( \frac{f_0^n(x)}{f_1^n(T_p(x))} \right)^t \exp \left\{ -\frac{t}{2} (|x|^2 - |T_p(x)|^2) \right\}.$$
2.3  The case $p = 1$.

Following [7], for any $A \in \mathcal{B}(n)$,
\[
\gamma_n(T_{p,t}(A)) = \int_A \det(\nabla T_{p,t}(x)) \exp\left\{-\frac{1}{2} |T_{p,t}(x)|^2\right\} \, dx
\]
\[
= \int_A \det(\nabla T_{p,t}(x)) \exp\left\{-\frac{1}{2} (|T_{p,t}(x)|^2 - |x|^2)\right\} \, d\gamma_n(x)
\]
\[
\geq \int_A \left(\frac{f_n^1(x)}{f_0^1(T_p(x))}\right)^t \exp\left\{\frac{1}{2} \|x - T_p(x)\|_2^2 (t - t^2)\right\} \, d\gamma_n(x)
\]
\[
\geq \frac{1}{M^t} \int_A f_n^1(x)^t d\gamma_n(x) = \frac{1}{M^t} \int_A f_0^1(x)^{t-1} \, d\rho_0^n(x) \geq \frac{1}{M} \rho_0^n(A).
\]

Since $(Id \times T_p)\#\rho_0^n$ converges weakly to $(Id \times T^n)\#\rho_0^n$, letting $p1$, proceeding as in [7], or in [7], we obtain
\[
\gamma_n(T_{n,t}(A)) \geq \frac{1}{M} \rho_0^n(A).
\]

Let $T_t(x, y) = (1 - t)x + ty$. Then the above result can be reformulated by
\[
\gamma_n(T_t(\Gamma_n \cap (A \times V_n))) \geq M \rho_0^n(A).
\]

Coming back to the Wiener space, we have the following result:

**Proposition 2.2**  Assume that the density of $\rho_0$ and $\rho_1$ with respect to $\mu$ are bounded by $M > 0$; then for any compact subset $A \subset X$, we have:
\[
\mu(T_t(\Gamma \cap A \times X)) \geq M \rho_0(A).
\]

The proof, given again in [7], holds true in a quite general setting, provided the cost is at least lower semi-continuous. Again following [7] step by step, we get the following result.

**Theorem 2.2** Let $\rho_0$ and $\rho_1$ be two probability measures on $X$ of finite entropy. Then there exists an optimal coupling between $\rho_0$ and $\rho_1$ which is concentrated on a graph of some Borel map $T : X \rightarrow X$.

Note that by Young inequality
\[
\|x\|_{k, \gamma}^2 f_0(x) \leq e^{\alpha \|x\|_{k, \gamma}^2} + \frac{f_0(x)}{\alpha} \log\left(\frac{f_0(x)}{\alpha}\right),
\]
we get
\[
\int_X \|x\|_{k, \gamma}^2 f_0(x) \, d\mu(x) \leq \int_X e^{\alpha \|x\|_{k, \gamma}^2} \, d\mu(x) + \mu(\rho_0/\alpha),
\]
which is finite if $\mu(\rho_0) < +\infty$, since by Fernique’s theorem
\[
\int_X e^{\alpha \|x\|_{k, \gamma}^2} \, d\mu(x) < +\infty
\]
for $\alpha$ small enough. Therefore any probability measure on $X$ has finite second moment with respect to $\|\cdot\|_{k, \gamma}^2$. 

---

13
3 Convexity of relative entropy on Wiener space

This part is split in two sections. The first one contains recalls in finite dimensional cases and an extension of known results for uniform and \( \| \cdot \|_{k,\gamma} \) norms on \( \mathbb{R}^d \). The second one will need this result to show our purpose on the Wiener space. Throughout all of this section, the cost \( c \) will be a distance induced by a norm. So that it takes sense to consider Wasserstein distance.

We consider the relative entropy with respect to a reference measure \( \gamma \) defined as

\[
\text{Ent}_\gamma(\rho) := \begin{cases} 
\int f \log(f) d\gamma & \text{if } \rho \text{ admits } f \text{ for density w.r.t } \gamma \\
+\infty & \text{otherwise}
\end{cases}
\] (3.8)

We need to recall the notion of geodesics on the space of probability measures. A \((c_{\text{onstant speed)}) \) geodesic, provided

\[
W_{2,c}(\rho_t^{(c)}, \rho_s^{(c)}) = (t-s)W_{2,c}(\rho_0, \rho_1) \quad \forall 0 \leq s \leq t \leq 1.
\]

One can obtain a constant speed geodesic by letting \( \rho_t^{(c)} := ((1-t)Pr_1 + tPr_2)\#\Pi \), \( \forall t \in [0,1] \), if \( \Pi \) is an optimal coupling for \( c \). In fact each optimal transference plan involves a constant speed geodesic (see for instance [?]). Moreover on non branching space, whenever the optimal transference plan is unique, there is a unique constant speed geodesic between \( \rho_0 \) and \( \rho_1 \). For example, Banach space with strictly convex norm is non branching, while Banach space with non strictly convex norm is branching.

Thanks to this definition, one can consider the notion of convexity (and weak convexity) along \((c_{\text{onstant speed)}) \) geodesics.

**Definition 3.1** For \( \rho_0, \rho_1 \) two probability measures with second finite moments (for a cost \( c \)), we say relative entropy with respect to a reference measure \( \mu \), is weakly \( K \)-convex along \((c_{\text{onstant speed)}) \) geodesics on \((\mathcal{P}_2(.), c) \) if there exists a \((c_{\text{onstant speed)}) \) geodesic \( \rho_t^{(c)} \) induced by an optimal transference plan belonging to \( \Gamma_0(\rho_0, \rho_1) \) such that:

\[
\text{Ent}_\mu(\rho_t^{(c)}) \leq (1-t)\text{Ent}_\mu(\rho_0) + t\text{Ent}_\mu(\rho_1) - \frac{Kt(1-t)}{2}W_{2,c}^2(\rho_0, \rho_1) \quad \forall t \in [0,1].
\]

We say that relative entropy is \( K \)-convex along \((c_{\text{onstant speed)}) \) geodesics on \((\mathcal{P}_2(.), c) \) (not weakly) if the latter inequality holds for all \((c_{\text{onstant speed)}) \) geodesics \( \rho_t^{(c)} \).

3.1 Finite dimensional cases

We want to show the relative entropy (with respect to the standard Gaussian measure \( \gamma_d \)) is \( K \)-convex along \((c_{\text{onstant speed)}) \) geodesics in the finite dimensional space \( \mathbb{R}^d \) endowed with the uniform norm and the norm \( \| \cdot \|_{k,\gamma} \). For the first one, we need to consider the cases where \( \mathbb{R}^d \) is endowed with different \( p \)-norms. For \( p \) positive integer we set

\[
c_p(x, y) := |x - y|_p = \left( \sum_{i=1}^{d} |x_i|^p \right)^{1/p}.
\]

Let us begin to recall \( K \)-convexity along geodesics for suitable norm on \( \mathbb{R}^d \).
3.1 Finite dimensional cases

**Proposition 3.1** Let \( \| \cdot \| \) be a strictly convex and differentiable norm on \( \mathbb{R}^d \setminus \{0\} \). If \( \sqrt{K} \| \cdot \| \leq \| \cdot \|_2 \) then relative entropy w.r.t. \( \gamma_d \) on \( (\mathbb{R}^d, \| \cdot \|) \) is \( K \)-convex along (constant speed) geodesics on \( (\mathcal{P}_2(\mathbb{R}^d), W_2) \), where \( W_2 \) is the Wasserstein distance induced by \( \| \cdot \| \).

By assumption on \( \| \cdot \| \), here \( (\mathbb{R}^d, \| \cdot \|) \) is a non branching space.

**Proof:**

Let \( \rho_0 \) and \( \rho_1 \) be two probability measures absolutely continuous with respect to \( \gamma_d \) (hence \( L \)) (otherwise the result is trivial). For \( i = 0, 1 \) let \( d\rho_0 = f_0 dL \) and \( d\rho_1 = f_1 dL \), then the density of probability of \( \rho_i \) with respect to \( \gamma_d \) is \( \frac{d\rho}{\gamma_d} = f_i (2\pi)^{\frac{d}{2}} e^{-\frac{|x|^2}{2}} \). Write

\[
\text{Ent}_{\gamma_d}(\rho_i) = \int_{\mathbb{R}^d} f_i(x)(2\pi)^{\frac{d}{2}} e^{-\frac{|x|^2}{2}} \log \left( f_i(x)(2\pi)^{\frac{d}{2}} e^{-\frac{|x|^2}{2}} \right) d\gamma_d(x)
\]

\[
= \int_{\mathbb{R}^d} f_i(x)(2\pi)^{\frac{d}{2}} e^{-\frac{|x|^2}{2}} \log(f_i(x)) d\gamma_d(x) + \int_{\mathbb{R}^d} f_i(x)(2\pi)^{\frac{d}{2}} e^{-\frac{|x|^2}{2}} \log((2\pi)^{\frac{d}{2}}) d\gamma_d(x)
\]

\[
+ \int_{\mathbb{R}^d} f_i(x)(2\pi)^{\frac{d}{2}} e^{-\frac{|x|^2}{2}} \left( \frac{|x|^2}{2} \right) d\gamma_d(x)
\]

\[
= \text{Ent}_L(\rho_i) + \frac{1}{2} \int |x|^2 d\rho_i(x) + \frac{d}{2} \log(2\pi).
\]

* First term of the latter equality is relative entropy with respect to \( L \). Since \( \| \cdot \|^2 \) is strictly convex and differentiable, it suffices to follow [?] to see that it is convex along geodesics on \( \mathcal{P}_2(\mathbb{R}^d, \| \cdot \|) \).

* Let us show that the second term of the latter equality \( \mu \to \int \frac{1}{2} |x|^2 d\mu(x) \) is \( 1 \)-convex along geodesics on \( \mathcal{P}_2(\mathbb{R}^d, \| \cdot \|) \). We know that the map \( x \to \frac{1}{2} |x|^2 \) is \( 1 \)-convex i.e for all \( x_1, x_2 \in \mathbb{R}^d \) and \( t \in [0, 1] \);

\[
\frac{1}{2} \| (1-t)x_1 + tx_2 \|^2 \leq \frac{1-t}{2} |x_1|^2 + \frac{t}{2} |x_2|^2 - \frac{t(1-t)}{2} |x_1 - x_2|^2 \leq \frac{1-t}{2} |x_1|^2 + \frac{t}{2} |x_2|^2 - \frac{Kt(1-t)}{2} \| x_1 - x_2 \|^2.
\]

Consider an optimal coupling \( \Pi \) (for \( \| \cdot \|^2 \) between \( \rho_0 \) and \( \rho_1 \). Then integrating the previous inequality w.r.t. \( \Pi \), it comes:

\[
\int \frac{1}{2} \| (1-t)x_1 + tx_2 \|^2 d\Pi(x_1, x_2) \leq \frac{1-t}{2} \int |x_1|^2 d\rho_0(x_1) + \frac{t}{2} \int |x_2|^2 d\rho_1(x_2) - \frac{Kt(1-t)}{2} W_2^2(\rho_0, \rho_1).
\]

* Finally relative entropy with respect to \( \gamma_d \) is \( 1 \)-convex along (constant speed) geodesics as sum of \( 0 \)-convex and \( 1 \)-convex maps.

We will apply Proposition 3.1 in the following cases:

* For the norm \( \| \cdot \|_p \) with \( 2 \leq p < \infty \) (\( K = 1 \)).

* For the norm induced by \( \| \cdot \|_{k, \gamma} \) (\( K = 1/C_{k, \gamma}^2 \)).

We shall extend the result for the uniform norm \( \| \cdot \|_\infty \). This latter fact is a priori not obvious since \( |x - y|_\infty \) is neither strictly convex nor differentiable on \( \mathbb{R}^d \setminus \{0\} \).
Now the question is: what is happening when \( p \) equals \(+\infty\)? When one changes the cost function, two items change in the above inequality: Wasserstein distances but also (constant speed) geodesics which depend implicitly on an optimal transference plan, which depends itself on cost function.

Let \( \rho_0 \) and \( \rho_1 \) be definitively fixed, as probability measures on \( \mathbb{R}^d \) with finite second moments (for \(|.|_{\infty}\)). In particular: \( W_{2,p}(\rho_0, \rho_1) < \infty \) for all \( p \geq 2 \). We know that for \( p \geq 2 \), there exists a unique optimal transference plan \( \Pi_0^{(p)} \) between \( \mu_0 \) and \( \mu_1 \) (for cost function \( c^2(p) \)). Thus we can watch behavior of the sequence \( (\Pi_0^{(p)})_p \). When \( p \) varies, cost function varies too. In fact \( (c_p)_p \) converges to \( c \) and it would be interesting if the sequence of optimal coupling converge to an optimal coupling for the uniform cost. Indeed it appears that up to a subsequence, \( (\Pi_0^{(p)})_p \) weakly converges to a probability measure which will be an optimal transference plan for the uniform cost. This fact combined with the property of lower semicontinuity of the relative entropy, which one adds the nonincreasing of the following sequel

\[ p \in \mathbb{N} \mapsto W_{2,p}^2(\mu_0, \mu_1) \]

will yield \( 1 \)-convexity of relative entropy along geodesics on \( (\mathbb{R}^d, |.|_{\infty}) \).

To prove the weak convergence of \( (\Pi_0^{(p)})_p \), a first (easy but useful) remark is that the sequel is tight. This yields, thanks to Prokohov's Theorem, there exists a subsequence \( (\Pi_0^{(p_k)})_p \) that we will denote by \( (\Pi_0^{(p_k)})_p \) again, converging weakly to a measure \( \Pi_\infty \). It is easy to check that \( \Pi_\infty \) is a coupling of \( \mu_0 \) and \( \mu_1 \). And we wish to see that \( \Pi_\infty \) is optimal for the cost \( c^2 \), where \( c(x,y) := |x-y|_{\infty} \). The difficulty lies in the fact that \( (c_p)_p \) converges to \( c \) but \textbf{not uniformly}. Nevertheless, it will work because the convergence is uniform on all compact subsets of \( \mathbb{R}^d \).

The next Lemma which appears to be an essential point for this purpose, is taken from [?]. Recall that for a probability measure \( \mu \), its support \( \text{Supp}(\mu) \), is defined as the smallest closed set on which \( \mu \) is concentrated (i.e. \( \mathbb{R}^d \setminus \text{Supp}(\mu) \) is \( \mu \)-negligible).

**Lemma 3.1** For all \( x \in \text{Supp}(\Pi_\infty) \), there exist \( x_p \in \text{Supp}(\Pi_0^{(p)}) \) such that \( \lim(x_p) = x \).

For self contained paper, here we include the proof.

**Proof:**

Let \( x \in \text{Supp}(\Pi_\infty) \subset \mathbb{R}^d \times \mathbb{R}^d \). Thus \( \Pi_\infty \) attributes a nonzero mass for all open ball of center \( x \). So let \( k \in \mathbb{N}^* \), we have \( \Pi_\infty(B(x,1/k)) > 0 \), where \( B(x,1/k) \) is the open ball centered at \( x \) and radius \( 1/k \). Thanks to the weakly convergence, we have:

\[ \liminf_{p \to +\infty} \Pi_0^{(p)}(B(x,1/k)) \geq \Pi_\infty(B(x,1/k)) > 0. \]

This inequality let us define an increasing sequence \( (j_k)_k \) such that: \( j_0 := 0 \) and for \( k > 0 \)

\[ j_k := \min\{ p \in \mathbb{N}, p > j_{k-1}, \forall n \geq p : \text{Supp}(\Pi_0^{(n)}) \cap B(x,1/k) \neq \emptyset \}. \]
The increase yields that for all \( k \in \mathbb{N} \), there exists \( j_k \leq p < j_{k+1} \) such that we can pick up a point \( x_p \in \text{Supp}(\Pi_0^{(p)}) \cap B(x, 1/k) \). The sequence \((x_p)_p\) converges to \( x \) and for all \( p \in \mathbb{N} \), we have \( x_p \in \text{Supp}(\Pi_0^{(p)}) \). □

Remark: In fact this proof (hence this lemma) is true for all sequence of measures converging weakly to another measure, on any metric space.

**Proposition 3.2** \( \Pi^\infty \) is optimal for the cost \( c^2 \).

Here we use the equivalence between optimality for \( c^2 \) and \( c^2 \)-cyclically monotonicity (see [?]) for continuous and real valued costs \( c \).

**Proof:**
It is then sufficient to prove that \( \text{Supp}(\Pi^\infty) \) is \( c \)-cyclically monotone. Let \( N \in \mathbb{N}^* \) and \((x_1,y_1), \ldots, (x_N,y_N) \in \text{Supp}(\Pi^\infty) \). Since \((\Pi_0^{(p)})_p\) converges weakly to \( \Pi^\infty \), we can apply the Lemma 3.1: for all \( i = 1, \ldots, N \), there exists \((x_{i}^p,y_{i}^p) \in \text{Supp}(\Pi_0^{(p)})\) such that \( \lim(x_{i}^p,y_{i}^p) = (x_i,y_i) \). Thus \((x_1^p,y_1^p), \ldots, (x_N^p,y_N^p) \in \text{Supp}(\Pi_0^{(p)})\) which is \( c^2 \)-cyclically monotone, since \( \Pi_0^{(p)} \) is optimal for the cost \( c_p \). Then the inequality

\[
\sum_{i=1}^{N} c_i^2(x_i^p,y_i^p) \leq \sum_{i=1}^{N} c_i^2(x_i^{p+1},y_i^{p+1})
\]

holds, with \( y_{N+1} := y_1 \). And it is elementary to check that the sets

\[
\bigcup_{p \geq 2} \{(x_1^p,y_1^p), \ldots, (x_N^p,y_N^p)\} \bigcup \{(x_1,y_1), \ldots, (x_N,y_N)\},
\]

are compact of \( \mathbb{R}^d \times \mathbb{R}^d \). But since \((c_p)_p\) converges uniformly on compact subsets of \( \mathbb{R}^d \times \mathbb{R}^d \) to \( c \), we get from (3.9), taking the limit with \( p \to +\infty \):

\[
\sum_{i=1}^{N} c^2(x_i,y_i) \leq \sum_{i=1}^{N} c^2(x_i,y_{i+1}).
\]

That is exactly the definition of \( c^2 \)-cyclically monotone for \( \text{Supp}(\Pi^\infty) \). □

Because of non strict convexity of \( |.|_\infty \), \((\mathbb{R}^d,|.|_\infty)\) is a branching space: there exists many constant speed geodesics between two probability measures. Finally one can conclude this section with the following result:

**Proposition 3.3** Relative convexity w.r.t. \( \gamma_d \) on \((\mathbb{R}^d,|.|_\infty)\) is weakly 1-convex along (constant speed) geodesics on \((\mathcal{P}_2(\mathbb{R}^d),W^2_{2,\infty})\).

**Proof:**
For \( p \in (2, +\infty) \) we consider optimal transfer plans \( \Pi_0^{(p)} \in \Gamma_0^{(p)}(\rho_0,\rho_1) \) and constant speed geodesics \( \rho_i^{(p)} \) associated. Applying Proposition 3.1 with \( |.|_p \) norms, we get

\[
\text{Ent}_{\gamma_d}(\rho_i^{(p)}) \leq (1-t)\text{Ent}_{\gamma_d}(\rho_0) + t\text{Ent}_{\gamma_d}(\rho_1) - \frac{t(1-t)}{2} W^2_{2,p}(\rho_0,\rho_1),
\]

(3.10)
for all \( p \geq 2 \). But for \( x \in \mathbb{R}^d \) and for \( p \geq 2 \), \(|x|_p \geq |x|_\infty\). Then
\[
\int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 d\Pi_0^{(p)}(x,y) \geq \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 d\Pi_0^{(p)}(x,y),
\]
\[
W^2_{2,p}(\rho_0, \rho_1) \geq \inf_{\Pi \in \Gamma(\mu, \mu_\gamma)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 d\Pi(x,y) = W^2_{2,\infty}(\rho_0, \rho_1),
\]
\[
\implies \liminf_p W^2_{2,p}(\rho_0, \rho_1) \geq W^2_{2,\infty}(\rho_0, \rho_1).
\]
Moreover, the sequel \((\Pi_0^{(p)})_p\) weakly converges to \(\Pi^\infty\) and this coupling is optimal for \(|\cdot|^2_\infty\) thanks to the previous proposition. Hence the sequel \((\rho_t^{(p)})_p\) weakly converges to \(\rho_t^\infty\) for all \( t \in [0,1] \). But since the relative entropy is lower semi-continuous, we have:
\[
\liminf_p \text{Ent}_{\gamma_d}(\rho_t^{(p)}) \geq \text{Ent}_{\gamma_d}(\rho_t^\infty).
\]
Finally, combining this two arguments, taking the liminf in the inequality (3.10) with respect to \( p \):
\[
\text{Ent}_{\gamma_d}(\rho_t^\infty) \leq (1-t)\text{Ent}_{\gamma_d}(\rho_0) + t\text{Ent}_{\gamma_d}(\rho_1) - \frac{t(1-t)}{2} W^2_{2,\infty}(\rho_0, \rho_1).
\]

### 3.2 Wiener space case

Consider the same notations of the previous settings and denote by \( \mathcal{A}_n \) the sub \( \sigma \)-field on \( W \) generated by \( \pi_n \). Let us recall the following results:

1. \((\pi_n)\#\mu = \gamma_n\) the standard Gaussian measure on \( V_n \).
2. For \( i = 0,1 \), \( \rho_i^n := (\pi_n)\#\rho_i \) is absolutely continuous with respect to \( \gamma_n \) with density \( f_i^n \) where \( f_i^n \circ \pi_n = \mathbb{E}[f_i|\mathcal{A}_n] = : f_i^n \).
3. For \( i = 0,1 \), \( f_i^n \rightharpoonup f_i \) in \( L^1(W, \mu) \).

**Theorem 3.1** Relative entropy w.r.t. \( \mu \) on \((W,|\cdot|_\infty)\) is weakly \( 1 \)-convex along (constant speed) geodesic on \((\mathcal{P}_2(W), W_{2,\infty})\).

**Proof:**

**Step 1.** Since \( \hat{\rho}_i^n = f_i^n \circ \pi_n \rightharpoonup f_i \), we have:
\[
\hat{\rho}_i^n := (f_i^n \circ \pi_n)\mu \rightharpoonup f_i \mu = \rho_i \quad i = 0,1.
\]

Then in [?], it is proved that for \( \hat{\Pi}_n^\infty \in \Gamma_0^\infty(\hat{\rho}_0^n, \hat{\rho}_1^n) \) there exists a subsequence of \((\hat{\Pi}_n^\infty)_n\) denoted by \((\hat{\Pi}_n^\infty)_k\) which weakly converges to \( \hat{\Pi}^\infty \) and we know this limit point belongs to \( \Gamma_0^\infty(\rho_0, \rho_1) \).
3.2 Wiener space case

Convexity of relative entropy on Wiener space

**Step 2.** \( W_{2,\infty}(\tilde{\rho}_0^n, \tilde{\rho}_1^n) \leq W_{2,\infty}(\rho_0^n, \rho_1^n) \)

We have the decomposition \( W = V_n \oplus V_{n}^\perp \). Then define a probability measure \( \Pi \in \Gamma(\hat{\rho}_0^n, \hat{\rho}_1^n) \) as

\[
\int_{W \times W} \varphi(x, y) d\Pi(x, y) := \int_{V_{n}^\perp} \int_{V_n \times V_n} \varphi(x_1 + z, x_2 + z) d\Pi_n(x_1, x_2) d\rho_0^n(z),
\]

for \( \varphi \) bounded continuous function, and \( \Pi_n \in \Gamma_0(\rho_0^n, \rho_1^n) \). So

\[
\int_{W \times W} |x - y|_\infty^2 d\Pi(x, y) = \int_{V_n \times V_n} |x_1 - x_2|_\infty^2 d\Pi_n(x_1, x_2) = W_{2,\infty}^2(\rho_0^n, \rho_1^n),
\]

\[
\implies W_{2,\infty}(\tilde{\rho}_0^n, \tilde{\rho}_1^n) \leq W_{2,\infty}^2(\rho_0^n, \rho_1^n).
\]

**Step 3.**

Now we have \( \Pi_{n_k}^\infty := (\pi_{n_k}, \pi_{n_k}) \# \hat{\Pi}_{n_k}^\infty \in \Gamma_0^\infty(\rho_0^{n_k}, \rho_1^{n_k}) \) and if we consider the associated constant speed geodesic \( \rho_t^{n_k} \), we can apply Proposition 3.3 to obtain:

\[
\text{Ent}_{\gamma(n_k)}(\rho_t^{n_k}) \leq (1 - t) \text{Ent}_{\gamma(n_k)}(\rho_0^{n_k}) + t \text{Ent}_{\gamma(n_k)}(\rho_1^{n_k}) - \frac{t(1 - t)}{2} W_{2,\infty}^2(\rho_0^{n_k}, \rho_1^{n_k}) \quad \forall t \in [0, 1].
\]

Let \( t \in [0, 1] \mapsto \tilde{\rho}_t^{n_k} := ((1 - t)P_{R_1} + tP_{R_2}) \# \hat{\Pi}_{n_k}^\infty \) be the constant speed geodesic associated to \( \hat{\Pi}_{n_k}^\infty \). We have:

\[
\text{Ent}_{\gamma(n_k)}(\rho_0^{n_k}) = \int_{V_{n_k}} f_0^{n_k}(x) \log(f_0^{n_k}(x)) d\gamma_{n_k}(x)
\]

\[
= \int_W f_0^{n_k}(\pi_{n_k}(y)) \log(f_0^{n_k}(\pi_{n_k}(y))) d\mu(y)
\]

\[
= \int_W \hat{\rho}_0^{n_k}(y) \log(f_0^{n_k}(y)) d\mu(y) = \text{Ent}_{\mu}(\hat{\rho}_0^{n_k}).
\]

And on the same way, we can easily prove this for \( \rho_1^{n_k} \) and \( \rho_t^{n_k} \) where \( t \in (0, 1) \). Thus (3.11) and Step 2. yield

\[
\text{Ent}_{\mu}(\hat{\rho}_t^{n_k}) \leq (1 - t) \text{Ent}_{\mu}(\rho_0^{n_k}) + t \text{Ent}_{\mu}(\rho_1^{n_k}) - \frac{t(1 - t)}{2} W_{2,\infty}^2(\rho_0^{n_k}, \rho_1^{n_k}) \quad \forall t \in [0, 1].
\]

**Step 4.**

Thanks to Step 1., we have:

\[
W_{2,\infty}^2(\rho_0, \rho_1) = \int_{W \times W} |x - y|_\infty^2 d\hat{\Pi}^\infty(x, y)
\]

\[
\leq \liminf_k \int_{W \times W} |x - y|_\infty^2 d\hat{\Pi}_{n_k}^\infty(x, y) = \liminf_k W_{2,\infty}^2(\rho_0^{n_k}, \rho_1^{n_k}).
\]

Hence for \( \epsilon > 0 \), there exists \( N \in \mathbb{N} \) such that:

\[
W_{2,\infty}^2(\rho_0^{n_k}, \rho_1^{n_k}) + \epsilon \geq W_{2,\infty}^2(\rho_0, \rho_1) \quad \forall k \geq N.
\]
Now fix \( k \geq N \). By the Jensen’s inequality, we have for \( i = 0, 1 \),
\[
\Ent_\mu(\tilde{\rho}_i^{n_k}) = \int_W \tilde{f}_i^{n_k}(y) \log(\tilde{f}_i^{n_k}(y)) d\mu(y) \\
= \int_W \mathbb{E}[f_i|A_{n_k}](y) \log(\mathbb{E}[f_i|A_{n_k}](y)) d\mu(y) \\
\leq \int_W \mathbb{E}[f_i(y)]d\mu(y)
\]
and then
\[
\Ent_\mu(\tilde{\rho}_i^{n_k}) \leq (1-t)\Ent_\mu(\tilde{\rho}_0^{n_k}) + t\Ent_\mu(\tilde{\rho}_1^{n_k}) - \left( \frac{(1-t)}{2} W_{2,\infty}(\rho_0, \rho_1) - \epsilon \right)
\]
\[
\leq (1-t)\Ent_\mu(\rho_0) + t\Ent_\mu(\rho_1) - \left( \frac{(1-t)}{2} W_{2,\infty}(\rho_0, \rho_1) - \epsilon \right).
\]
Moreover,
\[
\text{Since } \hat{\Pi}_n^{\infty} \xrightarrow{\text{weakly}} \hat{\Pi}^{\infty} \text{ which is optimal due to Lemma 3.1,}
\]
we have \( \tilde{\rho}_t^{n_k} \xrightarrow{\text{weakly}} \hat{\rho}_t := ((1-t) Pr_1 + t Pr_2)_# \hat{\Pi}^{\infty} \).

Denote by \( R \) the right hand side of the previous inequality, and by compactness of \( \{ \nu = \rho \mu \in \mathcal{P}_2(W) \mid \Ent_\mu(\nu) \leq R \} \) with respect to the weak topology, we have:
\[
\Ent_\mu(\hat{\rho}_t) \leq (1-t)\Ent_\mu(\rho_0) + t\Ent_\mu(\rho_1) - \left( \frac{(1-t)}{2} W_{2,\infty}(\rho_0, \rho_1) - \epsilon \right).
\]
Finally letting \( \epsilon \to 0 \) we get:
\[
\Ent_\mu(\hat{\rho}_t) \leq (1-t)\Ent_\mu(\rho_0) + t\Ent_\mu(\rho_1) - \frac{(1-t)}{2} W_{2,\infty}(\rho_0, \rho_1) \quad \forall t \in [0, 1].
\]

The same proof holds for the cost \( \| \cdot \|_{k,\gamma} \) by applying in Step 3. Proposition 3.1 with \( \| \cdot \|_{k,\gamma} \leq C_{k,\gamma} \| \cdot \|_2 \) instead of Proposition 3.3. Even better, since the latter norm is strictly convex, \((W, \| \cdot \|_{k,\gamma})\) is a non-branching space. Hence we have the

**Theorem 3.2** Relative entropy w.r.t. \( \mu \) on \((W, \| \cdot \|_{k,\gamma})\) is \(1/C_{k,\gamma}^2\)-convex along (constant speed) geodesics on \((\mathcal{P}_2(W), W_{2,(k,\gamma)})\).

- In Step 2, first proof we have in fact equality \( W_{2,\infty}(\tilde{\rho}_0^{n_k}, \tilde{\rho}_1^{n_k}) = W_{2,\infty}(\rho_0^{n_k}, \rho_1^{n_k}) \).
  This is provided by the fact that \( |\pi_n(w)|_{\infty} \leq |w|_{\infty} \) for all \( w \in W \). Indeed, we know that \( \hat{\Pi}_n^{\infty} \) is optimal. Define \( \Pi_n := (\pi_n, \pi_n)_# \hat{\Pi}_n^{\infty} \in \mathcal{P}(V_n \times V_n) \). Since the
norm on $V_n$ is less than the norm on $W$, we have:

$$
\int_{V_n \times V_n} |x - y|^2 d\Pi_n(x,y) = \int_{W \times W} |\pi_n(x) - \pi_n(y)|^2 d\Pi_n^\infty(x,y)
\leq \int_{W \times W} |x - y|^2 d\Pi_n(x,y) = W^2_{2,\infty}(\rho_0^n, \rho_1^n),
\Rightarrow W^2_{2,\infty}(\rho_0^n, \rho_1^n) \leq W^2_{2,\infty}(\hat{\rho}_0^n, \hat{\rho}_1^n).
$$

- In the second norm, it is still not clear if $\|\pi_n(w)\|_{k,\gamma} \leq \|w\|_{k,\gamma}$ for any $w \in W$.

### Appendices

**Theorem 3.3** Rademacher’s Theorem. If $\varphi \in L^p(W)$ is a $\|\cdot\|_{k,\gamma}$-convex map then $\varphi$ belongs to $D^p(W)$.

**Definition 2.4** We say that a function $f : W \rightarrow W$ is locally $\|\cdot\|_{k,\gamma}$-Lipschitz if for all $R > 0$, there exists $C_R > 0$ such that:

$$
|f(w) - f(w')| \leq C_R \|w - w'\|_{k,\gamma} \quad \forall w, w' \in B_{k,\gamma}(R) := \{w \in W; \|w\|_{k,\gamma} \leq R\}.
$$

**Proposition 4.4** Let $(f_i)_{i \in I}$ be a uniform family of locally $\|\cdot\|_{k,\gamma}$-Lipschitz real-valued functions defined on an open subset $U$ of $W$. If the function

$$
f(x) := \sup_{i \in I} f_i(x)
$$

is finite everywhere then $f : U \rightarrow \mathbb{R}$ is also locally $\|\cdot\|_{k,\gamma}$-Lipschitz.

**Proof:**

Let $R > 0$ such that $B_{k,\gamma}(R) \subset U$. There exists $C_R > 0$ (not depending on $i$) such that for all $w, w' \in B_{k,\gamma}(R)$,

$$
|f_i(w) - f_i(w')| \leq C_R \|w - w'\|_{k,\gamma} \quad \forall i \in I.
$$

For each $w \in B_{k,\gamma}(R)$ by definition of $f$ there is a sequence $(i_n)_n$ (depending on $w$) such that $\lim_n f_{i_n}(w) = f(w)$. Moreover $f_{i_n}(w') \leq f(w')$ for all $w' \in B_{k,\gamma}(R)$ then

$$
f_{i_n}(w) - f(w') \leq f_{i_n}(w) - f_{i_n}(w') \leq C_R \|w - w'\|_{k,\gamma}.
$$

Passing to the limit in the previous inequality, we get:

$$
f(w) - f(w') \leq C_R \|w - w'\|_{k,\gamma}.
$$

Exchanging $w$ and $w'$, this yields $f$ is locally $\|\cdot\|_{k,\gamma}$-Lipschitz.

**Acknowledgements.** I would like to thank Professor Shizan Fang, for his suggestion, and so Nicolas Juillet for useful discussions about general facts of optimal transportation.