Monge-Kantorovitch Measure Transportation and Monge-Ampère Equation on Wiener Space

D. Feyel and A. S. Üstünel

Abstract: Let (W, μ, H) be an abstract Wiener space assume two ν_i, $i = 1, 2$ probabilities on $(W, \mathcal{B}(W))$ [1]. We give some conditions for the Wasserstein distance between ν_1 and ν_2 with respect to the Cameron-Martin space

$$d_H(\nu_1, \nu_2) = \sqrt{\inf_\beta \int_{W\times W} |x-y|_H^2 d\beta(x,y)}$$

to be finite, where the infimum is taken on the set of probability measures β on $W \times W$ whose first and second marginals are ν_1 and ν_2. In this case we prove the existence of a unique (cyclically monotone) map $T = I_W + \xi$, with $\xi : W \to H$, such that T maps ν_1 to ν_2. Moreover, if $\nu_2 \ll \mu$ [2], then T is stochastically invertible, i.e., there exists $S : W \to W$ such that $S \circ T = I_W$ ν_1 a.s. and $T \circ S = I_W$ ν_2 a.s. If, in addition, $\nu_1 = \mu$, then there exists a 1-convex function ϕ in the Gaussian Sobolev space $\mathbb{D}_{2,1}$, such that $\xi = \nabla\phi$. These results imply that the quasi-invariant transformations of the Wiener space with finite Wasserstein distance from μ can be written as the composition of a transport map T and a rotation, i.e., a measure preserving map. We give also 1-convex sub-solutions and Ito-type solutions of the Monge-Ampère equation on W. [3]

1 Introduction

In 1781, Gaspard Monge has published his celebrated memoire about the most economical way of earth-moving [23]. The configurations of excavated earth and remblai were modelized as two measures of equal mass, say ρ and ν, that Monge had supposed absolutely continuous with respect to the volume measure. Later Ampère has studied an analogous question about the electricity current in a media with varying conductivity. In modern language of measure theory we can express the problem in the following terms: let W be a Polish space on which are given two positive measures ρ and ν, of finite, equal mass. Let $c(x, y)$ be a cost function on $W \times W$, which is, usually, assumed

[1] cf. Theorem 6.1 for the precise hypothesis about ν_1 and ν_2.

[2] In fact this hypothesis is too strong, cf. Theorem 6.1.

[3] Mathematics Subject Classification (2000): 60H07, 60H05,60H25, 60G15, 60G30, 60G35, 46G12, 47H05, 47H1, 35J60, 35B65,35A30, 46N10, 49Q20, 58E12, 26A16, 28C20.

positive. Does there exist a map $T : W \to W$ such that $T\rho = \nu$ and T minimizes the integral

$$\int_W c(x, T(x)) d\rho(x)$$

between all such maps? The problem has been further studied by Appell [3, 4] and by Kantorovitch [19]. Kantarovitch has succeeded to transform this highly nonlinear problem of Monge into a linear problem by replacing the search for T with the search of a measure γ on $W \times W$ with marginals ρ and ν such that the integral

$$\int_{W\times W} c(x, y) d\gamma(x, y)$$

is the minimum of all the integrals

$$\int_{W\times W} c(x, y) d\beta(x, y)$$

where β runs over the set of measures on $W \times W$ whose marginals are ρ and ν. Since then the problem adressed above is called the Monge problem and the quest of the optimal measure is called the Monge-Kantorovitch problem.

Historically, the problem of Monge was stated with $W = \mathbb{R}^m$ and with the Euclidean cost $c(x, y) = |x - y|$. The case of the quadratic cost $c(x, y) = |x - y|^2$ has come later. Even in this case, it is by no means obvious that the Monge problem is equivalent to the Monge-Kantorovitch problem. Nevertheless, suppose that the latter has a solution γ whose support lies in the graph of a map T, then this graph yields also a solution of the Monge problem. Many authors have attempted to find such a measure (cf. [6, 19, 21, 26]). Recently Brenier [6] and McCann [21] have proved the existence and the uniqueness of it when ρ is absolutely continuous with respect to the Lebesgue measure. They have also shown that this measure was supported by the graph of a mapping, so that the Monge problem was also solved.

In this paper we study the Monge-Kantorovitch and the Monge problems in the frame of an abstract Wiener space with a singular quadratic cost. In other words, let W be a separable Fréchet space with its Borel sigma algebra $\mathcal{B}(W)$ and assume that there is a separable Hilbert space H which is injected densely and continuously into W, hence in general the topology of H is stronger than the topology induced by W. The cost function $c : W \times W \to \mathbb{R}_+ \cup \{\infty\}$ is defined as

$$c(x, y) = |x - y|_H^2 ,$$

we suppose that $c(x, y) = \infty$ if $x - y$ does not belong to H. Clearly, this choice of the function c is not arbitrary, in fact it is closely related to Ito Calculus, hence also to the problems originating from Physics, quantum chemistry, large deviations, etc. Since for all the interesting measures on W, the Cameron-Martin space is a negligeable set, the cost function will be infinity very frequently. Let $\Sigma(\rho, \nu)$ denote the set of probability measures on $W \times W$ with given marginals ρ and ν. It is a convex, compact set under the weak topology $\sigma(\Sigma, C_b(W \times W))$. As explained above, the problem of Monge consists of finding a measurable map $T : W \to W$, called the optimal transport of ρ to ν, i.e., $T\rho = \nu$ [4] which minimizes the cost

$$U \to \int_W |x - U(x)|_H^2 d\rho(x)\,,$$

among all the maps $U : W \to W$ such that $U\rho = \nu$. The Monge-Kantorovitch problem will consist of finding a measure on $W \times W$, which minimizes the function $\theta \to J(\theta)$, defined by

$$J(\theta) = \int_{W\times W} \left|x - y\right|_H^2 d\theta(x,y)\,, \tag{1.1}$$

where θ runs over $\Sigma(\rho,\nu)$. Note that $\inf\{J(\theta) :\ \theta \in \Sigma(\rho,\nu)\}$ is the square of Wasserstein metric $d_H(\rho,\nu)$ with respect to the Cameron-Martin space H.

Any solution γ of the Monge-Kantorovitch problem will give a solution to the Monge problem provided that its support is included in the graph of a map. Hence our work consists of realizing this program. Although in the finite dimensional case this problem is well-studied in the path-breaking papers of Brenier [6] and McCann [21, 22] things do not come up easily in our setting and the difficulty is due to the fact that the cost function is not continuous with respect to the Fréchet topology of W, for instance the weak convergence of the probability measures does not imply the convergence of the integrals of the cost function. In other words the function $|x-y|_H^2$ takes the value plus infinity "very often". On the other hand the results we obtain seem to have important applications to several problems of stochastic analysis.

The most notable results of this paper can be stated as follows: assume that $d_H(\rho,\nu) < \infty$ and that along a sequence of finite dimensional projections increasing to the identity of H and having continuous extensions to W, the measure ρ disintegrates "nicely". Then the problem of Monge-Kantorovitch and the problem of Monge have unique solutions for the singular cost function $c(x,y) = |x-y|_H^2$, the solution of the Monge problem is of the form $T = I_W + \xi$, where ξ is an H-valued random variable, such that the map T is cylindrically cylically monotone. If the target measure ν has also a similar property, then the solution of the Monge problem is "almost surely invertible" and its inverse is the solution of the reverse problem of Monge. If, furthermore, ρ is equal to the Wiener measure and if ν is absolutely continuous with respect to ρ, then ξ above is of the form $\nabla\phi$, where ϕ is in the Gaussian Sobolev space $\mathbb{D}_{2,1}$ and it is a 1-convex function [15]. Finally, we give the strong sub-solutions and the Ito-type solutions of the Monge-Ampère equation. We shall explain the applications of these results while enumerating the contents of the paper.

In Section 2, we explain some basic results about the functional analysis constructed on the Wiener space (cf., for instance [14, 30]) and the probabilistic theory of convex functions recently developed in [15]. For the paper to be more self-contained, we have also added the most relevant results about the finite dimensional problems of Monge and of Monge-Kantorovitch. Section 3 is devoted to the derivation of some inequalities which control the Wasserstein distance. In particular, with the help of the Girsanov theorem, we give a very simple proof of an inequality, initially discovered by Talagrand

[4]We denote the push-forward of ρ by T, i.e., the image of ρ under T, by $T\rho$.

([27]); the ease to prove this result gives already an idea about the efficiency of the infinite dimensional techniques for the Monge-Kantorovitch problem[5]. We indicate some simple consequences of this inequality to control the measure of subsets of the Wiener space in terms of the second moment of their gauge functionals defined with respect to the Cameron-Martin distance. These inequalities are quite useful in the theory of large deviations. Using a different representation of the target measure, namely by constructing a flow of diffeomorphisms of the Wiener space (cf. Chapter V of [31]) which maps the Wiener measure to the target measure, we obtain also a new control of the Kantorovitch-Rubinstein metric of order one. The method we employ for this inequality generalizes directly to a more general class of measures, namely those for which one can define a reasonable divergence operator.

In Section 4, we solve directly the original problem of Monge when the first measure is the Wiener measure and the second one is given with a density, in such a way that the Wasserstein distance between these two measures is finite. We prove the existence and the uniqueness of a transformation of W of the form $T = I_W + \nabla\phi$, where ϕ is a 1-convex function in the Gaussian Sobolev space $\mathbb{D}_{2,1}$ such that the measure $\gamma = (I_W \times T)\mu$ is the unique solution of the problem of Monge-Kantorovitch. This result gives a new insight to the question of representing an integrable, positive random variable whose expectation is unity, as the Radon-Nikodym derivative of the image of the Wiener measure under a map which is a perturbation of identity, a problem which has been studied by X. Fernique and by one of us with M. Zakai (cf., [12, 13, 31]). In [31], Chapter II, it is shown that such random variables are dense in $\mathbb{L}^1_{1,+}(\mu)$ (the lower index 1 means that the expectations are equal to one), here we prove that this set of random variables contains the elements L of $\mathbb{L}^1_{1,+}(\mu)$ such that the Wasserstein distance between the measure $Ld\mu$ and the Wiener measure $d\mu$ is finite. In fact even if this distance is infinite, we show that there is a solution to this problem if we enlarge W slightly by taking $\mathbb{N} \times W$.

Section 5 is devoted to the immediate implications of the existence and the uniqueness of the solutions of Monge-Kantorovitch and Monge problems constructed in Section 4. Indeed the uniqueness implies at once that the absolutely continuous transformations of the Wiener space, at finite (Wasserstein) distance, have a unique decomposition in the sense that they can be written as the composition of a measure preserving map in the form of a perturbation of identity with another one which is the perturbation of identity by the Sobolev derivative of a 1-convex function. This means in particular that the class of 1-convex functions is as basic as the class of adapted processes in the setting of Wiener space.

In Section 6 we prove the existence and uniqueness of solutions of the Monge-Kantorovitch and Monge problems for general measures which are at finite Wasserstein distance from each other. The fundamental hypothesis we use is that the regular conditional probabilities which are obtained by the disintegration of one of the measures along the orthogonals of a sequence of regular, finite dimensional projections vanish on the sets of co-dimension one. In particular, this hypothesis is satisfied if the measure under question is absolutely continuous with respect to the Wiener measure. The

[5]In Section 7 we shall see another illustration of this phenomena.

method we use in this section is totally different from the one of Section 4; it is based on the notion of cyclic monotonicity of the supports of the regular conditional probabilities obtained through some specific disintegrations of the optimal measures. The importance of cyclic monotonicity has first been remarked by McCann and used abundantly in [21] and in [17] for the finite dimensional case. Here the things are much more complicated due to the singularity of the cost function, in particular, contrary to the finite dimensional case, the cyclic monotonicity is not compatible with the weak convergence of probability measures. A curious reader may ask why we did not treat first the general case and then attack the subject of Section 4. The answer is twofold: even if we had done so, we would have needed similar calculations as in Section 4 in order to show the Sobolev regularity of the transport map, hence concerning the volume, the order that we have chosen does not change anything. Secondly, the construction used in Section 4 has an interest of its own since it explains interesting relations between the transport map and its inverse and the optimal measure in a more concrete situation, in this sense this construction is rather complementary to the material of Section 6.

Section 7 studies the Monge-Ampère equation for measures which are absolutely continuous with respect to the Wiener measure. First we briefly indicate the notion of second order Alexandroff derivative and the Alexandroff version of the Ornstein-Uhlenbeck operator applied to a 1-convex function in the finite dimensional case. With the help of these observations, we write the corresponding Jacobian using the modified Carleman-Fredholm determinant which is natural in the infinite dimensional case (cf., [31]). Afterwards we attack the infinite dimensional case by proving that the absolutely continuous part of the Ornstein-Uhlenbeck operator applied to the finite rank conditional expectations of the transport function is a submartingale which converges almost surely. Hence the only difficulty lies in the calculation of the limit of the Carleman-Fredholm determinants. Here we have a major difficulty which originates from the pathology of the Radon-Nikodym derivatives of vector measures with respect to a scalar measure as explained in [28]: in fact even if the second order Sobolev derivative of a Wiener function is a vector measure with values in the space of Hilbert-Schmidt operators, its absolutely continuous part has no reason to take values in the space of Hilbert-Schmidt operators. Hence the Carleman-Fredholm determinant may not exist, however due to the 1-convexity, all the determinants of the approximating sequence take values in the interval $[0, 1]$. Consequently we can construct the subsolutions with the help of Fatou's lemma.

Last but not least, in section 7.1, we prove that all these difficulties can be overcome thanks to the natural renormalization of the Ito stochastic calculus. In fact using the Ito representation theorem and the Wiener space analysis extended to the distributions, cf. [29], we can give the explicit solution of the Monge-Ampère equation. This is a remarkable result in the sense that such techniques do not exist in the finite dimensional case.

2 Preliminaries and notations

Let W be a separable Fréchet space equipped with a Gaussian measure μ of zero mean whose support is the whole space. The corresponding Cameron-Martin space is denoted by H. Recall that the injection $H \hookrightarrow W$ is compact and its adjoint is the natural injection $W^\star \hookrightarrow H^\star \subset L^2(\mu)$. The triple (W, μ, H) is called an abstract Wiener space. Recall that $W = H$ if and only if W is finite dimensional. A subspace F of H is called regular if the corresponding orthogonal projection π_F has a continuous extension to W, denoted again by the same letter. It is well-known that there exists an increasing sequence of regular subspaces $(F_n, n \geq 1)$, called total, such that $\cup_n F_n$ is dense in H and in W. Let $V_n = \sigma(\pi_{F_n})$ be the σ-algebra generated by π_{F_n} [6], then for any $f \in L^p(\mu)$, the martingale sequence $(E[f|V_n], n \geq 1)$ converges to f (strongly if $p < \infty$) in $L^p(\mu)$. Observe that the function $f_n = E[f|V_n]$ can be identified with a function on the finite dimensional abstract Wiener space (F_n, μ_n, F_n), where $\mu_n = \pi_n \mu$. A typical example for $(F_n, n \geq 1)$ can be constructed with the help of a sequence $(e_n, n \geq 1) \subset W^\star$ (i.e., the continuous dual of W), whose image in H forms a complete, orthonormal basis of H (cf.[18]). In this case F_n is the subspace generated by $\{e_1, \ldots, e_n\}$.

Since the translations of μ by elements of H induce measures equivalent to μ, the Gâteaux derivative in H direction of the random variables is a closable operator on $L^p(\mu)$-spaces and this closure will be denoted by ∇ cf., for example [14, 30]. The corresponding Sobolev spaces (the equivalence classes) of the real random variables will be denoted as $\mathbb{D}_{p,k}$, where $k \in \mathbb{N}$ is the order of differentiability and $p > 1$ is the order of integrability. If the random variables take values in some separable Hilbert space, say Φ, then we shall define similarly the corresponding Sobolev spaces and they are denoted $\mathbb{D}_{p,k}(\Phi)$, $p > 1$, $k \in \mathbb{N}$. Since $\nabla : \mathbb{D}_{p,k} \to \mathbb{D}_{p,k-1}(H)$ is a continuous and linear operator its adjoint is a well-defined operator which we represent by δ. In the case of classical Wiener space, i.e., when $W = C_0(\mathbb{R}_+, \mathbb{R}^d)$, then δ coincides with the Ito integral of the Lebesgue density of the adapted elements of $\mathbb{D}_{p,k}(H)$ (cf.[30]).

For any $t \geq 0$ and measurable $f : W \to \mathbb{R}_+$, we note by

$$P_t f(x) = \int_W f\left(e^{-t}x + \sqrt{1 - e^{-2t}}y\right) \mu(dy),$$

it is well-known that $(P_t, t \in \mathbb{R}_+)$ is a hypercontractive semigroup on $L^p(\mu), p > 1$, which is called the Ornstein-Uhlenbeck semigroup (cf.[14, 30]). Its infinitesimal generator is denoted by $-\mathcal{L}$ and we call $\mathcal{L}$ the Ornstein-Uhlenbeck operator (sometimes called the number operator by the physicists). The norms defined by

$$\|\phi\|_{p,k} = \|(I + \mathcal{L})^{k/2}\phi\|_{L^p(\mu)} \tag{2.2}$$

are equivalent to the norms defined by the iterates of the Sobolev derivative ∇. This observation permits us to identify the duals of the space $\mathbb{D}_{p,k}(\Phi); p > 1$, $k \in \mathbb{N}$ by $\mathbb{D}_{q,-k}(\Phi')$, with $q^{-1} = 1 - p^{-1}$, where the latter space is defined by replacing k in (2.2) with $-k$, this gives us the distribution spaces on the Wiener space W (in fact we can take as k any real number). An easy calculation shows that, formally, $\delta \circ \nabla =$

[6] For typographical reasons, in the sequel we shall denote π_{F_n} by π_n.

$\mathcal{L}$, and this permits us to extend the divergence and the derivative operators to the distributions as linear, continuous operators. In fact $\delta : \mathbb{D}_{q,k}(H \otimes \Phi) \to \mathbb{D}_{q,k-1}(\Phi)$ and $\nabla : \mathbb{D}_{q,k}(\Phi) \to \mathbb{D}_{q,k-1}(H \otimes \Phi)$ continuously, for any $q > 1$ and $k \in \mathbb{R}$, where $H \otimes \Phi$ denotes the completed Hilbert-Schmidt tensor product (cf., for instance [30]).

Let us recall some facts from convex analysis. Given the Hilbert space K, a subset S of $K \times K$ is called cyclically monotone if any finite subset $\{(x_1, y_1), \ldots, (x_N, y_N)\}$ of S satisfies the following algebraic condition:

$$\langle y_1, x_2 - x_1 \rangle + \langle y_2, x_3 - x_2 \rangle + \cdots + \langle y_{N-1}, x_N - x_{N-1} \rangle + \langle y_N, x_1 - x_N \rangle \leq 0 \,,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product of K. It turns out that S is cyclically monotone if and only if for all $N \geq 2$

$$\sum_{i=1}^{N} \langle y_i, x_{\sigma(i)} - x_i \rangle \leq 0 \,,$$

for any permutation σ of $\{1, \ldots, N\}$ and for any finite subset $\{(x_i, y_i) : i = 1, \ldots, N\}$ of S. Note that S is cyclically monotone if and only if any translate of it is cyclically monotone. By a theorem of Rockafellar, any cyclically monotone set is contained in the graph of the subdifferential of a convex function in the sense of convex analysis ([25]) and even if the function may not be unique its subdifferential is unique. We shall also call a mapping $F : K \to K$ cyclically monotone if

$$\sum_{i=1}^{N} \langle F(x_i), x_{\sigma(i)} - x_i \rangle \leq 0 \,,$$

for any permutation σ of $\{1, \ldots, N\}$ and for any finite subset $\{x_i : i = 1, \ldots, N\}$ of K. Note that, this condition, for $N = 2$ implies in particular that the mapping F is monotone.

Let now (W, μ, H) be an abstract Wiener space; a measurable function $f : W \to \mathbb{R} \cup \{\infty\}$ is called 1-convex if the map

$$h \to f(x + h) + \frac{1}{2}|h|_H^2 = F(x, h)$$

is convex on the Cameron-Martin space H with values in $L^0(\mu)$. Note that this notion is compatible with the μ-equivalence classes of random variables thanks to the Cameron-Martin theorem. It is proven in [15] that this definition is equivalent the following condition: Let $(\pi_n, n \geq 1)$ be a sequence of regular, finite dimensional, orthogonal projections of H, increasing to the identity map I_H. Denote also by π_n its continuous extension to W and define $\pi_n^\perp = I_W - \pi_n$. For $x \in W$, let $x_n = \pi_n x$ and $x_n^\perp = \pi_n^\perp x$. Then f is 1-convex if and only if

$$x_n \to \frac{1}{2}|x_n|_H^2 + f(x_n + x_n^\perp)$$

is $\pi_n^\perp \mu$-almost surely convex for every $n \geq 1$.

2.1 Monge and Monge-Kantorovitch problems in finite dimension: A concise survey

For the sake of simplicity, we shall represent respectively "le déblai" and " le remblai" with the measures having positive densities with respect to the Lebesgue measure of $\mathbb{R}^m$: $\rho(dx) = f(x)dx$ and $\nu(dy) = g(y)dy$. It is assumed that $\rho(\mathbb{R}^m) = \nu(\mathbb{R}^m) = M$ and in general the constant M is assumed to be one. Let $c(x, y)$ be the cost function defined on $\mathbb{R}^m \times \mathbb{R}^m$, which is in general positive. In this situation the problem of Monge is to find a mapping $T : \mathbb{R}^m \to \mathbb{R}^m$ which transports "le déblai" ρ to the "le remblai" ν and which minimizes the overall cost

$$\begin{aligned} \inf_U \int_{\mathbb{R}^m} c(x, U(x))\rho(dx) &= \int_{\mathbb{R}^m} c(x, T(x))\rho(dx) \\ &= \int_{\mathbb{R}^m} c(x, T(x))f(x)dx\,, \end{aligned}$$

where the infimum is taken between all the measurable maps U of $\mathbb{R}^m$ such that $U\rho = \nu$. The transport or the Monge-Ampère equation is the following identity, which is a consequence of the Jacobi formula:

$$f(x) = g \circ T(x)|J_T(x)|\,,$$

where J_T is the Jacobian of T provided that the derivative of T has a meaning. In this latter case T is said to be a (strong) solution of the Monge-Ampère equation for (f, g). We shall explain only the case where $c(x, y) = |x-y|^2$, where $|\cdot|$ denotes the Euclidean norm of $\mathbb{R}^m$.

Let $\Sigma(\rho, \nu)$ be the set of probability measures on $\mathbb{R}^m \times \mathbb{R}^m$, whose first and second marginals are respectively ρ and ν. Clearly $\Sigma(\rho, \nu)$ is a convex, compact set in the weak topology of measures. For $\theta \in \Sigma(\rho, \nu)$, we define

$$J(\theta) = \int_{\mathbb{R}^m \times \mathbb{R}^m} |x - y|^2 \, \theta(dx, dy) \leq \infty\,.$$

The problem of Monge-Kantorovitch (PMK) consists of finding a measure $\gamma \in \Sigma(\rho, \nu)$ which minimizes J with $J(\gamma) < \infty$. In case $J(\gamma) = \infty$, the problem is called degenerate. Evidently the PMK is not equivalent to the problem of Monge, however each solution of PMK, whose support is contained in the graph of a mapping T will provide a solution of the problem of Monge. Rachev [24] has proven that under some specific conditions, PMK has a unique solution. Sudakov [26] had already discovered that for certain cost functions the solution of PMK was supported by graphs of certain mappings $T : \mathbb{R}^m \to \mathbb{R}^m$ which are automatically the solutions of the Monge problem. The following results are important for further understanding of the subject:

Proposition 2.1 (McCann) *Suppose that $\gamma \in \Sigma(\rho, \nu)$ has its support in the subgradient $\partial F = \{(x, y) \in \mathbb{R}^m \times \mathbb{R}^m : y \in \partial F(x)\}$ of some convex function F. If ρ vanishes on the sets of Hausdorff dimension $(m-1)$, then F is proper, $\partial F(x)$ is ρ-almost surely univalent and $\partial F(x) = \nabla F(x)$ ρ-a.s. Moreover we have*

$$\gamma = (I \times \nabla F)\rho\,.$$

Finally, such a convex function F is unique upto an additive constant, hence ∇F is unique.

Let us remark that in this proposition we may have $J(\gamma)=\infty$. In the finite dimensional case, since the quadratic cost function is continuous, to construct such measures γ is easy as it can be seen from the next statement:

Proposition 2.2 (McCann) *Assume that $(\gamma_k, k\geq 1)\subset \mathcal{M}_1(\mathbb{R}^m\times\mathbb{R}^m)$ have cyclically monotone supports and that the sequence converges weakly to a probability γ. Then the support of γ is also cyclically monotone. In particular, for any two probability measures ρ and ν, on $\mathbb{R}^m$ there exists a $\gamma\in\Sigma(\rho,\nu)$ with cyclically monotone support.*

In the case γ is optimal with $J(\gamma)<\infty$, we also have (cf. [1, 17]):

Proposition 2.3 *If $\gamma\in\Sigma(\rho,\nu)$ is optimal (with $J(\gamma)<\infty$), then the support of γ is cyclically monotone.*

Hence combining Propositions 2.1, 2.2 and 2.3 and using the strict convexity and the linearity of the cost function, we obtain at once the existence and the uniquness of the solutions of PMK and of the problem of Monge. Let us give a final result, whose proof is rather immediate, which we shall use in the sequel:

Proposition 2.4 *Assume that the measure ρ satisfies the hypothesis of Proposition 2.1 and let F be the proper, convex function constructed through the propositions above, denote by $F^\star$ its Legendre-Fenchel transformation (i.e.,its convex conjugate) and let $\gamma\in\Sigma(\rho,\nu)$ be the solution of PMK which is given by $\gamma=(I\times\nabla F)\rho$. Then we have*

$$F(x)+F^\star(y)\geq (x,y)_{\mathbb{R}^m}$$

for any $x,y\in\mathbb{R}^m$ and

$$F(x)+F^\star(y)=(x,y)_{\mathbb{R}^m}$$

γ-almost surely.

3 Some Inequalities

Definition 3.1 *Assume that ξ and η are two probability measures on $(W,\mathcal{B}(W))$. We shall denote by $\Sigma(\xi,\eta)$ the set of all the probability measures β on $W\times W$ such that $\pi_1\beta=\xi$ and $\pi_2\beta=\eta$, where π_i, $i=1,2$ denote the projections of $W\times W$ onto W.*

Definition 3.2 *Let ξ and η be two probabilities on $(W,\mathcal{B}(W))$. We say that a probability $\gamma\in\Sigma(\xi,\eta)$ is a solution of the Monge-Kantorovitch problem associated to the couple (ξ,η), if*

$$J(\gamma)=\int_{W\times W}|x-y|_H^2 d\gamma(x,y)=\inf\left\{\int_{W\times W}|x-y|_H^2 d\beta(x,y):\,\beta\in\Sigma(\xi,\eta)\right\}.$$

We shall denote the Wasserstein distance between ξ and η, which is the positive squareroot of this infimum, by $d_H(\xi,\eta)$.

Remark: Since the set of probability measures on $W\times W$ is weakly compact and since the integrand in the definition is lower semi-continuous and strictly convex, the infimum in the definition is always attained even if the functional J is identically infinity.
The following result, which has already been published in [16] (cf. also [10]) is an extension of an inequality due to Talagrand [27] and it gives a sufficient condition for the Wasserstein distance to be finite:

Theorem 3.1 *Let $L\in \mathbb{L}\log\mathbb{L}(\mu)$ be a positive random variable with $E[L]=1$ and let ν be the measure $d\nu=Ld\mu$. We then have*

$$d_H^2(\nu,\mu)\leq 2E[L\log L]\,. \tag{3.3}$$

Proof: Without loss of generality, we may suppose that W is equipped with a filtration of sigma algebras in such a way that it becomes a classical Wiener space as $W = C_0(\mathbb{R}_+,\mathbb{R}^d)$ (cf. [31], Chapter 2.6). Let $(W_t, t\geq 0)$ be the associated canonical Wiener process. Assume first that L is a strictly positive and bounded random variable. We can represent it as

$$L=\exp\left[-\int_0^\infty(\dot{u}_s,dW_s)-\frac{1}{2}|u|_H^2\right]\,,$$

where $u=\int_0^\cdot \dot{u}_s ds$ is an H-valued, adapted random variable. Define τ_n as

$$\tau_n(x)=\inf\left\{t\in\mathbb{R}_+:\int_0^t|\dot{u}_s(x)|^2ds>n\right\}\,.$$

τ_n is a stopping time with respect to the canonical filtration $(\mathcal{F}_t, t\in\mathbb{R}_+)$ of the Wiener process $(W_t, t\in\mathbb{R}_+)$ and $\lim_n\tau_n=\infty$ almost surely. Define u^n as

$$u^n(t,x)=\int_0^t \mathbf{1}_{[0,\tau_n(x)]}(s)\dot{u}_s(x)ds\,.$$

Let $U_n: W\to W$ be the map $U_n(x)=x+u^n(x)$, then the Girsanov theorem says that $(t,x)\to U_n(x)(t)=x(t)+\int_0^t\dot{u}^n_s ds$ is a Wiener process under the measure $L_n d\mu$, where $L_n=E[L|\mathcal{F}_{\tau_n}]$. Therefore

$$\begin{aligned}
E[L_n\log L_n] &= E\left[L_n\left\{-\int_0^\infty(\dot{u}^n_s,dW_s)-\frac{1}{2}|u^n|_H^2\right\}\right]\\
&= \frac{1}{2}E[L_n|u^n|_H^2]\\
&= \frac{1}{2}E[L|u^n|_H^2]\,.
\end{aligned}$$

Define now the measure β_n on $W\times W$ as

$$\int_{W\times W}f(x,y)d\beta_n(x,y)=\int_W f(U_n(x),x)L_n(x)d\mu(x)\,.$$

Then the first marginal of β_n is μ and the second one is $L_n.\mu$. Consequently

$$\begin{aligned}
&\inf\left\{\int_{W\times W}|x-y|_H^2 d\theta : \pi_1\theta=\mu,\ \pi_2\theta=L_n.\mu\right\}\\
&\leq \int_W |U_n(x)-x|_H^2 L_n d\mu\\
&= 2E[L_n\log L_n]\,.
\end{aligned}$$

Hence we obtain

$$d_H^2(L_n.\mu,\mu)=J(\gamma_n)\leq 2E[L_n\log L_n]\,,$$

where γ_n is a solution of the Monge-Kantorovitch problem in $\Sigma(L_n.\mu,\mu)$. Let now γ be any cluster point of the sequence $(\gamma_n, n\geq 1)$, since $\gamma\to J(\gamma)$ is lower semi-continuous with respect to the weak topology of probability measures, we have

$$\begin{aligned}
J(\gamma) &\leq \liminf_n J(\gamma_n)\\
&\leq \sup_n 2E[L_n\log L_n]\\
&\leq 2E[L\log L]\,,
\end{aligned}$$

since $\gamma\in\Sigma(L.\mu,\mu)$, it follows that

$$d_H^2(L.\mu,\mu)\leq 2E[L\log L]\,.$$

For the general case we stop the martingale $E[L|\mathcal{F}_t]$ appropriately to obtain a bounded density L_n, then replace it by $P_{1/n}L_n$ to obtain the strict positivity, where $(P_t, t\geq 0)$ denotes the Ornstein-Uhlenbeck semigroup. From Jensen's inequality,

$$E[P_{1/n}L_n\log P_{1/n}L_n]\leq E[L\log L]\,,$$

therefore, using the same reasoning as above

$$\begin{aligned}
d_H^2(L.\mu,\mu) &\leq \liminf_n d_H^2(P_{1/n}L_n.\mu,\mu)\\
&\leq 2E[L\log L]\,,
\end{aligned}$$

and this completes the proof. □

Corollary 3.1 *Assume that $\nu_i\,(i=1,2)$ have Radon-Nikodym densities $L_i\,(i=1,2)$ with respect to the Wiener measure μ which are in $\mathbb{L}\log\mathbb{L}(\mu)$. Then*

$$d_H(\nu_1,\nu_2)<\infty\,.$$

Proof: This is a simple consequence of the triangle inequality (cf. [5]):

$$d_H(\nu_1,\nu_2)\leq d_H(\nu_1,\mu)+d_H(\nu_2,\mu)\,.$$

□

Let us give a simple application of the above result along the lines of [20]:

Corollary 3.2 *Assume that $A \in \mathcal{B}(W)$ is any set of positive Wiener measure. Define the H-gauge function of A as*

$$q_A(x) = \inf(|h|_H : h \in (A - x) \cap H) .$$

Then we have

$$E[q_A^2] \le 2 \log \frac{1}{\mu(A)} ,$$

in other words

$$\mu(A) \le \exp \left\{ -\frac{E[q_A^2]}{2} \right\} .$$

For $\varepsilon > 0$, let A_ε be the H-neighbourhood of A defined as

$$A_\varepsilon = \{x \in W : q_A(x) \le \varepsilon\} ,$$

then

$$\mu(A_\varepsilon^c) \le \frac{1}{\mu(A)} e^{-\varepsilon^2/4} .$$

Consequently, if A and B are H-separated, i.e., if $A_\varepsilon \cap B = \emptyset$, then

$$\mu(A)\,\mu(B) \le \exp\left(-\frac{\varepsilon^2}{4} \right) .$$

Remark: We already know that, from the $0 - 1$–law, q_A is almost surely finite and it satisfies $|q_A(x+h) - q_A(x)| \le |h|_H$, hence $E[\exp \lambda q_A^2] < \infty$ for any $\lambda < 1/2$ (cf. [31]). In fact all these assertions can also be proved with the technique used below.

Proof: Let ν_A be the measure defined by

$$d\nu_A = \frac{1}{\mu(A)} 1_A d\mu .$$

Let γ_A be the solution of the Monge-Kantorovitch problem, it is easy to see that the support of γ_A is included in $W \times A$, hence

$$|x - y|_H \ge \inf\{|x - z|_H : z \in A\} = q_A(x) ,$$

γ_A-almost surely. This implies in particular that q_A is almost surely finite. It follows now from the inequality (3.3)

$$E[q_A^2] \le -2 \log \mu(A) ,$$

hence the proof of the first inequality follows. For the second let $B = A_\varepsilon^c$ and let γ_{AB} be the solution of the Monge-Kantorovitch problem corresponding to ν_A, ν_B. Then we have from the triangle inequality of Corollary 3.1,

$$d_H^2(\nu_A, \nu_B) \le -4 \log \mu(A)\mu(B) .$$

Moreover the support of the measure γ_{AB} is in $A \times B$, hence γ_{AB}-almost surely $|x - y|_H \ge \varepsilon$ and the proof follows. □

For Rubinstein's distance defined by

$$d_1(\nu,\mu) = \inf\left\{\int_{W\times W} |x-y|_H \, d\theta : \theta \in \Sigma(\mu,\nu)\right\}$$

we have the following control:

Theorem 3.2 *Let $L \in \mathbb{L}^1_+(\mu)$ with $E[L]=1$. Then we have*

$$d_1(L.\mu,\mu) \le E\left[\left|(I+\mathcal{L})^{-1}\nabla L\right|_H\right] . \tag{3.4}$$

Remark 3.1 *In the inequality (3.4), we do not need to assume that L is in some positively indexed Sobolev space since $(I+\mathcal{L})^{-1}$ is a smoothing (pseudo-differential) operator. In fact we may even assume that it is in some $\mathbb{D}_{p,-1}$.*

Proof: To prove the theorem we shall use a technique developed in [8]. Using the conditioning with respect to the sigma algebra $V_n = \sigma\{\delta e_1, \ldots, \delta e_n\}$, where $(e_i, i \ge 1)$ is a complete, orthonormal basis of H, we reduce the problem to the finite dimensional case. Moreover, we can assume that L is a smooth, strictly positive function on $\mathbb{R}^n$. Define now $\sigma = (I+\mathcal{L})^{-1}\nabla L$ and

$$\sigma_t(x) = \frac{\sigma(x)}{t+(1-t)L} ,$$

for $t \in [0,1]$. Let $(\phi_{s,t}(x), s \le t \in [0,1])$ be the flow of diffeomorphisms defined by the following differential equation:

$$\phi_{s,t}(x) = x - \int_s^t \sigma_\tau(\phi_{s,\tau}(x)) d\tau .$$

From standart results (cf. [31], Chapter V), it follows that $x \to \phi_{s,t}(x)$ is Gaussian under the probability $\Lambda_{s,t}.\mu$, where

$$\Lambda_{s,t} = \exp \int_s^t (\delta\sigma_\tau)(\phi_{s,\tau}(x)) d\tau$$

is the Radon-Nikodym density of $\phi^{-1}_{s,t}\mu$ with respect to μ. Define

$$H_s(t,x) = \Lambda_{s,t}(x)\left\{t + (1-t)L\circ\phi_{s,t}(x)\right\} .$$

It is easy to see that

$$\frac{d}{dt}H_s(t,x) = 0$$

for $t \in (s,1)$. Hence the map $t \to H_s(t,x)$ is constant, this implies that

$$\Lambda_{s,1}(x) = s + (1-s)L(x) .$$

We have, as in the proof of Theorem 3.1,

$$\begin{aligned}
d_1(L.\mu,\mu) &\le E[|\phi_{0,1}(x)-x|_H\Lambda_{0,1}] \\
&\le E\left[\Lambda_{0,1}\int_0^1 |\sigma_t(\phi_{0,t}(x))|_H dt\right] \\
&= E\left[\int_0^1 \left|\sigma_t(\phi_{0,t}\circ\phi_{0,1}^{-1})(x)\right|_H dt\right] \\
&= E\left[\int_0^1 \left|\sigma_t(\phi_{t,1}^{-1}(x))\right|_H dt\right] \\
&= E\left[\int_0^1 |\sigma_t(x)|_H\Lambda_{t,1} dt\right] \\
&= E[|\sigma|_H]\,,
\end{aligned}$$

and the proof is complete. □

4 Construction of the transport map for the Gauss measure

In this section we give the construction of the transport map in the Gaussian case. We begin with the following lemma:

Lemma 4.1 *Let (W,μ,H) be an abstract Wiener space, assume that $f: W\to \mathbb{R}$ is a measurable function such that it is Gâteaux differentiable in the direction of the Cameron-Martin space H, i.e., there exists some $\nabla f: W\to H$ such that*

$$f(x+h)=f(x)+\int_0^1(\nabla f(x+\tau h),h)_H d\tau\,,$$

μ-almost surely, for any $h\in H$. If $|\nabla f|_H\in L^2(\mu)$, then f belongs to the Sobolev space $\mathbb{D}_{2,1}$.

Proof: Since $|\nabla|f||_H\le|\nabla f|_H$, we can assume that f is positive. Moreover, for any $n\in\mathbb{N}$, the function $f_n=\min(f,n)$ has also a Gâteaux derivative such that $|\nabla f_n|_H\le|\nabla f|_H$ μ-almost surely. It follows from the Poincaré inequality that the sequence $(f_n-E[f_n],n\ge 1)$ is bounded in $L^2(\mu)$, hence it is also bounded in $L^0(\mu)$. Since f is almost surely finite, the sequence $(f_n,n\ge1)$ is bounded in $L^0(\mu)$, consequently the deterministic sequence $(E[f_n],n\ge1)$ is also bounded in $L^0(\mu)$. This means that $\sup_n E[f_n]<\infty$, hence the monotone convergence theorem implies that $E[f]<\infty$ and the proof is complete. □

Theorem 4.1 *Let ν be the measure $d\nu=Ld\mu$, where L is a positive random variable, with $E[L]=1$. Assume that $d_H(\mu,\nu)<\infty$ (for instance $L\in\mathbb{L}\log\mathbb{L}(\mu)$). Then* *there exists a 1-convex function $\phi \in \mathbb{D}_{2,1}$, unique upto a constant, such that the map $T = I_W + \nabla\phi$ is the unique solution of the problem of Monge. Moreover, its graph supports the unique solution of the Monge-Kantorovitch problem γ. Consequently*

$$(I_W \times T)\mu = \gamma$$

In particular T maps μ to ν and T is almost surely invertible, i.e., there exists some T^{-1} such that $T^{-1}\nu = \mu$ and that

$$\begin{aligned} 1 &= \mu\left\{x : T^{-1} \circ T(x) = x\right\} \\ &= \nu\left\{y \in W : T \circ T^{-1}(y) = y\right\} . \end{aligned}$$

Proof: Let $(\pi_n, n \geq 1)$ be a sequence of regular, finite dimensional orthogonal projections of H increasing to I_H. Denote their continuous extensions to W by the same letters. For $x \in W$, we define $\pi_n^\perp x =: x_n^\perp = x - \pi_n x$. Let ν_n be the measure $\pi_n\nu$. Since ν is absolutely continuous with respect to μ, ν_n is absolutely continuous with respect to $\mu_n := \pi_n\mu$ and

$$\frac{d\nu_n}{d\mu_n} \circ \pi_n = E[L|V_n] =: L_n ,$$

where V_n is the sigma algebra $\sigma(\pi_n)$ and the conditional expectation is taken with respect to μ. On the space H_n, the Monge-Kantorovitch problem, which consists of finding the probability measure which realizes the following infimum

$$d_H^2(\mu_n, \nu_n) = \inf\left\{J(\beta) : \beta \in M_1(H_n \times H_n), p_1\beta = \mu_n, p_2\beta = \nu_n\right\}$$

where

$$J(\beta) = \int_{H_n \times H_n} |x - y|^2 d\beta(x, y) ,$$

has a unique solution γ_n and where p_i, $i = 1, 2$ denote the projections $(x_1, x_2) \to x_i$, $i = 1, 2$ from $H_n \times H_n$ to H_n and $M_1(H_n \times H_n)$ denotes the set of probability measures on $H_n \times H_n$. The measure γ_n may be regarded as a measure on $W \times W$, by taking its image under the injection $H_n \times H_n \hookrightarrow W \times W$ which we shall denote again by γ_n. It results from the finite dimensional results of Brenier [6] and of McCann [21], summarized in Section 2.1, that there are two convex continuous functions (hence γ_n-almost everywhere differentiable) Φ_n and Ψ_n on H_n such that

$$\Phi_n(x) + \Psi_n(y) \geq (x, y)_H$$

for all $x, y \in H_n$ and that

$$\Phi_n(x) + \Psi_n(y) = (x, y)_H$$

γ_n-almost everywhere. Hence the support of γ_n is included in the graph of the derivative $\nabla\Phi_n$ of Φ_n, hence $\nabla\Phi_n\mu_n = \nu_n$ and the inverse of $\nabla\Phi_n$ is equal to $\nabla\Psi_n$. Let

$$\begin{aligned} \phi_n(x) &= \Phi_n(x) - \frac{1}{2}|x|_H^2 \\ \psi_n(y) &= \Psi_n(y) - \frac{1}{2}|y|_H^2 . \end{aligned}$$

Then ϕ_n and ψ_n are 1-convex functions and they satisfy the following relations:

$$\phi_n(x) + \psi_n(y) + \frac{1}{2}|x-y|_H^2 \geq 0\,, \tag{4.5}$$

for all $x, y \in H_n$ and

$$\phi_n(x) + \psi_n(y) + \frac{1}{2}|x-y|_H^2 = 0\,, \tag{4.6}$$

γ_n-almost everywhere. From what we have said above, it follows that γ_n-almost surely $y = x + \nabla\phi_n(x)$, consequently

$$J(\gamma_n) = E[|\nabla\phi_n|_H^2]\,. \tag{4.7}$$

Let $q_n : W \times W \to H_n \times H_n$ be defined as $q_n(x,y) = (\pi_n x, \pi_n y)$. If γ is any solution of the Monge-Kantorovitch problem, then $q_n\gamma \in \Sigma(\mu_n, \nu_n)$, hence

$$J(\gamma_n) \leq J(q_n\gamma) \leq J(\gamma) = d_H^2(\mu,\nu)\,. \tag{4.8}$$

Combining the relation (4.7) with the inequality (4.8), we obtain the following bound

$$\begin{aligned} \sup_n J(\gamma_n) &= \sup_n d_H^2(\mu_n, \nu_n) \\ &= \sup_n E[|\nabla\phi_n|_H^2] \\ &\leq d_H^2(\mu,\nu) = J(\gamma)\,. \end{aligned} \tag{4.9}$$

For $m \leq n$, $q_m\gamma_n \in \Sigma(\mu_m, \nu_m)$, hence we should have

$$\begin{aligned} J(\gamma_m) &= \int_{W\times W} |\pi_m x - \pi_m y|_H^2 d\gamma_m(x,y) \\ &\leq \int_{W\times W} |\pi_m x - \pi_m y|_H^2 d\gamma_n(x,y) \\ &\leq \int_{W\times W} |\pi_n x - \pi_n y|_H^2 d\gamma_n(x,y) \\ &= \int_{W\times W} |x-y|_H^2 d\gamma_n(x,y) \\ &= J(\gamma_n)\,, \end{aligned}$$

where the third equality follows from the fact that we have denoted the γ_n on $H_n \times H_n$ and its image in $W \times W$ by the same letter. Let now γ be a weak cluster point of the sequence of measures $(\gamma_n, n \geq 1)$ [7], where the word " weak" refers to the weak convergence of measures on $W \times W$. Since $(x,y) \to |x-y|_H$ is lower semi-continuous,

[7] We caution the reader that γ_n is not the projection of γ on $W_n \times W_n$.

we have

$$
\begin{aligned}
J(\gamma) &= \int_{W\times W} |x-y|_H^2 d\gamma(x,y) \\
&\leq \liminf_n \int_{W\times W} |x-y|_H^2 d\gamma_n(x,y) \\
&= \liminf_n J(\gamma_n) \\
&\leq \sup_n J(\gamma_n) \\
&\leq J(\gamma) = d_H^2(\mu,\nu)\,,
\end{aligned}
$$

from the relation (4.9). Consequently

$$
J(\gamma) = \lim_n J(\gamma_n)\,. \tag{4.10}
$$

Again from (4.9), if we replace ϕ_n with $\phi_n - E[\phi_n]$ and ψ_n with $\psi_n + E[\phi_n]$ we obtain a bounded sequence $(\phi_n, n\geq 1)$ in $\mathbb{D}_{2,1}$, in particular it is bounded in the space $L^2(\gamma)$ if we inject it into latter by $\phi_n(x)\to\phi_n(x)\otimes 1(y)$. Consider now the sequence of the positive, lower semi-continuous functions $(F_n, n\geq 1)$ defined on $W\times W$ as

$$
F_n(x,y) = \phi_n(x) + \psi_n(y) + \frac{1}{2}|x-y|_H^2\,.
$$

We have, from the relation (4.6)

$$
\begin{aligned}
\int_{W\times W} F_n(x,y)d\gamma(x,y) &= \int_W \phi_n d\mu + \int_W \psi_n(y)d\nu + \frac{1}{2}J(\gamma) \\
&= \frac{1}{2}\left(J(\gamma) - J(\gamma_n)\right) \to 0\,.
\end{aligned}
$$

Consequently the sequence $(F_n, n\geq 1)$ converges to zero in $L^1(\gamma)$, therefore it is uniformly integrable. Since $(\phi_n, n\geq 1)$ is uniformly integrable as explained above and since $|x-y|^2$ has a finite expectation with respect to γ, it follows that $(\psi_n, n\geq 1)$ is also uniformly integrable in $L^1(\gamma)$ hence also in $L^1(\nu)$. Let ϕ' be a weak cluster point of $(\phi_n, n\geq 1)$, then there exists a sequence $(\phi'_n, n\geq 1)$ whose elements are the convex combinations of some elements of $(\phi_k, k\geq n)$ such that $(\phi'_n, n\geq 1)$ converges in the norm topology of $\mathbb{D}_{2,1}$ and μ-almost everywhere. Therefore the sequence $(\psi'_n, n\geq 1)$, constructed from $(\psi_k, k\geq n)$, converges in $L^1(\nu)$ and ν-almost surely. Define ϕ and ψ as

$$
\begin{aligned}
\phi(x) &= \limsup_n \phi'_n(x) \\
\psi(y) &= \limsup_n \psi'_n(y)\,,
\end{aligned}
$$

hence we have

$$
G(x,y) = \phi(x) + \psi(y) + \frac{1}{2}|x-y|_H^2 \geq 0
$$

for all $(x,y) \in W \times W$, also the equality holds γ-almost everywhere. Let now h be any element of H, since $x-y$ is in H for γ-almost all $(x,y) \in W \times W$, we have

$$|x+h-y|_H^2 = |x-y|_H^2 + |h|_H^2 + 2(h, x-y)_H$$

γ-almost surely. Consequently

$$\phi(x+h) - \phi(x) \geq -(h, x-y)_H - \frac{1}{2}|h|_H^2$$

γ-almost surely and this implies that

$$y = x + \nabla\phi(x)$$

γ-almost everywhere. Define now the map $T: W \to W$ as $T(x) = x + \nabla\phi(x)$, then

$$\begin{aligned} \int_{W\times W} f(x,y)d\gamma(x,y) &= \int_{W\times W} f(x,T(x))d\gamma(x,y) \\ &= \int_W f(x,T(x))d\mu(x)\,, \end{aligned}$$

for any $f \in C_b(W \times W)$, consequently $(I_W \times T)\mu = \gamma$, in particular $T\mu = \nu$.

Let us notice that any weak cluster point of $(\phi_n, n \geq 1)$, say $\tilde{\phi}$, satisfies

$$\nabla\tilde{\phi}(x) = y - x$$

γ-almost surely, hence μ-almost surely we have $\tilde{\phi} = \phi$. This implies that $(\phi_n, n \geq 1)$ has a unique cluster point ϕ, consequently the sequence $(\phi_n, n \geq 1)$ converges weakly in $\mathbb{D}_{2,1}$ to ϕ. Besides we have

$$\begin{aligned} \lim_n \int_W |\nabla\phi_n|_H^2 d\mu &= \lim_n J(\gamma_n) \\ &= J(\gamma) \\ &= \int_{W\times W} |x-y|_H^2 d\gamma(x,y) \\ &= \int_W |\nabla\phi|_H^2 d\mu\,, \end{aligned}$$

hence $(\phi_n, n \geq 1)$ converges to ϕ in the norm topology of $\mathbb{D}_{2,1}$. Let us recapitulate what we have done till here: we have taken an arbitrary optimal $\gamma \in \Sigma(\mu,\nu)$ and an arbitrary cluster point ϕ of $(\phi_n, n \geq 1)$ and we have proved that γ is carried by the graph of $T = I_W + \nabla\phi$. This implies that γ and ϕ are unique and that the sequence $(\gamma_n, n \geq 1)$ has a unique cluster point γ.

Certainly $(\psi_n, n \geq 1)$ converges also in the norm topology of $L^1(\nu)$. Moreover, from the finite dimensional situation, we have $\nabla\phi_n(x) + \nabla\psi_n(y) = 0$ γ_n-almost everywhere. Hence

$$E_\nu[|\nabla\psi_n|_H^2] = E[|\nabla\phi_n|_H^2]$$

this implies the boundedness of $(\nabla\psi_n, n \geq 1)$ in $L^2(\nu, H)$ (i.e., H-valued functions). To complete the proof we have to show that, for some measurable, H-valued map, say η, it holds that $x = y + \eta(y)$ γ-almost surely. For this let F be a finite dimensional, regular subspace of H and denote by π_F the projection operator onto F which is continuously extended to W, put $\pi_F^\perp = I_W - \pi_F$. We have $W = F \oplus F^\perp$, with $F^\perp = \ker \pi_F = \pi_F^\perp(W)$. Define the measures $\nu_F = \pi_F(\nu)$ and $\nu_F^\perp = \pi_F^\perp(\nu)$. From the construction of ψ, we know that, for any $v \in F^\perp$, the partial map $u \to \psi(u+v)$ is 1-convex on F. Let also $A = \{y \in W : \psi(y) < \infty\}$, then A is a Borel set with $\nu(A) = 1$ and it is easy to see that, for $\nu_F^\perp$-almost all $v \in F^\perp$, one has

$$\nu(A|\pi_F^\perp = v) > 0\,.$$

It then follows from Lemma 3.4 of [15], and from the fact that the regular conditional probability $\nu(\cdot\,|\pi_F^\perp = v)$ is absolutely continuous with respect to the Lebesgue measure of F, that $u \to \psi(u+v)$ is $\nu(\cdot\,|\pi_F^\perp = v)$-almost everywhere differentiable on F for $\nu_F^\perp$-almost all $v \in F^\perp$. It then follows that, ν-almost surely, ψ is differentiable in the directions of F, i.e., there exists $\nabla_F\psi \in F$ ν-almost surely. Since we also have

$$\psi(y+k) - \psi(y) \geq (x-y,k)_H - \frac{1}{2}|k|_H^2\,,$$

we obtain, γ-almost surely

$$(\nabla_F\psi(y), k)_H = (x-y,k)_H\,,$$

for any $k \in F$. Consequently

$$\nabla_F\psi(y) = \pi_F(x-y)$$

γ-almost surely. Let now $(F_n, n \geq 1)$ be a total, increasing sequence of regular subspaces of H, we have a sequence $(\nabla_n\psi, n \geq 1)$ bounded in $L^2(\nu)$ hence also bounded in $L^2(\gamma)$. Moreover $\nabla_n\psi(y) = \pi_n x - \pi_n y$ γ-almost surely. Since $(\pi_n(x-y), n \geq 1)$ converges in $L^2(\gamma, H)$, $(\nabla_n\psi, n \geq 1)$ converges in the norm topology of $L^2(\gamma, H)$. Let us denote this limit by η, then we have $x = y + \eta(y)$ γ-almost surely. Note that, since $\pi_n\eta = \nabla_n\psi$, we can even write in a weak sense that $\eta = \nabla\psi$. If we define $T^{-1}(y) = y + \eta(y)$, we see that

$$\begin{aligned} 1 &= \gamma\{(x,y) \in W \times W : T \circ T^{-1}(y) = y\} \\ &= \gamma\{(x,y) \in W \times W : T^{-1} \circ T(x) = x\}\,, \end{aligned}$$

and this completes the proof of the theorem. □

Remark 4.1 *Assume that the operator ∇ is closable with respect to ν, then we have $\eta = \nabla\psi$. In particular, if ν and μ are equivalent, then we have*

$$T^{-1} = I_W + \nabla\psi\,,$$

where ψ is a 1-convex function.

Remark 4.2 *Assume that $L \in \mathbb{L}^1_+(\mu)$, with $E[L] = 1$ and let $(D_k, k \in \mathbb{N})$ be a measurable partition of W such that on each D_k, L is bounded. Define $d\nu = L\,d\mu$ and $\nu_k = \nu(\cdot|D_k)$. It follows from Theorem 3.1, that $d_H(\mu, \nu_k) < \infty$. Let then T_k be the map constructed in Theorem 4.1 satisfying $T_k\mu = \nu_k$. Define $n(dk)$ as the probability distribution on $\mathbb{N}$ given by $n(\{k\}) = \nu(D_k)$, $k \in \mathbb{N}$. Then we have*

$$\int_W f(y)d\nu(y) = \int_{W\times\mathbb{N}} f(T_k(x))\mu(dx)n(dk)\,.$$

A similar result is given in [13], the difference with that above lies in the fact that we have a more precise information about the probability space on which T is defined.

5 Polar factorization of the absolutely continuous transformations of the Wiener space

Assume that $V = I_W + v : W \to W$ be an absolutely continuous transformation and let $L \in \mathbb{L}^1_+(\mu)$ be the Radon-Nikodym derivative of $V\mu$ with respect to μ. Let $T = I_W + \nabla\phi$ be the transport map such that $T\mu = L.\mu$. Then it is easy to see that the map $s = T^{-1} \circ V$ is a rotation, i.e., $s\mu = \mu$ (cf. [31]) and it can be represented as $s = I_W + \alpha$. In particular we have

$$\alpha + \nabla\phi \circ s = v\,. \tag{5.11}$$

Since ϕ is a 1-convex map, we have $h \to \frac{1}{2}|h|_H^2 + \phi(x+h)$ is almost surely convex [15]. Let $s' = I_W + \alpha'$ be another rotation with $\alpha' : W \to H$. By the 1-convexity of ϕ, we have

$$\frac{1}{2}|\alpha'|_H^2 + \phi\circ s' \geq \frac{1}{2}|\alpha|_H^2 + \phi\circ s + \Big(\alpha + \nabla\phi\circ s, \alpha' - \alpha\Big)_H\,,$$

μ-almost surely. Taking the expectation of both sides, using the fact that s and s' preserve the Wiener measure μ and the identity (5.11), we obtain

$$E\left[\frac{1}{2}\,|\alpha|_H^2 - (v,\alpha)_H\right] \leq E\left[\frac{1}{2}|\alpha'|_H^2 - (v,\alpha')_H\right]\,.$$

Hence we have proven the existence part of the following

Proposition 5.1 *Let $\mathcal{R}_2$ denote the subset of $L^2(\mu, H)$ whose elements η are defined by the property that $x \to x + \eta(x)$ is a rotation, i.e., it preserves the Wiener measure. Then there exists a unique element α of $\mathcal{R}_2$ which is defined by the relation (5.11), minimizing the functional*

$$\eta \to M_v(\eta) = E\left[\frac{1}{2}|\eta|_H^2 - (v,\eta)_H\right]\,.$$

Proof: To show the uniqueness, assume that $\eta \in \mathcal{R}_2$ be another map minimizing M_v. Let β be the measure on $W \times W$, defined as

$$\int_{W\times W} f(x,y)d\beta(x,y) = \int_W f(x+\eta(x), V(x))d\mu\,.$$

Then the first marginal of β is μ and the second marginal is $L.\mu$. Since $\gamma = (I_W \times T)\mu$ is the unique solution of the Monge-Kantorovitch problem, we should have

$$\int |x-y|_H^2 d\beta(x,y) > \int |x-y|_H^2 d\gamma(x,y) = E[|\nabla\phi|_H^2]\,.$$

However we have

$$\begin{aligned}
\int_{W\times W} |x-y|_H^2 d\beta(x,y) &= E\left[|v-\eta|_H^2\right] \\
&= E\left[|v|_H^2\right] + 2M_v(\eta) \\
&= E\left[|v|_H^2\right] + 2M_v(\alpha) \\
&= E\left[|v-\alpha|_H^2\right] \\
&= E\left[|\nabla\phi\circ s|_H^2\right] \\
&= E\left[|\nabla\phi|_H^2\right] \\
&= \int_{W\times W} |x-y|_H^2 d\gamma(x,y) \\
&= J(\gamma)
\end{aligned}$$

and this gives a contradiction to the uniqueness of γ. □

The following theorem, whose proof is rather easy, gives a better understanding of the structure of absolutely continuous transformations of the Wiener measure:

Theorem 5.1 *Assume that $U: W \to W$ be a measurable map and $L \in \mathbb{L}\log\mathbb{L}(\mu)$ a positive random variable with $E[L]=1$. Assume that the measure $\nu = L\cdot\mu$ is a Girsanov measure for U, i.e., that one has*

$$E[f\circ U\, L] = E[f]\,,$$

for any $f \in C_b(W)$. Then there exists a unique map $T = I_W + \nabla\phi$ with $\phi \in \mathbb{D}_{2,1}$ is 1-convex, and a measure preserving transformation $R: W\to W$ such that $U\circ T = R$ μ-almost surely and $U = R\circ T^{-1}$ ν-almost surely.

Proof: By Theorem 4.1 there is a unique map $T = I_W + \nabla\phi$, with $\phi\in\mathbb{D}_{2,1}$, 1-convex such that T transports μ to ν. Since $U\nu = \mu$, we have

$$\begin{aligned}
E[f\circ U\, L] &= E[f\circ U\circ T] \\
&= E[f]\,.
\end{aligned}$$

Therefore $x \to U\circ T(x)$ preserves the measure μ. The rest is obvious since T^{-1} exists ν-almost surely. □

Another version of Theorem 5.1 can be stated as follows:

Theorem 5.2 *Assume that $Z : W \to W$ is a measurable map such that $Z\mu \ll \mu$, with $d_H(Z\mu, \mu) < \infty$. Then Z can be decomposed as*

$$Z = T \circ s \,,$$

where T is the unique transport map of the Monge-Kantorovitch problem for $\Sigma(\mu, Z\mu)$ and s is a rotation.

Proof: Let L be the Radon-Nikodym derivative of $Z\mu$ with respect to μ. We have, from Theorem 4.1,

$$\begin{aligned} E[f] &= E[f \circ T^{-1} \circ T] \\ &= E[f \circ T^{-1} \, L] \\ &= E[f \circ T^{-1} \circ Z] \,, \end{aligned}$$

for any $f \in C_b(W)$. Hence $T^{-1} \circ Z = s$ is a rotation. Since T is uniquely defined, s is also uniquely defined. □

Although the following result is a translation of the results of this section, it is interesting from the point of view of stochastic differential equations:

Theorem 5.3 *Let (W, μ, H) be the standard Wiener space on $\mathbb{R}^d$, i.e., $W = C_0(\mathbb{R}_+, \mathbb{R}^d)$. Assume that there exists a probability $P \ll \mu$ which is the weak solution of the stochastic differential equation*

$$dy_t = dW_t + b(t, y)dt \,,$$

such that $d_H(P, \mu) < \infty$. Then there exists a process $(T_t, t \in \mathbb{R}_+)$ which is a pathwise solution of some stochastic differential equation whose law is equal to P.

Proof: Let T be the transport map constructed in Theorem 4.1 corresponding to $dP/d\mu$. Then it has an inverse T^{-1} such that $\mu\{T^{-1} \circ T(x) = x\} = 1$. Let ϕ be the 1-convex function such that $T = I_W + \nabla\phi$ and denote by $(D_s\phi, s \in \mathbb{R}_+)$ the representation of $\nabla\phi$ in $L^2(\mathbb{R}_+, ds)$. Define $T_t(x)$ as the trajectory $T(x)$ evaluated at $t \in \mathbb{R}_+$. Then it is easy to see that $(T_t, t \in \mathbb{R}_+)$ satifies the stochastic differential equation

$$T_t(x) = W_t(x) + \int_0^t l(s, T(x))ds \ , \ t \in \mathbb{R}_+ \,,$$

where $W_t(x) = x(t)$ and $l(s, x) = D_s\phi \circ T^{-1}(x)$. □

6 Construction and uniqueness of the transport map in the general case

In this section we call optimal every probability measure[8] γ on $W \times W$ such that $J(\gamma) < \infty$ and that $J(\gamma) \le J(\theta)$ for every other probability θ having the same marginals

[8]In fact the results of this section are essentially true for bounded, positive measures.

as those of γ. We recall that a finite dimensional subspace F of W is called regular if the corresponding projection is continuous. Similarly a finite dimensional projection of H is called regular if it has a continuous extension to W.

We begin with the following lemma which answers all kind of questions of measurability that we may encounter in the sequel:

Lemma 6.1 *Consider two uncountable Polish spaces X and T. Let $t \to \gamma_t$ be a Borel family of probabilities on X and let $\mathcal{F}$ be a separable sub-σ-algebra of the Borel σ-algebra $\mathcal{B}$ of X. Then there exists a Borel kernel*

$$N_t f(x) = \int_X f(y) N_t(x, dy)\,,$$

such that, for any bounded Borel function f on X, the following properties hold true:

i) $(t, x) \to N_t f(x)$ is Borel measurable on $T \times X$.

ii) For any $t \in T$, $N_t f$ is an $\mathcal{F}$-measurable version of the conditional expectation $E_{\gamma_t}[f|\mathcal{F}]$.

Proof: Assume first that $\mathcal{F}$ is finite, hence it is generated by a finite partition $\{A_1, \ldots, A_k\}$. In this case it suffices to take

$$N_t f(x) = \sum_{i=1}^{k} \frac{1}{\gamma_t(A_i)} \left(\int_{A_i} f d\gamma_t \right) 1_{A_i}(x) \ \left(\text{with } 0 = \frac{0}{0} \right).$$

For the general case, take an increasing sequence $(\mathcal{F}_n, n \geq 1)$ of finite sub-σ-algebras whose union generates $\mathcal{F}$. Without loss of generality we can assume that $(X, \mathcal{B})$ is the Cantor set (Kuratowski Theorem, cf., [9]). Then for every clopen set (i.e., a set which is closed and open at the same time) G and any $t \in T$, the sequence $(N_t^n 1_G, n \geq 1)$ converges γ_t-almost everywhere. Define

$$H_G(t, x) = \limsup_{m,n \to \infty} |N_t^n 1_G(x) - N_t^m 1_G(x)|\,.$$

H_G is a Borel function on $T \times X$ which vanishes γ_t-almost all $x \in X$, moreover, for any $t \in T$, $x \to H_G(t, x)$ is $\mathcal{F}$-measurable. As there exist only countably many clopen sets in X, the function

$$H(t, x) = \sup_G H_G(t, x)$$

inherits all the measurability properties. Let θ be any probability on X, for any clopen G, define

$$\begin{aligned} N_t 1_G(x) &= \lim_n N_t^n 1_G(x) \quad \text{if} \quad H(t,x) = 0\,, \\ &= \theta(G) \quad \text{if} \quad H(t,x) > 0\,. \end{aligned}$$

Hence, for any $t \in T$, we get an additive measure on the Boolean algebra of clopen sets of X. Since such a measure is σ-additive and extends uniquely as a σ-additive measure on $\mathcal{B}$, the proof is complete. □

Remark 6.1 1. In fact due to the Theorem of Lusin, the map $(t,x) \to N_t(x,dy)$ is also measurable as a measure-valued map.

2. This result holds in fact for Lusin spaces since they are Borel isomorphic to the Cantor set.

3. The particular case where $T = \mathcal{M}_1(X)$, i.e., the space of probability measures on X under the weak topology and $t \to \gamma_t$ being the identity map, is particularly important for the sequel. In this case we obtain a kernel N such that $(x,\gamma) \to N_\gamma f(x)$ is measurable and $N_\gamma f$ is an $\mathcal{F}$-measurable version of $E_\gamma[f|\mathcal{F}]$.

Lemma 6.2 *Let ρ and ν be two probability measures on W such that*

$$d_H(\rho,\nu) < \infty$$

and let $\gamma \in \Sigma(\rho,\nu)$ be an optimal measure, i.e., $J(\gamma) = d_H^2(\rho,\nu)$, where J is given by (1.1). Assume that F is a regular finite dimensional subspace of W with the corresponding projection π_F from W to F and let $\pi_F^\perp = I_W - \pi_F$. Define p_F as the projection from $W \times W$ onto F with $p_F(x,y) = \pi_F x$ and let $p_F^\perp(x,y) = \pi_F^\perp x$. Consider the Borel disintegration

$$\begin{aligned} \gamma(\cdot) &= \int_{F^\perp \times W} \gamma(\,\cdot|x^\perp)\gamma^\perp(dz^\perp) \\ &= \int_{F^\perp} \gamma(\,\cdot|x^\perp)\rho^\perp(dx^\perp) \end{aligned}$$

along the projection of $W \times W$ on $F^\perp$, where $\rho^\perp$ is the measure $\pi_F^\perp \rho$, $\gamma(\cdot\,|x^\perp)$ denotes the regular conditional probability $\gamma(\cdot\,|p_F^\perp = x^\perp)$ and $\gamma^\perp$ is the measure $p_F^\perp\gamma$. Then, $\rho^\perp$ and $\gamma^\perp$-almost surely $\gamma(\,\cdot|x^\perp)$ is optimal on $(x^\perp + F) \times W$.

Proof: Let p_1, p_2 be the projections of $W \times W$ defined as $p_1(x,y) = \pi_F(x)$ and $p_2(x,y) = \pi_F(y)$. Note first the following obvious identity:

$$p_1\gamma(\cdot\,|x^\perp) = \rho(\cdot\,|x^\perp)\,,$$

$\rho^\perp$ and $\gamma^\perp$-almost surely. Define the sets $B \subset F^\perp \times \mathcal{M}_1(F \times F)$ and C as

$$\begin{aligned} B &= \{(x^\perp,\theta) : \theta \in \Sigma(p_1\gamma(\cdot\,|x^\perp), p_2\gamma(\cdot\,|x^\perp))\} \\ C &= \{(x^\perp,\theta) \in B : J(\theta) < J(\gamma(\cdot\,|x^\perp)\}\,, \end{aligned}$$

where $\mathcal{M}_1(F \times F)$ denotes the set of probability measures on $F \times F$. Let K be the projection of C on $F^\perp$. Since B and C are Borel measurable, $K$is a Souslin set, hence it is $\rho^\perp$-measurable. The selection theorem (cf. [9]) implies the existence of a measurable map

$$x^\perp \to \theta_{x^\perp}$$

from K to $\mathcal{M}_1(F \times F)$ such that, $\rho^\perp$-almost surely, $(x^\perp, \theta_{x^\perp}) \in C$. Define

$$\theta(\cdot) = \int_K \theta_{x^\perp}(\cdot)d\rho^\perp(x^\perp) + \int_{K^c} \gamma(\cdot\,|x^\perp)d\rho^\perp(x^\perp)\,.$$

Then $\theta \in \Sigma(\rho, \nu)$ and we have

$$\begin{aligned} J(\theta) &= \int_K J(\theta_{x^\perp}) d\rho^\perp(x^\perp) + \int_{K^c} J(\gamma(\cdot\,|x^\perp)) d\rho^\perp(x^\perp) \\ &< \int_K J(\gamma(\cdot\,|x^\perp)) d\rho^\perp(x^\perp) + \int_{K^c} J(\gamma(\cdot\,|x^\perp)) d\rho^\perp(x^\perp) \\ &= J(\gamma)\,, \end{aligned}$$

hence we obtain $J(\theta) < J(\gamma)$ which is a contradiction to the optimality of γ. □

Lemma 6.3 *Assume that the hypothesis of Lemma 6.2 holds and let F be any regular finite dimensional subspace of W. Denote by π_F the projection operator associated to it and let $\pi_F^\perp = I_W - \pi_F$. If $\pi_F^\perp \rho$-almost surely, the regular conditional probability $\rho(\cdot\,|\pi_F^\perp = x^\perp)$ vanishes on the subsets of $x^\perp + F$ whose Hausdorff dimension are at most equal to $\dim(F) - 1$, then there exists a map $T_F : F \times F^\perp \to F$ such that*

$$\gamma\left(\left\{(x, y) \in W \times W : \pi_F y = T_F(\pi_F x, \pi_F^\perp x)\right\}\right) = 1\,.$$

Proof: Let $C_{x^\perp}$ be the support of the regular conditional probability $\gamma(\cdot\,|x^\perp)$ in $(x^\perp + F) \times W$. We know from Lemma 6.2 that the measure $\gamma(\cdot\,|x^\perp)$ is optimal in $\Sigma(\pi_1\gamma(\cdot\,|x^\perp), \pi_2\gamma(\cdot\,|x^\perp))$, with $J(\gamma(\cdot\,|x^\perp)) < \infty$ for $\rho^\perp$-almost everywhere $x^\perp$. From Theorem 2.3 of [17] and from [1], the set $C_{x^\perp}$ is cyclically monotone, moreover, $C_{x^\perp}$ is a subset of $(x^\perp + F) \times H$, hence the cyclic monotonicity of it implies that the set $K_{x^\perp} \subset F \times F$, defined as

$$K_{x^\perp} = \{(u, \pi_F v) \in F \times F : (x^\perp + u, v) \in C_{x^\perp}\}$$

is cyclically monotone in $F \times F$. Therefore $K_{x^\perp}$ is included in the subdifferential of a convex function defined on F. Since, by hypothesis, the first marginal of $\gamma(\cdot\,|x^\perp)$, i.e., $\rho(\cdot\,|x^\perp)$ vanishes on the subsets of $x^\perp + F$ of co-dimension one, the subdifferential under question, denoted as $U_F(u, x^\perp)$ is $\rho(\cdot\,|x^\perp)$-almost surely univalent (cf. [2, 21]). This implies that

$$\gamma(\cdot\,|x^\perp)\left(\left\{(u, v) \in C_{x^\perp} : \pi_F v = U_F(u, x^\perp)\right\}\right) = 1\,,$$

$\rho^\perp$-almost surely. Let

$$K_{x^\perp, u} = \{v \in W : (u, v) \in K_{x^\perp}\}\,.$$

Then $K_{x^\perp, u}$ consists of a single point for almost all u with respect to $\rho(\cdot\,|x^\perp)$. Let

$$N = \left\{(u, x^\perp) \in F \times F^\perp : \mathrm{Card}(K_{x^\perp, u}) > 1\right\}\,,$$

note that N is a Souslin set, hence it is universally measurable. Let σ be the measure which is defined as the image of ρ under the projection $x \to (\pi_F x, \pi_F^\perp x)$. We then have

$$\begin{aligned} \sigma(N) &= \int_{F^\perp} \rho^\perp(dx^\perp) \int_F \mathbf{1}_N(u, x^\perp) \rho(du|x^\perp) \\ &= 0\,. \end{aligned}$$

Hence $(u, x^\perp) \mapsto K_{x^\perp,u} = \{y\}$ is ρ and γ-almost surely well-defined and it suffices to denote this map by T_F to obtain the proof. $\square$

Theorem 6.1 *Suppose that ρ and ν are two probability measures on W such that*

$$d_H(\rho, \nu) < \infty\,.$$

Let $(\pi_n, n \geq 1)$ be a total increasing sequence of regular projections (of H, converging to the identity map of H). Suppose that, for any $n \geq 1$, the regular conditional probabilities $\rho(\cdot\,|\pi_n^\perp = x_n^\perp)$ vanish $\pi_n^\perp\rho$-almost surely on the subsets of $x_n^\perp + \pi_n(W)$ with Hausdorff dimension $(n-1)$. Then there exists a unique solution of the Monge-Kantorovitch problem, denoted by $\gamma \in \Sigma(\rho, \nu)$ and γ is supported by the graph of a Borel map $T : W \to W$ of the form $T = I_W + \xi$ with $\xi \in H$ almost surely. Moreover $T = I_W + \xi$ is the unique solution of the Monge problem and for all $n \geq 1$, $\pi_n^\perp\rho$-almost all $x_n^\perp$, the map $u \to u + \xi(u + x_n^\perp)$ is cyclically monotone on $\pi_n(W)$. Finally, if, for any $n \geq 1$, $\pi_n^\perp\nu$-almost surely, $\nu(\cdot\,|\pi_n^\perp = y^\perp)$ also vanishes on the $(n-1)$-Hausdorff dimensional subsets of $x_n^\perp + \pi_n(W)$, then T is invertible, i.e, there exists $S : W \to W$ of the form $S = I_W + \eta$ such that $\eta \in H$ satisfies a similar cyclic monotonicity property as ξ and that

$$\begin{aligned} 1 &= \gamma\left\{(x,y) \in W \times W : T \circ S(y) = y\right\} \\ &= \gamma\left\{(x,y) \in W \times W : S \circ T(x) = x\right\}\,. \end{aligned}$$

In particular we have

$$\begin{aligned} d_H^2(\rho, \nu) &= \int_{W\times W} |x-y|_H^2 d\gamma(x,y) \\ &= \int_W |S(y) - y|_H^2 d\nu(y)\,. \end{aligned}$$

Remark 6.2 *For the measures ρ which are absolutely continuous with respect to the Wiener measure μ, the second hypothesis is satisfied, i.e., the measure $\rho(\cdot\,|\pi_n^\perp = x_n^\perp)$ vanishes on the subsets of Hausdorff dimension $(n-1)$.*

Proof: Let γ be any optimal measure, i.e., $J(\gamma) = d_H^2(\rho, \nu)$ and let $(F_n, n \geq 1)$ be the increasing sequence of regular subspaces associated to $(\pi_n, n \geq 1)$, whose union is dense in W. From Lemma 6.3, for any F_n, there exists a map T_n, such that $\pi_n y = T_n(\pi_n x, \pi_n^\perp x)$ for γ-almost all (x, y), where $\pi_n^\perp = I_W - \pi_n$. Write T_n as $I_n + \xi_n$, where I_n denotes the identity map on F_n. Then we have the following representation:

$$\pi_n y = \pi_n x + \xi_n(\pi_n x, \pi_n^\perp x)\,,$$

γ-almost surely. Since

$$\begin{aligned} \pi_n y - \pi_n x &= \pi_n(y - x) \\ &= \xi_n(\pi_n x, \pi_n^\perp x) \end{aligned}$$

and since $y - x \in H$ γ-almost surely, $(\pi_n y - \pi_n x, n \geq 1)$ converges γ-almost surely. Consequently $(\xi_n, n \geq 1)$ converges γ, hence ρ almost surely to a measurable ξ. Consequently we obtain

$$\gamma\left(\{(x,y) \in W \times W : y = x + \xi(x)\}\right) = 1 .$$

Since $J(\gamma) < \infty$, ξ takes its values almost surely in the Cameron-Martin space H. The cyclic monotonicity of ξ is obvious. To prove the uniqueness, assume that we have two optimal solutions γ_1 and γ_2 with the same marginals and $J(\gamma_1) = J(\gamma_2)$. Since $\beta \to J(\beta)$ is linear, the measure defined as $\gamma = \frac{1}{2}(\gamma_1 + \gamma_2)$ is also optimal and it has also the same marginals ρ and ν. Consequently, it is also supported by the graph of a map T. Note that γ_1 and γ_2 are absolutely continuous with respect to γ, let $L_1(x,y)$ be the Radon-Nikodym density of γ_1 with respect to γ. For any $f \in C_b(W)$, we then have

$$\begin{aligned}\int_W f d\rho &= \int_{W\times W} f(x) d\gamma_1(x,y) \\ &= \int_{W\times W} f(x) L_1(x,y) d\gamma(x,y) \\ &= \int_W f(x) L_1(x, T(x)) d\rho(x) .\end{aligned}$$

Therefore we should have ρ-almost surely, $L_1(x, T(x)) = 1$, hence also $L_1 = 1$ almost everywhere γ and this implies that $\gamma = \gamma_1 = \gamma_2$. This shows the uniqueness of the solution of PMK. To show the uniqueness of the solution of the problem of Monge, assume that there are T_1 and T_2 which solve both the Monge problem. Then γ_i, $i = 1, 2$ defined by $(I \times T_i)\rho$, $i = 1, 2$, are both the solutions of PMK, hence $\gamma_1 = \gamma_2$, this implies that $f \circ T_1 = f \circ T_2$ ρ-almost surely for any positive, Borel function f on W, hence $T_1 = T_2$ ρ-almost surely. To prove the last part of the theorem, we can construct symmetrically the map S such that $(S \times I)\nu = \gamma$. We then have

$$\begin{aligned}1 &= \gamma\{(x,y) : y = T(x)\} \\ &= \gamma\{(x,y) : x = S(y)\} ,\end{aligned}$$

hence the intersection of these two sets is again of full γ-measure, in other words we have

$$\begin{aligned}1 &= \gamma\{(x,y) \in W \times W : S \circ T(x) = x\} \\ &= \rho\{x \in W : S \circ T(x) = x\} \\ &= \gamma\{(x,y) \in W \times W : T \circ S(y) = y\} \\ &= \nu\{y \in W : T \circ S(y) = y\} ,\end{aligned}$$

and this completes the proof. □

Corollary 6.1 *Assume that ρ is equivalent to the Wiener measure μ, then for any $h_1, \ldots, h_N \in H$ and for any permutation τ of $\{1, \ldots, N\}$, we have, with the notations of Theorem 6.1,*

$$\sum_{i=1}^{N} \left(h_i + \xi(x + h_i), h_{\tau(i)} - h_i\right)_H \leq 0$$

ρ-almost surely.

Proof: Again with the notations of the theorem, $\rho_k^{\perp}$-almost surely, the graph of the map $x_k \to x_k + \xi_k(x_k, x_k^{\perp})$ is cyclically monotone on F_k. Hence, for the case $h_i \in F_n$ for all $i = 1, \ldots, N$ and $n \leq k$, we have

$$\sum_{i=1}^{N} \Big(h_i + x_k + \xi_k(x_k + h_i, x_k^{\perp}), h_{\tau(i)} - h_i\Big)_H \leq 0\,.$$

Since $\sum_i (x_k, h_{\tau(i)} - h_i)_H = 0$, we also have

$$\sum_{i=1}^{N} \Big(h_i + \xi_k(x_k + h_i, x_k^{\perp}), h_{\tau(i)} - h_i\Big)_H \leq 0\,.$$

We know that $\xi_k(x_k + h_i, x_k^{\perp})$ converges to $\xi(x + h_i)$ ρ-almost surely. Moreover $h \to \xi(x + h)$ is continuous from H to $L^0(\rho)$ and the proof follows. □

Remark 6.3 *We could have defined a notion of cyclical monotonicity of ξ (or of η) relative to a sequence of orthogonal projections $(\pi_n, n \geq 1)$ of H, increasing to the identity I_H of H and relative to the measure ρ by saying that ξ is $(\pi_n, n \geq 1)$-1-ρ-cyclically monotone if the map $u \to u + \xi(u + x_n^{\perp})$ is cyclically monotone on $F_n = \pi_n(H)$ for $\pi_n^{\perp}\rho$-almost all $x_n^{\perp}$ and for any $n \geq 1$. Although this notion seems interesting, we have avoided this option in order not to make the paper too technical.*

7 The Monge-Ampère equation

In this section we study the Monge-Ampère equation on the infinite dimensional Wiener space. Recall that, in the finite dimensional case, this equation consists, for two given positive functions f and g, whose Lebesgue integrals are equal to one, of finding a transformation T such that

$$g \circ T\, |J_T| = f\,, \tag{7.12}$$

where J_T denotes the Jacobian of T. Of course in general, such T is not unique, however, if we want it to be also the solution of the Monge problem, hence of the form $T = I + \nabla\phi$, then we can have the uniqueness. In the infinite dimensional case, the main difficulty stems from the absence of the Lebesgue measure. The good candidate to replace it is the Gauss measure, however the Jacobian term should be arranged in a more practical form. This can be done (cf. for further explanation [31]) by taking the Gaussian Jacobian

$$\Lambda_T = \det_2(I + \nabla^2\phi)\exp\left\{-\delta\nabla\phi - \frac{1}{2}|\nabla\phi|^2\right\}.$$

In this expression as well as in the sequel, the notation $\det_2(I_H + A)$ denotes the modified Carleman-Fredholm determinant of the operator $I_H + A$ on a Hilbert space H. If A is an operator of finite rank, then it is defined as

$$\det_2(I_H + A) = \prod_{i=1}^{n}(1 + l_i)e^{-l_i},$$

where $(l_i, i \le n)$ denotes the eigenvalues of A counted respecting their multiplicity. In fact this determinant has an analytic extension to the space of Hilbert-Schmidt operators on a separable Hilbert space, cf. [11] and Appendix A.2 of [31]. Consequently the modified Carleman-Fredholm determinant exists for the Hilbert-Schmidt operators while the ordinary determinant does not, since the existence of the latter requires the existence of the trace of A. Hence the modified Carleman-Fredholm determinant is particularly useful when one deals with the absolute continuity properties of the image of a Gaussian measure under non-linear transformations in the setting of infinite dimensional Banach spaces (cf., [31] for further information). Note also that, since $\delta \circ \nabla$ is equal to the Ornstein-Uhlenbeck operator $\mathcal{L}$, our new Jacobian takes the form

$$\Lambda_T = \det_2(I + \nabla^2\phi)\exp\left\{-\mathcal{L}\phi - \frac{1}{2}|\nabla\phi|^2\right\}.$$

There are some other difficulties concerning the second order differentiability of ϕ. In finite dimension, in order to avoid this difficulty, we use the notion of Alexandrov differentiability as explained below: Assume again that $W = \mathbb{R}^n$ and take a density $L \in \mathbb{L}\log\mathbb{L}(\mu)$. Let $\phi \in \mathbb{D}_{2,1}$ be the 1-convex function such that $T = I + \nabla\phi$ maps μ to $L \cdot \mu$. Let $S = I + \nabla\psi$ be its inverse with $\psi \in \mathbb{D}_{2,1}$. Let now $\nabla_a^2\phi$ be the second Alexandrov derivative of ϕ, i.e., the Radon-Nikodym derivative of the absolutely continuous part of the vector measure $\nabla^2\phi$ with respect to the Gaussian measure μ on $\mathbb{R}^n$. Since ϕ is 1-convex, it follows that $\nabla^2\phi \ge -I_{\mathbb{R}^n}$ in the sense of the distributions, consequently $\nabla_a^2\phi \ge -I_{\mathbb{R}^n}$ μ-almost surely. Define also the Alexandrov version $\mathcal{L}_a\phi$ of $\mathcal{L}\phi$ as the Radon-Nikodym derivative of the absolutely continuous part of the distribution $\mathcal{L}\phi$. In finite dimensional situation, there is an explicit expression for $\mathcal{L}_a\phi$ as

$$\mathcal{L}_a\phi(x) = (\nabla\phi(x), x)_{\mathbb{R}^n} - \mathrm{trace}\left(\nabla_a^2\phi\right).$$

Let Λ be the Gaussian Jacobian

$$\Lambda = \det_2\left(I_{\mathbb{R}^n} + \nabla_a^2\phi\right)\exp\left\{-\mathcal{L}_a\phi - \frac{1}{2}|\nabla\phi|^2_{\mathbb{R}^n}\right\}.$$

It follows from the change of variables formula given in Corollary 4.3 of [22], that, for any $f \in C_b(\mathbb{R}^n)$,

$$E[f \circ T\,\Lambda] = E\left[f\,1_{\partial\Phi(M)}\right],$$

where M is the set of non-degeneracy of $I_{\mathbb{R}^n} + \nabla_a^2\phi$,

$$\Phi(x) = \frac{1}{2}|x|^2 + \phi(x)$$

and $\partial\Phi$ denotes the subdifferential of the convex function Φ. Hence Λ appears to be a good candidate for the Jacobian for the Gaussian case.

Let us note that, in case $L > 0$ almost surely, T has a global inverse S, i.e., $S \circ T = T \circ S = I_{\mathbb{R}^n}$ μ-almost surely and $\mu(\partial\Phi(M)) = \mu(S^{-1}(M))$. Assume now that $\Lambda > 0$ almost surely, i.e., that $\mu(M) = 1$. Then, for any $f \in C_b(\mathbb{R}^n)$, we have

$$\begin{aligned} E[f \circ T] &= E\left[f \circ T \frac{\Lambda}{\Lambda \circ T^{-1} \circ T}\right] \\ &= E\left[f \frac{1}{\Lambda \circ T^{-1}} 1_{\partial\Phi(M)}\right] \\ &= E[f\, L]\,, \end{aligned}$$

where T^{-1} denotes the left inverse of T whose existence is guaranteed by Theorem 4.1. Since $T(x) \in \partial\Phi(M)$ almost surely, it follows from the above calculations

$$\frac{1}{\Lambda} = L \circ T\,,$$

almost surely. Hence T is the solution in the Alexandrov sense of the Monge-Ampère equation with $f = 1$ and $g = L$ with the notations of the equation (7.12).

Before proceeding to infinite dimensions let us give some preliminaries about the linear interpolations of the transport map T: let $t \in [0, 1)$, the map $x \to \frac{1}{2}|x|_H^2 + t\phi(x) = \Phi_t(x)$ is strictly convex and a simple calculation implies that the mapping $T_t = I + t\nabla\phi$ is $(1-t)$-monotone (cf. [31], Chapter 6), consequently it has a left inverse denoted by S_t. Let us denote by Ψ_t the Legendre transformation of Φ_t:

$$\Psi_t(y) = \sup_{x\in\mathbb{R}^n} \left\{(x, y) - \Phi_t(x)\right\}\,.$$

A simple calculation shows that

$$\begin{aligned} \Psi_t(y) &= \sup_x \left[(1-t)\left\{(x, y) - \frac{|x|^2}{2}\right\} + t\left\{(x, y) - \frac{|x|^2}{2} - \phi(x)\right\}\right] \\ &\leq (1-t)\frac{|y|^2}{2} + t\Psi_1(y)\,. \end{aligned}$$

Since Ψ_1 is the Legendre transformation of $\Phi_1(x) = |x|^2/2 + \phi(x)$ and since $L \in \mathbb{L}\log\mathbb{L}(\mu)$, it is finite on a convex set of full measure, hence it is finite everywhere. Consequently $\Psi_t(y) < \infty$ for any $y \in \mathbb{R}^n$. Since a finite, convex function is almost everywhere differentiable, $\nabla\Psi_t$ exists almost everywhere on and it is equal almost everywhere on $T_t(M_t)$ to the left inverse T_t^{-1}, where M_t is the set of non-degeneracy of $I_{\mathbb{R}^n} + t\nabla_a^2\phi$. Note that $\mu(M_t) = 1$. The strict convexity implies that T_t^{-1} is Lipschitz with a Lipschitz constant $\frac{1}{1-t}$. Let now Λ_t be the Gaussian Jacobian

$$\Lambda_t = \det{}_2\left(I_{\mathbb{R}^n} + t\nabla_a^2\phi\right)\exp\left\{-t\mathcal{L}_a\phi - \frac{t^2}{2}|\nabla\phi|_{\mathbb{R}^n}^2\right\}\,.$$

Since the domain of ϕ is the whole space $\mathbb{R}^n$, $\Lambda_t > 0$ almost surely, hence, as we have explained above, it follows from the change of variables formula of [22] that $T_t\mu$ is absolutely continuous with respect to μ and that

$$\frac{1}{\Lambda_t} = L_t \circ T_t \,,$$

μ-almost surely.
Let us come back to the infinite dimensional case: we first give an inequality which can be proven with the help of the finite dimensional observations that we have made above:

Theorem 7.1 *Assume that (W,μ,H) is an abstract Wiener space, assume that $K, L \in \mathbb{L}^1_+(\mu)$ with $K > 0$ almost surely and denote by $T : W \to W$ the transfer map $T = I_W + \nabla\phi$, which maps the measure $Kd\mu$ to the measure $Ld\mu$. Then the following inequality holds:*

$$\frac{1}{2}E[|\nabla\phi|_H^2] \leq E[-\log K + \log L \circ T] \,. \tag{7.13}$$

Proof: Let us define k as $k = K \circ T^{-1}$, then for any $f \in C_b(W)$, we have

$$\begin{aligned} \int_W f(y)L(y)d\mu(y) &= \int_W f \circ T(x)K(x)d\mu(x) \\ &= \int_W f \circ T(x)k \circ T(x)d\mu(x) \,, \end{aligned}$$

hence

$$T\mu = \frac{L}{k} \,.\mu \,.$$

It then follows from the inequality 3.3 that

$$\begin{aligned} \frac{1}{2}E\left[|\nabla\phi|_H^2\right] &\leq E\left[\frac{L}{k}\log\frac{L}{k}\right] \\ &= E\left[\log\frac{L \circ T}{k \circ T}\right] \\ &= E[-\log K + \log L \circ T] \,. \end{aligned}$$

□

To solve the Monge-Ampère equation in the infinite dimensional case we shall try to pass to the limit from finite to infinite dimensions. To do so we need some preparations which are explained along the following lines. Suppose that $\phi \in \mathbb{D}_{2,1}$ is a 1-convex Wiener functional. Let V_n be the sigma algebra generated by $\{\delta e_1, \dots, \delta e_n\}$, where $(e_n, n \geq 1)$ is an orthonormal basis of the Cameron-Martin space H. Then $\phi_n = E[\phi|V_n]$ is again 1-convex [15], hence $\mathcal{L}\phi_n$ is a measure. However the sequence $(\mathcal{L}\phi_n, n \geq 1)$ converges to $\mathcal{L}\phi$ only in $\mathbb{D}'$. Consequently, there is no reason for the limit $\mathcal{L}\phi$ to be a measure. In case this happens, we shall denote the Radon-Nikodym density with respect to μ, of the absolutely continuous part of this measure by $\mathcal{L}_a\phi$.

Lemma 7.1 *Let $\phi \in \mathbb{D}_{2,1}$ be 1-convex and let V_n be defined as above and define $F_n = E[\phi|V_n]$. Then the sequence $(\mathcal{L}_a F_n, n \geq 1)$ is a submartingale, where $\mathcal{L}_a F_n$ denotes the μ-absolutely continuous part of the measure $\mathcal{L}F_n$.*

Proof: Note that, due to the 1-convexity, we have $\mathcal{L}_a F_n \geq \mathcal{L}F_n$ for any $n \in \mathbb{N}$. Let $X_n = \mathcal{L}_a F_n$ and $f \in \mathbb{D}$ be a positive, V_n-measurable test function. Since $\mathcal{L}E[\phi|V_n] = E[\mathcal{L}\phi|V_n]$, we have

$$\begin{aligned} E[X_{n+1} f] &\geq \langle \mathcal{L}F_{n+1}, f\rangle \\ &= \langle \mathcal{L}F_n, f\rangle \,, \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ denotes the duality bracket for the dual pair $(\mathbb{D}', \mathbb{D})$. Consequently

$$E[f\, E[X_{n+1}|V_n]] \geq \langle \mathcal{L}F_n, f\rangle \,,$$

for any positive, V_n-measurable test function f, it follows that the absolutely continuous part of $\mathcal{L}F_n$ is also dominated by the same conditional expectation and this proves the submartingale property. □

Lemma 7.2 *Assume that $L \in \mathbb{L}\log\mathbb{L}(\mu)$ is a positive random variable whose expectation is one. Assume further that it is lower bounded by a constant $a > 0$. Let $T = I_W + \nabla\phi$ be the transport map such that $T\mu = L\,.\mu$ and let $T^{-1} = I_W + \nabla\psi$. Then $\mathcal{L}\psi$ is a Radon measure on $(W, \mathcal{B}(W))$. If L is upper bounded by $b > 0$, then $\mathcal{L}\phi$ is also a Radon measure on $(W, \mathcal{B}(W))$.*

Proof: Let $L_n = E[L|V_n]$, then $L_n \geq a$ almost surely. Let $T_n = I_W + \nabla\phi_n$ be the transport map which satisfies $T_n\mu = L_n\,.\mu$ and let $T_n^{-1} = I_W + \nabla\psi_n$ be its inverse. We have

$$L_n = \det{}_2\left(I_H + \nabla_a^2\psi_n\right)\exp\left[-\mathcal{L}_a\psi_n - \frac{1}{2}|\nabla\psi_n|_H^2\right] .$$

By the hypothesis $-\log L_n \leq -\log a$. Since ψ_n is 1-convex, it follows from the finite dimensional results that $\det_2\left(I_H + \nabla_a^2\psi_n\right) \in [0,1]$ almost surely. Therefore we have

$$\mathcal{L}_a\psi_n \leq -\log a \,,$$

moreover $\mathcal{L}\psi_n \leq \mathcal{L}_a\psi_n$ as distributions, consequently

$$\mathcal{L}\psi_n \leq -\log a$$

as distributions, for any $n \geq 1$. Since $\lim_n \mathcal{L}\psi_n = \mathcal{L}\psi$ in $\mathbb{D}'$, we obtain $\mathcal{L}\psi \leq -\log a$, hence $-\log a - \mathcal{L}\psi \geq 0$ as a distribution, hence $\mathcal{L}\psi$ is a Radon measure on W, c.f., [14], [30]. This proves the first claim. Note that whenever L is upperbounded, $\Lambda = 1/L \circ T$ is lowerbounded, hence the proof of the second claim is similar to that of the first one. □

Theorem 7.2 *Assume that* L *is a strictly positive bounded random variable with* $E[L] = 1$*. Let* $\phi \in \mathbb{D}_{2,1}$ *be the* 1*-convex Wiener functional such that*

$$T = I_W + \nabla\phi$$

is the transport map realizing the measure $L\,.\mu$ *and let* $S = I_W + \nabla\psi$ *be its inverse. Define* $F_n = E[\phi|V_n]$*, then the submartingale* $(\mathcal{L}_a F_n, n \geq 1)$ *converges almost surely to* $\mathcal{L}_a\phi$*. Let* $\lambda(\phi)$ *be the random variable defined as*

$$\begin{aligned}\lambda(\phi) &= \liminf_{n\to\infty} \Lambda_n \\ &= \left(\liminf_n \det{}_2\left(I_H + \nabla_a^2 F_n\right)\right) \exp\left\{-\mathcal{L}_a\phi - \frac{1}{2}|\nabla\phi|_H^2\right\}\end{aligned}$$

where

$$\Lambda_n = \det{}_2\left(I_H + \nabla_a^2 F_n\right) \exp\left\{-\mathcal{L}_a F_n - \frac{1}{2}|\nabla F_n|_H^2\right\}.$$

Then it holds true that

$$E[f \circ T\,\lambda(\phi)] \leq E[f] \tag{7.14}$$

for any $f \in C_b^+(W)$*, in particular* $\lambda(\phi) \leq \frac{1}{L\circ T}$ *almost surely. If* $E[\lambda(\phi)] = 1$*, then the inequality in (7.14) becomes an equality and we also have*

$$\lambda(\phi) = \frac{1}{L\circ T}.$$

Proof: Let us remark that, due to the 1-convexity, $0 \leq \det_2\left(I_H + \nabla_a^2 F_n\right) \leq 1$, hence the lim inf exists. Now, Lemma 7.2 implies that $\mathcal{L}\phi$ is a Radon measure. Let $F_n = E[\phi|V_n]$, then we know from Lemma 7.1 that $(\mathcal{L}_a F_n, n \geq 1)$ is a submartingale. Let $\mathcal{L}^+\phi$ denote the positive part of the measure $\mathcal{L}\phi$. Since $\mathcal{L}^+\phi \geq \mathcal{L}\phi$, we have also $E[\mathcal{L}^+\phi|V_n] \geq E[\mathcal{L}\phi|V_n] = \mathcal{L}F_n$. This implies that $E[\mathcal{L}^+\phi|V_n] \geq \mathcal{L}_a^+ F_n$. Hence we find that

$$\sup_n E[\mathcal{L}_a^+ F_n] < \infty$$

and this condition implies that the submartingale $(\mathcal{L}_a F_n, n \geq 1)$ converges almost surely. We shall now identify the limit of this submartingale. Let $\mathcal{L}_s G$ be the singular part of the measure $\mathcal{L}G$ for a Wiener function G such that $\mathcal{L}G$ is a measure. We have

$$\begin{aligned}E[\mathcal{L}\phi|V_n] &= E[\mathcal{L}_a\phi|V_n] + E[\mathcal{L}_s\phi|V_n] \\ &= \mathcal{L}_a F_n + \mathcal{L}_s F_n\,,\end{aligned}$$

hence

$$\mathcal{L}_a F_n = E[\mathcal{L}_a\phi|V_n] + E[\mathcal{L}_s\phi|V_n]_a$$

almost surely, where $E[\mathcal{L}_s\phi|V_n]_a$ denotes the absolutely continuous part of the measure $E[\mathcal{L}_s\phi|V_n]$. Note that, from the Theorem of Jessen (cf., for example Theorem 1.2.1 of [31]), $\lim_n E[\mathcal{L}_s^+\phi|V_n]_a = 0$ and $\lim_n E[\mathcal{L}_s^-\phi|V_n]_a = 0$ almost surely, hence we have

$$\lim_n \mathcal{L}_a F_n = \mathcal{L}_a\phi\,,$$

μ-almost surely. To complete the proof, an application of Fatou's lemma implies that

$$\begin{aligned} E[f \circ T\, \lambda(\phi)] &\leq E[f] \\ &= E\left[f \circ T \,\frac{1}{L \circ T}\right], \end{aligned}$$

for any $f \in C_b^+(W)$. Since T is invertible, it follows that

$$\lambda(\phi) \leq \frac{1}{L \circ T}$$

almost surely. Therefore, in the case $E[\lambda(\phi)] = 1$, we have

$$\lambda(\phi) = \frac{1}{L \circ T},$$

and this completes the proof. □

The following result gives the existence of the subsolutions of the Monge-Ampère equation:

Corollary 7.1 *Assume that K, L are two positive random variables with values in a bounded interval $[a, b] \subset (0, \infty)$ such that $E[K] = E[L] = 1$. Let $T = I_W + \nabla\phi$, $\phi \in \mathbb{D}_{2,1}$, be the transport map pushing $Kd\mu$ to $Ld\mu$, i.e, $T(Kd\mu) = Ld\mu$. We then have*

$$L \circ T\, \lambda(\phi) \leq K,$$

μ-almost surely. In particular, if $E[\lambda(\phi)] = 1$, then T is the solution of the Monge-Ampère equation.

Proof: Since $a > 0$,

$$\frac{dT\mu}{d\mu} = \frac{L}{K \circ T} \leq \frac{b}{a}.$$

Hence, Theorem 7.14 implies that

$$\begin{aligned} E[f \circ T\, L \circ T\, \lambda(\phi)] &\leq E[f\, L] \\ &= E[f \circ T\, K], \end{aligned}$$

consequently

$$L \circ T\, \lambda(\phi) \leq K,$$

the rest of the claim is now obvious. □

The following result is important for the interpolation of measures (cf. [22]):

Theorem 7.3 *Assume that L is a positive random variable of class $\mathbb{L}\log\mathbb{L}(\mu)$ such that $E[L] = 1$. Let $\phi \in \mathbb{D}_{2,1}$ be the 1-convex function corresponding to the transport map $T = I_W + \nabla\phi$. Define $T_t = I_W + t\nabla\phi$, where $t \in [0, 1]$. Then, for any $t \in [0, 1]$, $T_t\mu$ is absolutely continuous with respect to the Wiener measure μ.*

Proof: Let ϕ_n be defined as the transport map corresponding to $L_n = E[P_{1/n}L_n|V_n]$ and define T_n as $I_W + \nabla\phi_n$. For $t \in [0,1)$, let $T_{n,t} = I_W + t\nabla\phi_n$. It follows from the finite dimensional results which are summarized in the beginning of this section that $T_{n,t}\mu$ is absolutely continuous with respect to μ. Let $L_{n,t}$ be the corresponding Radon-Nikodym density and define $\Lambda_{n,t}$ as

$$\Lambda_{n,t} = \det{}_2\left(I_H + t\nabla_a^2\phi_n\right)\exp\left\{-t\mathcal{L}_a\phi_n - \frac{t^2}{2}|\nabla\phi_n|_H^2\right\}.$$

Moreover, for any $t \in [0,1)$,

$$\left((I_H + t\nabla_a^2\phi_n)h, h\right)_H > 0\,, \tag{7.15}$$

μ-almost surely for any $0 \neq h \in H$. Since ϕ_n is of finite rank, 7.15 implies that $\Lambda_{n,t} > 0$ μ-almost surely and we have shown at the beginning of this section

$$\Lambda_{n,t} = \frac{1}{L_{n,t}\circ T_{n,t}}$$

μ-almost surely. An easy calculation shows that $t \to \log\det_2(I + t\nabla_a^2\phi_n)$ is a non-increasing function. Since $\mathcal{L}_a\phi_n \geq \mathcal{L}\phi_n$, we have $E[\mathcal{L}_a\phi_n] \geq 0$. Consequently

$$\begin{aligned}
E\left[L_{t,n}\log L_{t,n}\right] &= E\left[\log L_{n,t}\circ T_{n,t}\right] \\
&= -E\left[\log \Lambda_{t,n}\right] \\
&= E\left[-\log\det{}_2\left(I_H + t\nabla^2\phi_n\right) + t\mathcal{L}_a\phi_n + \frac{t^2}{2}|\nabla\phi_n|_H^2\right] \\
&\leq E\left[-\log\det{}_2\left(I_H + \nabla^2\phi_n\right) + \mathcal{L}_a\phi_n + \frac{1}{2}|\nabla\phi_n|_H^2\right] \\
&= E\left[L_n\log L_n\right] \\
&\leq E[L\log L]\,,
\end{aligned}$$

by Jensen's inequality. Therefore

$$\sup_n E[L_{n,t}\log L_{n,t}] < \infty$$

and this implies that the sequence $(L_{n,t}, n \geq 1)$ is uniformly integrable for any $t \in [0,1]$. Consequently it has a subsequence which converges weakly in $L^1(\mu)$ to some L_t. Since, from Theorem 4.1, $\lim_n \phi_n = \phi$ in $\mathbb{D}_{2,1}$, where ϕ is the transport map associated to L, for any $f \in C_b(W)$, we have

$$\begin{aligned}
E[f\circ T_t] &= \lim_k E\left[f\circ T_{n_k,t}\right] \\
&= \lim_k E\left[f\,L_{n_k,t}\right] \\
&= E[f\,L_t]\,,
\end{aligned}$$

hence the theorem is proved. □

7.1 The solution of the Monge-Ampère equation via Ito-renormalization

As we have already seen, the strong solution of the Monge-Ampère equation corresponding to the probability densities K and L, consists of finding a map $T : W \to W$ such that

$$L \circ T \, J(T) = K$$

almost surely, where $J(T)$ is a kind of Jacobian to be written in terms of T. In Corollary 7.1, we have shown the existence of some $\lambda(\phi)$ which gives an inequality instead of the equality. The reason for this difficulty originates from the lack of the regularity of the function $\phi \in \mathbb{D}_{2,1}$. In fact, in order to calculate the Carleman-Fredholm determinant, we need generally that $\phi \in \mathbb{D}_{p,2}$ for some $p > 1$, or at least, that the absolutely continuous part of the vector measure $\nabla^2 \phi$ is with values in the space of Hilbert-Schmidt operators on the Cameron-Martin space H. Although in the finite dimensional case there are some regularity results about the transport map (cf., [7]), in the infinite dimensional case such techniques do not work. However, as we shall show in this section, **all these difficulties can be circumvented using the miraculous renormalization of the Ito calculus**. In fact assume that K and L satisfy the hypothesis of the corollary. First let us indicate that we can assume $W = C_0([0,1], \mathbb{R})$ (cf., [31], Chapter II, to see how one can pass from an abstract Wiener space to the standard one) and in this case the Cameron-Martin space H becomes $H^1([0,1])$, which is the space of absolutely continuous functions on $[0,1]$, with a square integrable Sobolev derivative. Let now

$$\Lambda = \frac{K}{L \circ T} \,,$$

where T is as constructed above. Then $\Lambda.\mu$ is a Girsanov measure for the map T. This means that the law of the stochastic process $(t, x) \to T_t(x)$ under $\Lambda.\mu$ is equal to the Wiener measure, where $T_t(x)$ is defined as the evaluation of the trajectory $T(x)$ at $t \in [0,1]$. In other words the process $(t, x) \to T_t(x)$ is a Brownian motion under the probability $\Lambda.\mu$. Let $(\mathcal{F}_t^T, t \in [0,1])$ be its filtration, the invertibility of T implies that

$$\bigvee_{t \in [0,1]} \mathcal{F}_t^T = \mathcal{B}(W) \,.$$

Λ is upper and lower bounded μ-almost surely, hence also $\Lambda.\mu$-almost surely. The Ito representation theorem implies that it can be represented as

$$\Lambda = E[\Lambda^2] \exp\left\{ -\int_0^1 \dot{\alpha}_s dT_s - \frac{1}{2} \int_0^1 |\dot{\alpha}_s|^2 ds \right\} \,,$$

where $\alpha(\cdot) = \int_0^{\cdot} \dot{\alpha}_s ds$ is an H-valued random variable. In fact α can be calculated explicitly using the Ito-Clark representation theorem (cf., [30]), and it is given as

$$\dot{\alpha}_t = \frac{E_\Lambda[D_t \Lambda | \mathcal{F}_t^T]}{E_\Lambda[\Lambda | \mathcal{F}_t^T]} \tag{7.16}$$

$dt \times \Lambda d\mu$-almost surely, where E_Λ denotes the expectation operator with respect to $\Lambda.\mu$ and $D_t\Lambda$ is the Lebesgue density of the absolutely continuous map $t \to \nabla\Lambda(t,x)$. From the relation (7.16), it follows that α is a function of T, hence we have obtained the strong solution of the Monge-Ampère equation. Let us state all this as

Theorem 7.4 *Assume that K and L are upper and lower bounded densities, let T be the transport map constructed in Theorem 6.1. Then T is also the strong solution of the Monge-Ampère equation in the Ito sense, namely*

$$E[\Lambda^2]\, L\circ T \exp\left\{-\int_0^1 \dot\alpha_s dT_s - \frac{1}{2}\int_0^1 |\dot\alpha_s|^2 ds\right\} = K\,,$$

μ-almost surely, where $\dot\alpha$ is given by (7.16).

Acknowledgement: The authors are grateful to Françoise Combelles for all the bibliographical help that she has supplied for the realization of this research. We thank also the anonymous referee for his particular attention and valuable remarks.

References

[1] T. Abdellaoui and H. Heinich: "Sur la distance de deux lois dans le cas vectoriel". CRAS, Paris Série I, Math., **319**, 397-400, 1994.

[2] R. D. Anderson and V.L. Klee, Jr.: "Convex functions and upper semicontinuous collections". Duke Math. Journal, **19**, 349-357, 1952.

[3] P. Appell: "Mémoire sur déblais et les remblais des systèmes continus ou discontinus". Mémoires présentées par divers savants à l'Académie des Sciences de l'Institut de France. Paris, I. N. **29**, 1-208, 1887.

[4] P. Appell: "Le problème géométrique des déblais et des remblais". Mémorial des Sciences Mathématiques, fasc. **XXVII**, Paris, 1928.

[5] P. J. Bickel and D. A. Freedman: "Some asymptotic theory for the bootstrap". The Annals of Statistics, Vol. **9**, No. 6, 1196-1217, 1981.

[6] Y. Brenier: "Polar factorization and monotone rearrangement of vector valued functions". Comm. pure Appl. Math, **44**, 375-417, 1991.

[7] L. A. Caffarelli: "The regularity of mappings with a convex potential". Jour. Amer. Math. Soc., **5**, 99-104, 1992.

[8] B. Dacarogna and J. Moser: "On a partial differential equation involving the Jacobian determinant". Ann. Inst. Henri Poincaré, Analyse non-linéaire, **7**, 1-26, 1990.

[9] C. Dellacherie and P. A. Meyer: *Probabilités et Potentiel, Ch. I à IV.* Paris, Hermann, 1975.

[10] H. Djellout, A. Guillin and L. Wu: "Transportation cost-information inequalities for random dynamical systems and diffusions". Preprint, December 2002.

[11] N. Dunford and J. T. Schwartz: *Linear Operators* **2**, Interscience 1963.

[12] X. Fernique: "Extension du théorème de Cameron-Martin aux translations aléatoires", Comptes Rendus Mathématiques, Vol. **335**, Issue 1, 65-68, 2002.

[13] X. Fernique: "Comparaison aux mesures gaussiennes, espaces autoreproduisants. Une application des propriétés isopérimétriques". Preprint.

[14] D. Feyel and A. de La Pradelle: "Capacités gaussiennes". Annales de l'Institut Fourier, t.41, f.1, 49-76, 1991.

[15] D. Feyel and A. S. Üstünel: "The notion of convexity and concavity on Wiener space". Journal of Functional Analysis, **176**, 400-428, 2000.

[16] D. Feyel and A. S. Üstünel: "Transport of measures on Wiener space and the Girsanov theorem". Comptes Rendus Mathématiques, Vol. **334**, Issue 1, 1025-1028, 2002.

[17] W. Gangbo and R. J. McCann: "The geometry of optimal transportation". Acta Mathematica, **177**, 113-161, 1996.

[18] K. Ito and M. Nisio: "On the convergence of sums of independent Banach space valued random variables". Osaka Journ. Math. **5**, 35-48, 1968.

[19] L. V. Kantorovitch: "On the transfer of masses". Dokl. Acad. Nauk. SSSR **37**, 227-229, 1942.

[20] K. Marton: "Bounding $\bar{d}$-distance by informational divergence: a method to prove measure concentration". Annals of Probability, **24**, no.2, 857-866, 1996.

[21] R. J. McCann: "Existence and uniqueness of monotone measure-preserving maps". Duke Math. Jour., **80**, 309-323, 1995.

[22] R. J. McCann: "A convexity principle for interacting gases". Advances in Mathematics, **128**, 153-179, 1997.

[23] G. Monge: "Mémoire sur la théorie des déblais et des remblais". Histoire de l'Académie Royale des Sciences, Paris, 1781.

[24] S. T. Rachev; "The Monge-Kantorovitch transference problem". Th. prob. Appl.**49**, 647-676, 1985.

[25] T. Rockafellar: *Convex Analysis*. Princeton University Press, Princeton, 1972.

[26] V. N. Sudakov: "Geometric problems in the theory of infinite dimensional probability distributions". Proc. Steklov Inst. Math., **141**, 1-178, 1979.

[27] M. Talagrand: "Transportation cost for Gaussian and other product measures". Geom. Funct. Anal., **6**, 587-600, 1996.

[28] E. Thomas: "The Lebesgue-Nikodym theorem for vector valued Radon measures". Memoirs of A.M.S., **139**, 1974.

[29] A. S. Üstünel: "Representation of distributions on Wiener space and Stochastic Calculus of Variations". Journal of Functional Analysis, **70**, 126-139, 1987.

[30] A. S. Üstünel: *Introduction to Analysis on Wiener Space*. Lecture Notes in Math. Vol. **1610**. Springer, 1995.

[31] A. S. Üstünel and M. Zakai: *Transformation of Measure on Wiener Space*. Springer Monographs in Mathematics. Springer Verlag, 1999.

- D. Feyel, Université d'Evry-Val-d'Essone, 91025 Evry Cedex, France. E-mail: feyel@maths.univ-evry.fr
- A. S. Üstünel, ENST, Dépt. Infres, 46, rue Barrault, 75634 Paris Cedex 13, France. E-mail: ustunel@enst.fr