Regularity of the backward Monge potential and the Monge–Ampère equation on Wiener space

by

MINE ÇAĞLAR and İHSAN DEMIREL (Istanbul)

Abstract. In this paper, the Monge–Kantorovich problem is considered in infinite dimensions on an abstract Wiener space $(W, H, \mu)$, where $H$ is the Cameron–Martin space and $\mu$ is the Gaussian measure. We study the regularity of optimal transport maps with a quadratic cost function assuming that both initial and target measures have a strictly positive Radon–Nikodym density with respect to $\mu$. Under some conditions on the density functions, the forward and backward transport maps can be written in terms of Sobolev derivatives of so-called Monge–Brenier maps, or Monge potentials. We show the Sobolev regularity of the backward potential under the assumption that the density of the initial measure is log-concave and prove that the backward potential solves the Monge–Ampère equation.

1. Introduction. The Monge problem is motivated by the problem of moving given piles of sand to fill up holes of the same volume with a minimum total cost of transportation\[1\] [25]. Piles and holes are modeled by probability measures $\rho$ and $\nu$ defined on some measurable sets $X$ and $Y$, respectively, with a measurable and non-negative cost function $c : X \times Y \to \mathbb{R}_+ \cup \{+\infty\}$. The aim is to find a transport map $T : X \to Y$, with $T\rho = \nu$, that minimizes the expected cost of moving the sand. A solution to the Monge problem may not always exist, as it does not allow split of mass. The Monge–Kantorovich problem overcomes this difficulty by the relaxation that mass from location $x \in X$ can be moved possibly to several locations in $Y$ with the objective of finding a transference plan $\gamma^* \in \Gamma(\rho, \nu)$ that minimizes the functional

$$J[\gamma] = \int_{X \times Y} C(x, y) \, d\gamma(x, y)$$

2020 Mathematics Subject Classification: Primary 60H07; Secondary 49J27.
Key words and phrases: Monge–Kantorovich problem, Wiener space, Monge–Ampère equation, optimal transport, logarithmic concave measure.
Received 6 September 2021; revised 14 April 2022.
Published online 3 October 2022.

DOI: 10.4064/sm210906-2-5  [1] © Instytut Matematyczny PAN, 2022
where \( \Gamma(\mu, \nu) \) is the set of joint probability measures on \( X \times Y \) with first and second marginals \( \rho \) and \( \nu \), respectively. If \( X \) and \( Y \) are Polish spaces and \( c \) is lower semicontinuous, then the Monge–Kantorovich problem admits a minimizer [2, Thm. 1.5]. In infinite dimensions, the Monge–Kantorovich problem has been studied on Wiener space first in [14] with a quadratic cost function. Later in [9, 12, 20], different cost functions were considered with other initial and target measures.

In this paper, we consider an abstract Wiener space \((W, H, \mu)\), where \( H \) is the Cameron–Martin space and \( \mu \) is the Gaussian measure, and define \( d\rho = e^{-f}d\mu \) and \( d\nu = e^{-g}d\mu \) for measurable \( f, g : W \to \mathbb{R} \). Under some conditions on \( f \) and \( g \), there exists a so-called forward Monge potential or Monge–Brenier map \( \phi \) such that \( T = I_W + \nabla \phi \) is the optimal transport map, where \( \nabla \) denotes the Gross–Sobolev derivative operator. Moreover, the inverse \( S \) of \( T \) is given by \( S = I_W + \nabla \psi \) for some \( \psi : W \to \mathbb{R} \), called a backward Monge potential, satisfying \((S \times I_W)\rho = (I_W \times T)\nu = \gamma\).

We aim to show the Sobolev regularity of the Monge–Brenier map \( \psi \) in such a way that we can write the Jacobian function associated with the corresponding transformation \( S \). In Theorem 4.6, assuming \( f \) is \((1 - c)\)-convex, equivalently \( e^{-f} \) is log-concave, we show the Sobolev regularity of \( \psi \) in the sense that \( \nabla^2 \psi \) is in \( L^2(\nu, H \otimes H) \) and provide an upper estimate for its norm as

\[
\mathbb{E}_\nu[\|\nabla^2 \psi\|_2^2] \leq \frac{3}{c}(\mathbb{E}_\rho[\|\nabla \phi\|_H^2] + \mathbb{E}_\nu[\|\nabla g\|_H^2] + \mathbb{E}_\rho[\|\nabla f\|_H^2]),
\]

where \( H \otimes H \) denotes the space of Hilbert–Schmidt operators on \( H \), and \( \| \cdot \|_2 \) denotes the Hilbert–Schmidt norm. We use the approximation approach of [23], where the initial measure \( \rho \) is taken as \( \mu \) and the target measure \( \nu \) is assumed to be absolutely continuous with respect to \( \mu \).

Monge potentials are closely related to the Monge–Ampère equation defined in Wiener space. On \( \mathbb{R}^n \), with \( F, G : \mathbb{R}^n \to \mathbb{R}_+ \), the Monge–Ampère equation in \( T \) can be written as

\[
(1.1) \quad F = G \circ T \det J_T
\]

where \( J_T \) is the Jacobian of \( T \). When \( F \) and \( G \) are the densities of \( \rho \) and \( \nu \) with respect to Lebesgue measure, respectively, the corresponding transport map \( T \) is a solution. In finite dimensions, regularity of solution of the Monge–Kantorovich problem and the Monge–Ampère equation have been studied in [8, 10, 11, 19].

In order to define an analogous equation in Wiener space, suppose \( F = e^{-f} \) and \( G = e^{-g} \). In other words, the Radon–Nikodym derivatives of \( \rho \) and \( \nu \) with respect to Lebesgue measure are proportional to \( e^{-f - |x|^2/2} \)
and \( e^{-g-|x|^2/2} \) respectively for \( x \in \mathbb{R}^n \) and (1.1) becomes

\[
e^{-f(x)-|x|^2/2} = e^{-g \circ T(x)-|T(x)|^2/2} \det J_T(x).
\]

On \( \mathbb{R}^n \), we have \( T = I_{\mathbb{R}^n} + \nabla \varphi \). Using the identities \( J_T = I_{\mathbb{R}^n} + \nabla^2 \varphi \) and \( \det(I_{\mathbb{R}^n} + \cdot) = \det_2(I_{\mathbb{R}^n} + \cdot)e^{\text{tr}(\cdot)} \), where \( \det_2 \) denotes the Carleman–Fredholm determinant, and \( \text{tr} \) denotes the trace of a matrix, we get

\[
e^{-f} = e^{-g \circ T} \det_2(I_{\mathbb{R}^n} + \nabla^2 \varphi) e^{\text{tr}(\nabla^2 \varphi)} e^{-x \nabla \varphi} e^{-|\nabla \varphi|^2/2}.
\]

In view of the identity \(-\mathcal{L} \varphi = \Delta \varphi - x \nabla \varphi = \text{tr}(\nabla^2 \varphi) - x \nabla \varphi \), where \( \mathcal{L} \) is the generator of the Ornstein–Uhlenbeck process on \( \mathbb{R}^n \) and \( \Delta \) denotes the Laplacian, the Monge–Ampère equation can now be written as

\[
e^{-f} = e^{-g \circ T} \det_2(I_{\mathbb{R}^n} + \nabla^2 \varphi) \exp[-\mathcal{L} \varphi - \frac{1}{2}|\nabla \varphi|^2].
\]

Since only \( \det_2 \) is well-defined in infinite dimensions, this form can be used to define the Monge–Ampère equation in Wiener space by replacing \(| \cdot |\) with the norm in \( H \). In [14], it is proved that the Monge potential \( \varphi \) solves the Monge–Ampère equation under the condition \( f \) and \( g \) are positive random variables with values in a bounded interval \([a, b]\). The Monge–Ampère equation for \( \varphi \) has been obtained in [15][16] with \( f = 0 \) and \( g \) an \( H \)-convex Wiener function. In [5][6], the authors have shown that \( \varphi \) satisfies the Monge–Ampère equation when \( g \in L^1(\mu) \) and \( ge^{-g} \in L^1(\mu) \), and when \( e^{-f/2} \in \mathbb{D}_{2,1} \) and \( g = 0 \). The case when both the initial and target measures are absolutely continuous with respect to \( \mu \) is studied in [12] for bounded \( f \) and lower bounded \( g \) with the additional assumption that the second order Sobolev derivative of \( g \) exists.

In the setting of the present paper, the backward potential is shown to be regular in Theorem 4.6 that is, \( \nabla^2 \psi \in L^2(\nu, H \otimes H) \) under the assumption that the initial measure \( \rho \) has a log-concave density with respect to the Gaussian measure \( \mu \). Therefore, we prove in Theorem 4.11 that the backward potential \( \psi \) solves the Monge–Ampère equation given by

\[
e^{-g} = e^{-f \circ S} \det_2(I_H + \nabla^2 \psi) \exp[-\mathcal{L} \psi - \frac{1}{2}||\nabla \psi||_H^2]
\]

\( \nu \)-almost surely.

The organization of the paper is as follows. In Section 2, we review the preliminary definitions and introduce the notation used in the paper. Section 3 focuses on the finite-dimensional and smooth case as a basis for the infinite-dimensional case. In Section 4, the regularity of the backward potential is shown and the Monge–Ampère equation is considered.

2. Preliminaries. Let \((W, H, \mu)\) be an abstract Wiener space. The corresponding Cameron–Martin space is denoted by \( H \). The norm in \( H \) will be denoted by \(| \cdot |_H \). If \( h \in H \), then there exists \((l_n) \subset W^*\) such that the image of this sequence under the injection \( W^* \hookrightarrow H \), say \((l_n)\), converges to \( h \) in \( H \).
Therefore, the sequence \((\langle l_n, \cdot \rangle_H)\) of random variables is Cauchy in \(L^p(\mu)\) for any \(p \geq 0\). We denote its limit by \(\delta h\), which is an \(N(0, |h|^2_H)\) random variable.

A function \(F : W \to \mathbb{R}\) is called a cylindrical Wiener functional if it is of the form

\[
F(\omega) = f(\delta h_1(\omega), \ldots, \delta h_n(\omega)), \quad h_1, \ldots, h_n \in H, \; f \in S(\mathbb{R}^n),
\]

for some \(n \in \mathbb{N}\), where \(S(\mathbb{R}^n)\) denotes the Schwartz space of rapidly decreasing functions on \(\mathbb{R}^n\) \([14, 22]\). We denote the collection of Wiener functionals by \(S(W)\). For \(F \in S(W)\) and \(h \in H\), we define

\[
\nabla_h F(\omega) = \frac{d}{d\epsilon} F(\omega + \epsilon h) \bigg|_{\epsilon=0}.
\]

For fixed \(\omega \in W\), \(h \mapsto \nabla_h F(\omega)\) is continuous and linear on \(H\). Therefore, there exists an element in \(H\), say \(\nabla F\), such that \(\nabla_h F = \langle \nabla F, h \rangle_H\). The operator \(\nabla : F \mapsto \nabla F\) is linear from \(S(W)\) into the space of \(H\)-valued Wiener functionals \(L^p(\mu; H)\) for any \(p > 1\). The operator \(\nabla\) is closable from \(L^p(\mu)\) into \(L^p(\mu; H)\) for any \(p > 1\). The completion of \(S(W)\) under the norm

\[
\| \cdot \|_{p,1} = \| \cdot \|_{L^p(\mu)} + \| \cdot \|_{L^p(\mu; H)}
\]

is denoted by \(\mathbb{D}_{p,1}\), which is a Banach space with norm \(\| \cdot \|_{p,1}\) \([21]\). The definition can also be extended to the collection of Wiener functionals which take values in a separable Hilbert space \(X\). The completion of the collection \(S(W; X)\) of \(X\)-valued Wiener functionals under the norm

\[
\| \cdot \|_{L^p(\mu; X)} + \| \cdot \|_{L^p(\mu; X \otimes H)}
\]

is denoted by \(\mathbb{D}_{p,1}(X)\). Higher order derivatives can also be defined, e.g. we say \(F \in \mathbb{D}_{p,2}\) if \(\nabla^2 F \in \mathbb{D}_{p,1}(H)\) and write \(\nabla^2 F = \nabla(\nabla F)\).

Let \(\nu\) be a measure on \(W\) absolutely continuous with respect to \(\mu\) with Radon–Nikodym derivative \(L\). If \(\int_W |\nabla L|_H^2 e^{-L} \, d\mu < \infty\), then the operator \(\nabla\) is closable over \(S(W)\) under the norm

\[
\| \cdot \|_{L^p(\nu)} + \| \cdot \|_{L^p(\nu; H)}.
\]

We will denote the completion of \(S(W)\) by \(\mathbb{D}_{p,1}(\nu)\). The adjoint of the continuous linear operator \(\nabla\) will be denoted by \(\delta\). That is, for suitable \(\xi : W \to H\) and any \(F \in \mathbb{D}_{p,1}\), we have

\[
\mathbb{E}[\langle \nabla F, \xi \rangle_H] = \mathbb{E}[F \cdot \delta \xi].
\]

The operator \(\delta : \xi \mapsto \delta \xi\) is called the divergence operator. We can also define \(\delta_\nu\) by this procedure, i.e. \(\delta_\nu\) is the adjoint of the Sobolev derivative \(\nabla\) under the measure \(\nu\).
For measurable $f : W \to \mathbb{R}$ and $t \geq 0$ the Ornstein–Uhlenbeck semigroup $(P_t)_{t \geq 0}$ is given by

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

Its infinitesimal generator is denoted by $-\mathcal{L}$ and we call $\mathcal{L}$ the Ornstein–Uhlenbeck operator. The norm given by $\| (I + \mathcal{L})^{r/2}(\cdot) \|_{L^p(\mu)}$ is equivalent to the norm $\| \cdot \|_{p,r}$ for any $p > 1$ and $r \in \mathbb{N}$. The completion of $S(W)$ with respect to the norm $\| (I + \mathcal{L})^{r/2} \|_{L^p(\mu)}$ is again denoted by $\mathbb{D}_{p,r}$ and has $\mathbb{D}_{q,-r}, q^{-1} = 1 - p^{-1},$ as its continuous dual for any $p > 1, r \in \mathbb{R}$. Similarly, we can define $\mathbb{D}_{p,r}(X)$ as the completion of $S(W; X)$ with respect to the norm $\| (I + \mathcal{L})^{k/2} F \|_{L^p(\mu; X)}$ and its continuous dual is $\mathbb{D}_{q,-r}(X^*)$, where $X^*$ is the dual of $X$.

A measurable function $f : W \to \mathbb{R} \cup \{\infty\}$ is called $\alpha$-convex, $\alpha \in \mathbb{R}$, if the map $h \mapsto f(\cdot + h) + \frac{\alpha}{2} |h|_H^2$ is convex on the Cameron–Martin space $H$ with values in $L^0(\mu)$ a.s.

For the Monge–Kantorovich problem in Wiener space, the cost function $C(x, y) : W \times W \to \mathbb{R}^+ \cup \{+\infty\}$ is defined by

$$C(x, y) = \begin{cases} |x - y|_H^2 & \text{if } x - y \in H, \\ +\infty & \text{otherwise} \end{cases}$$

in [14]. The following result is the starting point of the present work [12, Thm. 1.1].

**Theorem 2.1.** Let $\rho, \nu$ be probability measures on $W$ such that $d\rho = F d\mu$ and $d\nu = G d\mu$, where $\mu$ is the Wiener measure, $F : W \to \mathbb{R}$ and $G : W \to \mathbb{R}$ are measurable functions such that

$$\int_W \frac{\nabla F^2_H}{F} \, d\mu < \infty \quad \text{and} \quad \int_W \frac{\nabla G^2_H}{G} \, d\mu < \infty$$

and the measure $\rho$ satisfies the Poincaré inequality, that is, for every cylindrical functional $\xi : W \to \mathbb{R}$,

$$(1 - c) \int_W (\xi - \mathbb{E}_\rho[\xi])^2 \, d\rho \leq \int_W |\nabla \xi|^2_H \, d\rho$$

for some $c \in [0, 1)$. Then there exists a $\varphi \in \mathbb{D}_{2,1}(\rho)$ such that $T = I_W + \nabla \varphi$ is the unique solution of the Monge problem for $(\rho, \nu)$ and the probability measure $\gamma$ given by $\gamma = (I_W \times T) \rho$ is the unique solution of the Monge–Kantorovich problem for $(\rho, \nu)$. Moreover, $T$ is $\rho$-a.s. invertible and the inverse map $S = T^{-1}$ has the form $S = I_W + \nabla \psi$, where $\psi \in \mathbb{D}_{2,1}(\nu)$. 

Finally, let $\mathcal{P}(X)$ denote the probability measures on a measurable space $X$. For $m_1, m_2 \in \mathcal{P}(X)$ on a Polish space $(X,d)$, the Wasserstein distance of order $p \in [1, \infty)$ between $m_1$ and $m_2$ is defined by

\begin{equation}
    d_p(m_1, m_2) = \left( \inf_{\gamma \in \Gamma(m_1,m_2)} \int_X d(x,y)^p \, d\gamma(x,y) \right)^{1/p}.
\end{equation}

Note that $d_p$ is not a metric, since it can take the value of $+\infty$. However, its restriction to a subset of $\mathcal{P}(X) \times \mathcal{P}(X)$ where it takes finite values leads to a complete metric space. The relative entropy of $m_1$ with respect to $m_2$ is given by

\begin{equation}
    H(m_1 \mid m_2) = \begin{cases} \int_X \frac{dm_1}{dm_2} \log \left( \frac{dm_1}{dm_2} \right) \, dm_2 & \text{if } m_1 \ll m_2, \\ +\infty & \text{otherwise.} \end{cases}
\end{equation}

We take $X = W$ and $d(x,y) = |x - y|_H$ and $p = 2$. The following form of Talagrand’s inequality, derived in [14, Thm. 3.1], provides a nice relation between the Wasserstein distance and the relative entropy. For any probability measure $\mu_0$ on $W$, we have

\begin{equation}
    d_2(\mu_0, \mu)^2 \leq 2H(\mu_0 \mid \mu).
\end{equation}

### 3. An estimate for a smooth backward Monge potential.

In this section, we assume that a forward Monge potential $\varphi$ is smooth, that is, $\varphi \in D^2_{2,k}$ for all $k \in \mathbb{N}$. We first prove an identity for the forward Monge potential $\varphi$ and then find an upper bound for the $D^2_{2,2}(\nu)$-norm of $\psi$, which will be auxiliary for regularity in infinite dimensions as shown in the next section. Similar results have been obtained in [23] when the initial measure is Gaussian and the target measure is absolutely continuous with respect to the Gaussian measure.

**Remark 3.1.** When $f$ and $g$ are smooth and bounded from below, and $\mu$ is a standard Gaussian measure on $\mathbb{R}^n$, then $\varphi$ is smooth. This follows from [18] and [25, Thm. 4.14].

**Proposition 3.2.** When $W$ is finite-dimensional and $\varphi$ is smooth, the Monge potential $\varphi$ satisfies the relation

\[ \nabla \varphi + \nabla g \circ T - \nabla f = \delta_\rho[(I_H + \nabla^2 \varphi)^{-1} - I_H]. \]

**Proof.** For $\xi \in D_{2,1}(H)$ with $\|\nabla \xi\|_2$ is in the space of bounded random variables $L^\infty(\rho)$ and sufficiently small $\epsilon > 0$, let $T_\epsilon = I_W + \nabla \varphi + \epsilon \xi$. We have

\[ e^{-f} = e^{-g \circ T_\epsilon} A_{T_\epsilon}, \]

where $e^{-g_\epsilon} = dT_\epsilon \rho / d\rho$ [19] and

\[ A_{T_\epsilon} = \det_2(I_H + \nabla^2 \varphi + \epsilon \nabla \xi) \exp(-\mathcal{L}(\varphi) + \epsilon \delta \xi - \frac{1}{2} |\nabla \varphi + \epsilon \xi|_H^2). \]

Consider the change of variable formula

\begin{equation}
    \mathbb{E}[e^{-g}] = \mathbb{E}[e^{-g \circ T_\epsilon} A_{T_\epsilon}].
\end{equation}
Observe that
\[
\frac{d}{d\epsilon} \mathbb{E}_\rho[e^{-g}] \big|_{\epsilon=0} = 0, \quad \frac{d}{d\epsilon} \mathbb{E}_\rho[g \circ T_\epsilon] \big|_{\epsilon=0} = \mathbb{E}_\rho[\langle \nabla g \circ T, \xi \rangle_H],
\]
\[
\frac{d}{d\epsilon} \mathbb{E}_\rho[-\log \Lambda_{T_\epsilon}] \big|_{\epsilon=0} = \mathbb{E}_\rho[-\text{tr}((I_H + \nabla^2 \varphi)^{-1} - I_H) \cdot \nabla \xi] + \mathbb{E}_{\rho}[\delta \xi + \langle \nabla \varphi, \xi \rangle_H]
\]
\[
= \mathbb{E}_\rho[-\langle((I_H + \nabla^2 \varphi)^{-1} - I_H), \nabla \xi \rangle_2 + \langle \nabla \varphi, \xi \rangle_H] + \mathbb{E}_\rho[\langle \nabla f, \xi \rangle_H],
\]
where \(\langle \cdot, \cdot \rangle_2\) denotes the Hilbert–Schmidt operator inner product. Here, the first derivative is 0 as the term does not depend on \(\epsilon\), the second derivative follows by definition, and \([24, \text{Thm. A.2.2}]\) is used for the last derivative. Then, taking the derivative of both sides of the change of variable formula \((3.1)\) we obtain
\[
0 = \frac{d}{d\epsilon} \mathbb{E}[e^{-g\circ T_\epsilon}] \bigg|_{\epsilon=0} = \mathbb{E}[\langle \nabla g \circ T, \xi \rangle_H e^{-g\circ T} \Lambda_{T_\epsilon}] + \mathbb{E}[-\text{tr}((I_H + \nabla^2 \varphi)^{-1} - I_H) \cdot \nabla \xi] + \mathbb{E}[(\langle \nabla \varphi, \xi \rangle_H + \delta \xi) e^{-g\circ T} \Lambda_{T_\epsilon}]
\]
\[
= \mathbb{E}_\rho[\langle \nabla g \circ T, \xi \rangle_H] + \mathbb{E}_\rho[-\text{tr}((I_H + \nabla^2 \varphi)^{-1} - I_H) \cdot \nabla \xi] + \mathbb{E}_{\rho}[\delta \xi + \langle \nabla \varphi, \xi \rangle_H]
\]
\[
= \mathbb{E}_\rho[\langle \nabla g \circ T, \xi \rangle_H] + \mathbb{E}_\rho[-\langle((I_H + \nabla^2 \varphi)^{-1} - I_H), \nabla \xi \rangle_2 + \langle \nabla \varphi, \xi \rangle_H] + \mathbb{E}_\rho[\langle \nabla f, \xi \rangle_H],
\]
and the result follows. □

**Remark 3.3.** The proof of Proposition \([3.2]\) is quite simplified in view of the assumption that \(f\) and \(g\) are smooth. Otherwise, one can follow the proof of \([23, \text{Thm. 1}]\) based on the parametrization of a variational formula.

We need the following technical lemmas in order to prove the regularity of a backward Monge potential. We refer to \([23]\) for their proofs.

**Lemma 3.4.** Suppose \(W\) is finite-dimensional and \(\varphi\) is smooth. Let \(K = (I_H + \nabla^2 \phi)^{-1}\) and \(h \in H\). Then
\[
\text{tr}(K \nabla^3 \varphi K \cdot K \nabla^3 \varphi K h) \geq 0 \quad \mu\text{-a.s.}
\]

**Lemma 3.5.** If \(\xi : W \to H\) is smooth, then \(\delta \rho \xi = \delta \xi + \langle \nabla f, \xi \rangle_H\).

**Lemma 3.6.** If \(\xi : W \to H\) is smooth, then
\[
\mathbb{E}_\rho[(\delta \rho \xi)^2] = \mathbb{E}_\rho[(I_H + \nabla^2 g, \xi \circ \xi)_{H \otimes 2} + \text{tr}(\nabla \xi \cdot \nabla \xi)].
\]
PROPOSITION 3.7. Suppose $W$ is finite-dimensional, $\varphi$ is smooth, and the function $f$ is $(1-c)$-convex for some $c \in [0, 1)$. Then

$$c\mathbb{E}_\nu[|\nabla^2 \psi|^2_2] \leq 3(\mathbb{E}_\rho[|\nabla \varphi|^2_H] + \mathbb{E}_\nu[|\nabla g|^2_H] + \mathbb{E}_\rho[|\nabla f|^2_H]).$$

Proof. Taking the expectation of the square of the norm of the expression in Proposition 3.2 yields

$$\mathbb{E}_\rho[\delta_\rho((I_H + \nabla^2 \varphi)^{-1} - I_H)^2_H]$$

$$\leq 3(\mathbb{E}_\rho[|\nabla \varphi|^2_H] + \mathbb{E}_\rho[|\nabla g \circ T|^2_H] + \mathbb{E}_\rho[|\nabla f|^2_H])$$

$$= 3(\mathbb{E}_\rho[|\nabla \varphi|^2_H] + \mathbb{E}_\nu[|\nabla g|^2_H] + \mathbb{E}_\rho[|\nabla f|^2_H]).$$

Denote $(I_H + \nabla^2 \varphi)^{-1}$ by $M$. Applying Lemma 3.6, we get

$$\mathbb{E}_\rho[|\delta_\rho(M - I_H)|^2_H] = \sum_{k=1}^{\infty} \mathbb{E}_\rho[(\delta_\rho(M - I_H)(e_k))^2]$$

$$= \sum_{k=1}^{\infty} \mathbb{E}_\rho[(I_H + \nabla^2 f, (M - I_H)e_k \otimes (M - I_H)e_k)_{H \otimes H}]$$

$$+ \sum_{k=1}^{\infty} \mathbb{E}_\rho[\text{tr}((Me_k) \cdot \nabla(Me_k))].$$

By Lemma 3.4, the second term on the right hand side is positive, so

$$\mathbb{E}_\rho[|\delta_\rho(M - I_H)|^2_H]$$

$$\geq \sum_{k=1}^{\infty} \mathbb{E}_\rho[(I_H + \nabla^2 f, (M - I_H)e_k \otimes (M - I_H)e_k)_{H \otimes H}]$$

$$\geq c \sum_{k=1}^{\infty} \mathbb{E}_\rho[(M - I_H)e_k]^2_H] = c\mathbb{E}_\rho[(M - I_H)|^2_H].$$

Since $T = I_W + \nabla \varphi$ and $S = I_W + \nabla \psi$ are inverses of each other, we have

$$S = (I_H + \nabla^2 \psi)^{-1} = (I_H + \nabla^2 \psi) \circ T.$$  (3.4)

Combining (3.2)–(3.4), we get

$$c\mathbb{E}_\nu[|\nabla^2 \psi|^2_2] \leq 3(\mathbb{E}_\rho[|\nabla \varphi|^2_H] + \mathbb{E}_\nu[|\nabla g|^2_H] + \mathbb{E}_\rho[|\nabla f|^2_H]).$$

4. Regularity and Monge–Ampère equation. Recall that $f : W \to \mathbb{R}$ and $g : W \to \mathbb{R}$ are measurable functions such that $f, g \in D_{2,1}$ and

$$\int_W |\nabla f|^2 e^{-f} d\mu < \infty \quad \text{and} \quad \int_W |\nabla g|^2 e^{-g} d\mu < \infty$$

for the initial and target measures of the Monge–Kantorovich problem, which are $d\rho = e^{-f} d\mu$ and $d\nu = e^{-g} d\mu$, with $\rho$ satisfying the Poincaré inequality (2.1). In this section, we will show that the backward Monge potential $\psi$
solves the Monge–Ampère equation, that is, we have
\[ e^{-g} = e^{-f \circ S} \det_2 (I_H + \nabla^2 \psi) \exp \left[ -\mathcal{L} \psi - \frac{1}{2} |\nabla \psi|_H^2 \right] \]
where \( S = T^{-1} = I_W + \nabla \psi \). However, we only know \( \psi \in D_{2,1}(\nu) \) so far. Therefore, we first establish an upper bound for the \( D_{2,2}(\nu) \)-norm of \( \psi \), which implies that the second Sobolev derivative of \( \psi \) is well-defined, in Sections 4.1 and 4.2. The Monge–Ampère equation is considered in Section 4.3.

4.1. Approximation lemmas. In order to show the regularity of the backward Monge potential \( \psi \) in the general setting of this section, we will approximate \( f \) and \( g \) of (4.1). This will be accomplished through several lemmas. In Lemma 4.1, we define a sequence of probability measures \( \rho_n \) and \( \nu_n \) which are absolutely continuous with respect to the standard Gaussian measure on \( \mathbb{R}^n \), that is, in finite dimensions.

**Lemma 4.1.** Let \( (\varphi, \psi) \) be Monge potentials associated to the Monge–Kantorovich problem \( (\rho, \nu) \), where \( d\rho = e^{-f} d\mu \) and \( d\nu = e^{-g} d\mu \), and let \( f, g \in D_{2,1}(\nu) \) satisfy (4.1) and the measure \( \rho \) satisfy the Poincaré inequality (2.1). Define \( g_n \) and \( f_n \) as
\[ e^{-f_n} = \mathbb{E}[e^{-f}|V_n] \quad \text{and} \quad e^{-g_n} = \mathbb{E}[e^{-g}|V_n] \]
where \( V_n \) is generated by \{\( \delta e_1, \ldots, \delta e_n \)\} and \( \{e_i, i \geq 1\} \) is an orthogonal basis of \( H \). Let \( (\varphi_n, \psi_n) \) be Monge potentials associated with the Monge–Kantorovich problem \( (\rho_n, \nu_n) \), where
\[ d\rho_n = e^{-f_n} d\beta \quad \text{and} \quad d\nu_n = e^{-g_n} d\beta \]
and \( \beta \) is the standard Gaussian measure on \( \mathbb{R}^n \). Then \( (\varphi_n) \) converges to \( \varphi \) in \( D_{2,1}(\rho) \), \( (\psi_n) \) converges to \( \psi \) in \( L^1(\nu) \) and \( (\nabla \psi_n) \) converges to \( \nabla \psi \) in \( L^2(\nu) \).

**Proof.** This is proved in [12, Thm. 1.1]. ■

As the next step, we will define smooth functions \( f_m \) and \( g_m \) in the following lemma using the Ornstein–Uhlenbeck semigroup to approximate functions \( f \) and \( g \) given in finite dimensions. The sequences \( (f_m) \) and \( (g_m) \) will be used later to approximate the sequences of Lemma 4.1 which are in finite dimensions.

**Lemma 4.2.** Let \( \beta \) be the standard Gaussian measure on \( \mathbb{R}^n \), and let \( f \in D_{2,1}(\beta) \) and \( g \in D_{2,1}(\beta) \) be such that
\[ \int_{\mathbb{R}^d} |\nabla f|^2 e^{-f} d\beta < \infty \quad \text{and} \quad \int_{\mathbb{R}^d} |\nabla g|^2 e^{-g} d\beta < \infty. \]
Let \( (\varphi, \psi) \) be Monge potentials associated with the Monge–Kantorovich problem \( \Gamma(\rho, \nu) \), where
\[ d\rho = e^{-f} d\beta \quad \text{and} \quad d\nu = e^{-g} d\beta. \]
Define \( f_m \) and \( g_m \) via
\[
e^{-f_m} = Q_{1/m} e^{-f} \quad \text{and} \quad e^{-g_m} = Q_{1/m} e^{-g}
\]
where \((Q_t, t \geq 0)\) is the Ornstein–Uhlenbeck semigroup on \( \mathbb{R}^d\). Let \((\varphi_m, \psi_m)\) be Monge potentials associated with Monge–Kantorovich problem \( \Gamma(\rho_m, \nu_m)\), where
\[
d\rho_m = e^{-g_m} d\beta \quad \text{and} \quad d\nu_m = e^{-g_m} d\beta.
\]
Then \((\varphi_m)\) converges to \(\varphi\) in \(D_{2,1}(\rho)\), \((Q_{1/m} \psi_m)\) converges to \(\psi\) in \(L^1(\nu)\) and \((Q_{1/m} \nabla \psi_m)\) converges to \(\nabla \psi\) in \(L^2(\nu)\).

**Proof.** Let \(\gamma_m\) and \(\gamma\) be solutions of the Monge–Kantorovich problems for \((\rho_m, \nu_m)\) and \((\rho, \nu)\), respectively. Then
\[
\int_{\mathbb{R}^d} |\nabla \varphi_m|^2 e^{-f_m} d\beta = d_2(\rho_m, \nu_m)^2 \leq 4(H(\rho_m | \beta) + H(\nu_m | \beta))
\]
\[
\leq 4(H(\rho | \beta) + H(\nu | \beta))
\]
\[
= 4 \left( \int_{\mathbb{R}^n} -f e^{-f} d\beta + \int_{\mathbb{R}^n} -g e^{-g} d\beta \right)
\]
\[
\leq 2 \left( \int_{\mathbb{R}^n} |\nabla f|^2 e^{-f} d\beta + \int_{\mathbb{R}^n} |\nabla g|^2 e^{-g} d\beta \right) < \infty
\]
where we have used Talagrand’s inequality in the first line, Jensen’s inequality in the second and the following Gaussian Sobolev inequality [17, (1.2)]:
\[
\int_{\mathbb{R}^n} |\xi(x)|^2 \ln |\xi(x)| d\beta(x) \leq \int_{\mathbb{R}^n} |\nabla \xi(x)|^2 d\beta(x) + \|\xi\|_{L^2(\beta)}^2 \ln \|\xi\|_{L^2(\beta)}
\]
in the last line. Therefore, we have
\[
\sup_m \int_{\mathbb{R}^d} |\nabla \varphi_m|^2 e^{-f_m} d\beta < \infty.
\]
Using the Poincaré inequality, we get
\[
\int_{\mathbb{R}^d} |\varphi_m - E_{\rho_m}[\varphi_m]|^2 e^{-f_m} d\beta \leq \int_{\mathbb{R}^d} |\nabla \varphi_m|^2 e^{-f_m} d\beta.
\]
Then, we replace \(\varphi_m\) with \(\varphi_m - E_{\rho_m}[\varphi_m]\) and use the fact that \(\rho_m\) converges to \(\rho\) weakly to get
\[
(4.2) \quad \sup_m \|\varphi_m\|_{D_{2,1}(\rho)} < \infty,
\]
which implies \((\varphi_m)\) converges weakly in \(D_{2,1}(\rho)\). As in the proof of [14, Thm. 4.1], we have
\[
\varphi_m(x) + \psi_m(y) + \frac{1}{2} |x - y|^2 \geq 0.
\]
Applying the Ornstein–Uhlenbeck semigroup with respect to \(x\) and then \(y\),
Moreover, from a version of Young’s inequality given in [3, p. 15], we get
\[
\lim_{m \to \infty} (4.4)
\]
in view of the positivity improving property of the Ornstein–Uhlenbeck semi-group. Observe that
\[
\lim_{m}(4.3)
\]
We want to show that \( Q_{1/m}Q_{1/m}(|x - y|^2) \) converges to \(|x - y|^2\) in \( L^1(\gamma) \). Observe that
\[
Q_{1/m}Q_{1/m}(|x - y|^2)
\]
We integrate with respect to \( \gamma \), we get
\[
\int Q_{1/m}Q_{1/m}(|x - y|^2) \, d\gamma = \int \int ((e^{-1/m}x - e^{-1/m}y)^2(1 - e^{-2/m})t^2 + (1 - e^{-2/m})|z|^2) \, d\gamma \, \, d\gamma(x, y).
\]
It is easy to see that
\[
\lim_{m \to \infty} \int (1 - e^{-2/m})|z|^2 \, d\gamma = \lim_{m \to \infty} \int (1 - e^{-2/m})|t|^2 \, d\gamma = 0
\]
and
\[
|e^{-1/m}x + \sqrt{1 - e^{-2/m}t} - e^{-1/m}y - \sqrt{1 - e^{-2/m}z}|^2
\]
\[
\leq C(|x|^2 + |y|^2 + |z|^2 + |t|^2).
\]
Moreover, from a version of Young’s inequality given in [3, p. 15], we get
\[
\int |y|^2 \, d\gamma \leq \int e^{\alpha|y|^2} \, d\gamma + \frac{1}{\alpha} H(\nu|\beta),
\]
\[
\int |x|^2 \, d\gamma \leq \int e^{\alpha|x|^2} \, d\gamma + \frac{1}{\alpha} H(\rho|\beta).
\]
From the dominated convergence theorem, \( Q_{1/m}Q_{1/m}(|x - y|^2) \) converges to \(|x - y|^2\) in \( L^1(\gamma) \). Therefore, (4.3) reads
\[
(4.4)
\]
\[
\lim_{m}(4.4)
\]
\[
\lim_{m} -\frac{1}{2}d_2(\rho, \nu_m)^2 + \frac{1}{2}d_2(\rho, \nu)^2
\]
\[
\leq \lim_{m} (\frac{1}{2}d_2(\rho, \rho_m)^2 + \frac{1}{2}d_2(\nu_m, \nu)^2).
\]
Observe that
\[
\lim_m \int_{\mathbb{R}^d} |x|^2 \, d\rho_m = \lim_m \int_{\mathbb{R}^d} |x|^2 Q_{1/m} e^{-f} \, d\beta
\]
\[
= \lim_m \int_{\mathbb{R}^d} (Q_{1/m} \lambda |x|^2) \frac{e^{-f}}{\lambda} \, d\beta = \int |x|^2 \, d\rho,
\]
where we have used weak convergence of \(\rho_m\) to \(\rho\) in the first line and the dominated convergence theorem in the second. Indeed, Young’s inequality implies
\[
Q_{1/m} (\lambda |x|^2) \frac{e^{-g}}{\lambda} \leq \exp(Q_{1/m} \lambda |x|^2) + H(\rho | \beta) \quad \text{and for } \lambda < 1/2 \text{ Jensen’s inequality yields}
\]
\[
\int \exp(Q_{1/m} \lambda |x|^2) \, d\beta \leq \int Q_{1/m} \exp(\lambda |x|^2) \, d\beta = \int \exp(\lambda |x|^2) \, d\beta < \infty.
\]
Similarly, \(\lim \int_{\mathbb{R}^d} |y|^2 \, d\nu_m = \int_{\mathbb{R}^d} |y|^2 \, d\nu\). So \(d_2(\rho_m, \rho) \to 0\) and \(d_2(\nu_m, \nu) \to 0\) by using \([4, \text{Lem. 8.3}]\). Combining this result with (4.4), we find that
\[
(Q_{1/m} \varphi_m(x) + Q_{1/m} \psi_m(y) + \frac{1}{m} Q_{1/m} Q_{1/m}(|x - y|^2)) \text{ converges to } 0 \text{ in } L^1(\gamma)
\]
and it is uniformly integrable with respect to \(\gamma\). Moreover, \((Q_{1/m} \varphi_m)\) is uniformly integrable, so \((Q_{1/m} \psi_m)\) is uniformly integrable. Let \(a'\) and \(b'\) be weak limit points of \((Q_{1/m} \varphi_m)\) and \((Q_{1/m} \psi_m)\). The Cesàro means
\[
Q_{1/m} \varphi_m' = \frac{1}{n} \sum_{i=1}^n Q_{1/m} \varphi_m \quad \text{and} \quad Q_{1/m} \psi_m' = \frac{1}{n} \sum_{i=1}^n Q_{1/m} \psi_m
\]
converge, up to a subsequence, to \(a'\) and \(b'\) in \(L^1(\gamma)\), respectively. As a result, we have
\[
a' + b' + \frac{1}{2} |x - y|^2 = 0 \quad \gamma\text{-a.s.}
\]
Let \(a(x) = \limsup \varphi_m'(x)\) and \(b(y) = \limsup \psi_m'(y)\). Then \(a' = a\) and \(b' = b\) \(\gamma\text{-a.s.}\) and
\[
a(x) + b(y) + \frac{1}{2} |x - y|^2 \geq 0 \quad \text{for any } (x, y) \in \mathbb{R}^d \times \mathbb{R}^d,
\]
\[
a(x) + b(y) + \frac{1}{2} |x - y|^2 = 0 \quad \gamma\text{-a.s.}
\]
By uniqueness of solutions, we get \(a = \varphi\) and \(b = \psi\). Therefore, we deduce that \((Q_{1/m} \varphi_m)\) converges weakly to \(\varphi\) in \(L^1(\gamma)\).

On the other hand, from (4.2), there exists \(\varphi' \in L^2(\beta)\) such that \((\varphi_m)\) converges weakly to \(\varphi'\) in \(L^2(\rho)\) and in \(L^2(\gamma)\). For \(h \in L^2(\gamma)\), we have
\[
\int (\varphi_m - \varphi) h \, d\gamma = \int (\varphi_m - Q_{1/m} \varphi_m) h \, d\gamma + \int (Q_{1/m} \varphi_m - \varphi) h \, d\gamma.
\]
Since both integrals on the right hand side converge to zero as \(m \to \infty\), \((\varphi_m)\) converges weakly to \(\varphi\) in \(L^2(\gamma)\) and in \(L^2(\rho)\). Moreover, \(\lim \mathbb{E}_\rho [|
abla \varphi_m|^2] = \mathbb{E}_\rho [|
abla \varphi|^2]\), which implies \((\varphi_m)\) converges to \(\varphi\) in \(D_{2,1}(\rho)\). Similarly, \((Q_{1/m} \psi_m)\) converges to \(\psi\) in \(L^1(\nu)\), and since \(\nabla\) is closable in \(L^p(\nu)\) for \(p \geq 1\), \((\nabla Q_{1/m} \psi_m)\)
converges weakly to $\nabla \psi$ in $L^2(\nu)$. In addition, we have
\[
\lim_m \mathbb{E}_{\nu_m}[|\nabla \psi_m|^2] = \lim_m d_2(\rho_m, \nu_m)^2 = d_2(\rho, \nu)^2 = \mathbb{E}_\nu[|\nabla \psi|^2],
\]
which implies that $Q_{1/m} \nabla \psi_m$ converges to $\nabla \psi$ in $L^2(\nu)$. \qed

Now, let $\rho$ and $\nu$ be probability measures on $\mathbb{R}^n$ defined by
(4.5) \[ d\rho = F \, d\beta, \quad d\nu = G \, d\beta, \]
where $\beta$ is a Gaussian measure on $\mathbb{R}^n$ and $F, G \in L^1(\beta)$. Suppose that $\nu$ is also absolutely continuous with respect to $\rho$ with $d\nu/d\rho = L$. Define
\[ F_k = \frac{\theta_k F}{\mathbb{E}[\theta_k F]}, \quad G_k = \frac{\theta_k G}{\mathbb{E}[\theta_k G]}, \]
where $\theta_k \in C^\infty_c(\mathbb{R}^n)$ is a smooth function with compact support satisfying $\theta_k(x) = 0$ if $|x| \geq k$, $\theta_k(x) = 1$ if $|x| \leq k - 1$ with $0 \leq \theta_k \leq 1$ for each $k \in \mathbb{N}$, and $\sup_k |(\nabla \theta_k)^2/\theta_k| \leq 1$. Consider probability measures $\rho_k$ and $\nu_k$ given by
(4.6) \[ d\rho_k = F_k \, d\beta, \quad d\nu_k = G_k \, d\beta. \]
It is easy to see that $\nu_k$ is absolutely continuous with respect to $\rho_k$ with Radon–Nikodym derivative $d\nu_k/d\rho_k = L_k$, where
\[ L_k = \begin{cases} G_k/F_k & \text{on } \{\theta_k \neq 0\}, \\ 0 & \text{on } \{\theta_k = 0\}. \end{cases} \]
Observe that $L_k = b_k/a_k L$ on the set $\{\theta_k \neq 0\}$, where $a_k = 1/\mathbb{E}[\theta_k F]$ and $b_k = 1/\mathbb{E}[\theta_k G]$. Therefore, the relative entropy of $\nu_k$ with respect to $\rho_k$, $H(\nu_k|\rho_k)$, is finite if and only if $H(\nu|\rho)$ is finite. Note that $F_k$ and $G_k$ are bounded, being continuous functions with compact support. Using the following lemma, we will be able to approximate the smooth density functions of Lemma 4.2 by smooth and bounded sequences.

**Lemma 4.3.** Let $(\rho, \nu)$ and $(\rho_k, \nu_k)$ be probability measures on $\mathbb{R}^n$ defined as in (4.5) and (4.6), respectively, with $H(\nu|\rho) < \infty$. Let $(\varphi, \psi)$ and $(\varphi_k, \psi_k)$ be Monge potentials associated with the Monge–Kantorovich problems for $(\rho, \nu)$ and $(\rho_k, \nu_k)$, respectively, with quadratic cost. Then $(\varphi_k)$ converges to $\varphi$ in $\mathbb{D}_{2,1}(\rho)$, $(\theta_k \psi_k)$ converges to $\psi$ in $L^1(\nu)$ and $(\sqrt{\theta_k} \nabla \psi_k)$ converges to $\nabla \psi$ in $L^2(\nu)$.

**Proof.** Let $\gamma$ and $\gamma_k$ be the optimal transport plans for the Monge–Kantorovich problems for $(\rho, \nu)$ and $(\rho_k, \nu_k)$. From [13, Prop. 3.1], replacing $\varphi_k$ with $\varphi_k - \mathbb{E}[\varphi_k]$ and $\psi_k$ with $\psi_k + \mathbb{E}[\varphi_k]$, we have
(4.7) \[ F_k(x, y) = \varphi_n(x) + \psi_n(y) + \frac{1}{2} |x - y|^2 \geq 0 \]
for all \( x, y \in \mathbb{R}^n \). Since \( 0 \leq \theta_k \), we have

\[
0 \leq \int_{\mathbb{R}^n \times \mathbb{R}^n} \theta_k(y) F_k(x, y) \, d\gamma(x, y) = \int \theta_k(y) \varphi_k(x) \, d\gamma + \int \theta_k(y) \psi_k(y) \, d\gamma + \frac{1}{2} \int \theta_k(y) |x-y|^2 \, d\gamma. \tag{4.8}
\]

In the second integral above, we have

\[
\int \theta_k(y) \psi_k(y) \, d\gamma = b_k^{-1} \int \psi_k(y) \, d\nu_k(y) = b_k^{-1} \int \psi_k(y) \, d\gamma_k(y) = b_k^{-1} \left[ \left( \varphi_k(x) - \frac{1}{2} |x-y|^2 \right) \, d\gamma_k(x, y) \right] = b_k^{-1} \left[ \left( \varphi_k(x) \rho_k(x) - \frac{1}{2} |x-y|^2 \right) \, d\gamma_k(x, y) \right] = \frac{a_k}{b_k} \int -\theta_k(x) \varphi(x) \, d\rho - \frac{1}{2b_k} \int |x-y|^2 \, d\gamma_k(x, y). \tag{4.9}
\]

If we substitute (4.9) into (4.8), we get

\[
\int \theta_k(y) F_k(x, y) \, d\gamma = \int \left( \theta_k(y) - \frac{a_k}{b_k} \theta_k(x) \right) \varphi_k(x) \, d\gamma + \frac{1}{2} \int \theta_k(y) |x-y|^2 \, d\gamma(x, y) - \frac{1}{2b_k} \int |x-y|^2 \, d\gamma_k(x, y) =: I_k + II_k - III_k.
\]

The sequences \((a_k)\) and \((b_k)\) given by \(a_k = \mathbb{E}[\theta_k F]^{-1}\) and \(b_k = \mathbb{E}[\theta_k G]^{-1}\) converge to 1 decreasingly. Therefore, for given \( \epsilon > 0 \) there exists \( k_\epsilon > 0 \) such that \( 0 < \frac{b_k}{a_k} \leq \frac{1}{1-\epsilon} \) for each \( k \geq k_\epsilon \). Hence,

\[
\lim_k \int |\varphi_k|^2 \, d\gamma = \lim_k \int |\varphi_k|^2 \, d\beta \leq \lim_k \int |\nabla \varphi_k|^2 \, d\rho \leq \lim_k 2H(\nu_k|\rho_k) \leq 2 \left( \frac{b_k}{a_k} L \log L + \frac{b_k}{a_k} L \log \frac{b_k}{a_k} \right) \, d\rho \leq \frac{2}{1-\epsilon} H(\nu|\rho) + \frac{2}{1-\epsilon} < \infty \tag{4.10}
\]

for sufficiently small \( \epsilon \). We have used the Poincaré inequality in the first line and Talagrand’s inequality in the second. Therefore, from the Cauchy–Schwarz inequality and the dominated convergence theorem, we have \( I_k \to 0 \) as \( k \to \infty \). By monotone convergence, \( II_k \) converges to \( \frac{1}{2} \int |x-y|^2 \, d\gamma \). Moreover, \( III_k \) also converges to \( \frac{1}{2} \int |x-y|^2 \, d\gamma \). Indeed, Young’s inequality implies
that
\[ |x|^2 \theta_k(x) F \leq \frac{1}{1-\epsilon} \left( e^{|x|^2} + \frac{1}{\epsilon} F \log F \right), \]
\[ |y|^2 \theta_k(y) G \leq \frac{1}{1-\epsilon} \left( e^{|y|^2} + \frac{1}{\epsilon} G \log G \right). \]

In other words, the sequences \(|x|^2 \theta_k(x) F, k \geq k_\epsilon\) and \(|y|^2 \theta_k(y) G, k \geq k_\epsilon\) are bounded by a \(\beta\)-integrable function. Using the dominated convergence theorem, we get
\[ \lim_{k \to \infty} \int |x|^2 F_k \, d\beta = \lim_{k \to \infty} \int |x|^2 \, d\rho_k = \int |x|^2 \, d\rho, \]
\[ \lim_{k \to \infty} \int |y|^2 G_k \, d\beta = \lim_{k \to \infty} \int |y|^2 \, dv_k = \int |y|^2 \, dv. \]

Combining the above with \([4\text{ Lem. } 8.3]\) yields
\[ \lim_{k \to \infty} \frac{1}{2b_k} \int |x-y|^2 \, d\gamma_k(x, y) = \lim_{k \to \infty} \frac{1}{2b_k} d_2(\rho_k, \nu_k)^2 \]
\[ = \frac{1}{2} d_2(\rho, \nu)^2 = \frac{1}{2} \int |x-y|^2 \, d\gamma(x, y). \]

Therefore, we get

\[ (4.11) \quad \lim_{k \to \infty} \int \theta_k(y) F_k(x, y) \, d\gamma = 0. \]

The sequence \((\varphi_k)\) is bounded in \(\mathbb{D}_{2,1}(\rho)\). Therefore, it converges weakly to some \(\varphi' \in \mathbb{D}_{2,1}(\rho)\) up to a subsequence. Moreover, the sequence \((\theta_k(y) \varphi_k(x))\) is uniformly integrable and for any \(h \in L^\infty(\gamma)\),
\[ \int (\theta_k(y) \varphi_k(x) - \varphi'(x)) h \, d\gamma \]
\[ = \int (\theta_k(y) - 1) \varphi_k(x) h \, d\gamma + \int (\varphi_k(x) - \varphi'(y)) h \, d\gamma. \]

The second integral on the right hand side converges to zero since \((\varphi_k)\) converges to \(\varphi'\) weakly and the first integral converges to zero since \((\theta_k - 1)\) converges to zero in \(L^2(\gamma)\). Therefore, \((\theta_k \varphi_k)\) converges weakly to \(\varphi'\) in \(L^2(\gamma)\).

On the other hand, \((\theta_k F_k)\) converges to zero in \(L^1(\gamma)\) as given in \((4.11)\), so it is uniformly integrable. Since \((\theta_k \varphi_k)\) is uniformly integrable, \((\theta_k(y) \psi_k(x))\) is also uniformly integrable and converges weakly to some \(\psi'\) in \(L^1(\gamma)\) up to a subsequence. There exists a further subsequence such that the Cesàro means
\[ \varphi_k' = \frac{1}{k} \sum_{i=1}^{k} \theta_i(y) \varphi_i(x) \quad \text{and} \quad \psi_n' = \frac{1}{k} \sum_{i=1}^{k} \theta_i \psi_i \]
converge to \(\varphi'\) in \(L^2(\gamma)\) and \(\psi'\) in \(L^1(\gamma)\), respectively. Let \(a(x) = \limsup \varphi_k'(x)\) and \(b(y) = \limsup \psi_k'(y)\). Then \(\varphi' = a\) and \(\psi' = b\) \(\gamma\)-a.s. Therefore, \(\varphi' = a\)
and $\psi' = b \gamma$-a.s. and
\begin{align*}
a(x) + b(y) + \frac{1}{2}|x-y|^2 &\geq 0 \quad \text{for any } (x,y) \in \mathbb{R}^d \times \mathbb{R}^d, \\
a(x) + b(y) + \frac{1}{2}|x-y|^2 &= 0 \quad \gamma\text{-a.s.}
\end{align*}

By uniqueness of solutions, we have $\varphi' = \varphi$ and $\psi' = \psi$. Combining this result with $\lim \mathbb{E}_\rho[|\nabla \varphi_k|^2] = \mathbb{E}_\rho[|\nabla \varphi|^2]$, we see that $(\varphi_k)$ converges strongly to $\varphi$ in $\mathbb{D}_{2,1}(\rho)$. Finally, since $(\theta_k F_k)$ converges to zero and $(\frac{1}{2}\theta_k|x-y|^2)$ converges to $\frac{1}{2}|x-y|^2$ in $L^1(\gamma)$, $(\theta_k \psi_k)$ converges to $\psi$ in $L^1(\gamma)$. The only remaining part is to show the convergence of the sequence $(\sqrt{\theta_k} \nabla \psi_k)$. Observe that
\begin{align*}
\int \theta_k |\nabla \psi_k|^2 \, d\gamma &= \int \theta_k |\nabla \psi_k|^2 \, d\nu = \int b_k^{-1} |\nabla \psi_k|^2 \, d\nu_k(y) \\
&= \int b_k^{-1} |\nabla \varphi_k|^2 \, d\rho_k(x) = \int \frac{a_k}{b_k} \theta_k |\nabla \varphi_k|^2 \, d\rho(x).
\end{align*}

If we take the limit of both sides, we get
\begin{equation}
\tag{4.12}
\lim_{k \to \infty} \int \theta_k |\nabla \psi_k|^2 \, d\gamma = \int |\nabla \varphi|^2 \, d\rho(x) = \int |\nabla \psi|^2 \, d\nu(y).
\end{equation}

Next, we will show that the sequence $(\sqrt{\theta_k} \nabla \psi_k)$ converges weakly to $\nabla \psi$. Let $\xi$ be a bounded smooth vector field. Then
\begin{equation}
\tag{4.13}
\int \langle \sqrt{\theta_k} \nabla \psi_k, \xi \rangle \, d\gamma = \int_{\{\theta_k < c\}} \langle \sqrt{\theta_k} \nabla \psi_k, \xi \rangle \, d\gamma + \int_{\{\theta_k \geq c\}} \langle \sqrt{\theta_k} \nabla \psi_k, \xi \rangle \, d\gamma.
\end{equation}

Equation (4.12) implies that the sequence $(\sqrt{\theta_k} \nabla \psi_k)$ is uniformly integrable with respect to $\gamma$. Combining this with the boundedness of $\xi$ and the fact that $(\theta_k)$ converges to 1 imply that the first integral converges to 0 for $c < 1$. On the other hand, the second integral in (4.13) can be written as
\begin{align*}
\int_{\{\theta_k \geq c\}} \langle \sqrt{\theta_k} \nabla \psi_k, \xi \rangle \, d\nu &= b_k^{-1} \int_{\{\theta_k \geq c\}} \frac{\langle \nabla \psi_k, \xi \rangle}{\sqrt{\theta_k}} \, d\nu_k \\
&= b_k^{-1} \int_{\{\theta_k \circ T_k \geq c\}} -\frac{\langle \nabla \varphi_k, \xi \circ T_k \rangle}{\sqrt{\theta_k \circ T_k}} \, d\rho_k \\
&= \frac{a_k}{b_k} \int_{\{\theta_k \circ T_k \geq c\}} -\frac{\langle \nabla \varphi_k, \xi \circ T_k \rangle}{\sqrt{\theta_k \circ T_k}} \theta_k \, d\rho
\end{align*}

where $T_k = I_{\mathbb{R}^n} + \nabla \varphi_k$ is the optimal transport map of the Monge problem. Since the sequence $(L_k)$ is uniformly integrable, the sequence $(T_k)$ is equiconcentrated on a compact set. Hence, $\lim_{k \to \infty} \theta_k \circ T_k = 1$ and $\lim_{k \to \infty} \xi \circ T_k = \xi \circ T$ in $\rho$-probability. By using the dominated convergence theorem
and $L^2(\rho)$-boundedness of the sequence of $(\nabla \varphi_k)$, we get
\[
\lim_{k \to \infty} \frac{a_k}{b_k} \int_{\theta_k \circ T \geq c} -\frac{\langle \nabla \varphi_k, \xi \circ T \rangle}{\sqrt{\theta_k \circ T}} \theta_k \, d\rho = \int -\langle \nabla \varphi, \xi \circ T \rangle \, d\rho = \int (\nabla \psi_k \circ T, \xi \circ T) \, d\rho = \int (\nabla \psi, \xi) \, d\gamma.
\]
As a result, the limit of (4.13) is
\[
\lim_{k \to \infty} \int (\nabla \psi_k, \xi) \, d\gamma = \int (\nabla \psi_k, \xi) \, d\gamma,
\]
which implies that the sequence $(\sqrt{\theta_k \circ T} \nabla \psi_k)$ converges weakly to $\nabla \psi$ with respect to $\gamma$. In view of (4.12), $(\sqrt{\theta_k \circ T} \nabla \psi_k)$ converges to $\nabla \psi$ in $L^2(\nu).$ □

In the next lemma, we will use strictly positive sequences to approximate density functions such as those used in Lemma 4.3.

**Lemma 4.4.** Let $(\rho, \nu)$ be probability measures on $\mathbb{R}^n$ defined as in (4.5) with $H(\rho \mid \beta) < \infty$. Define, for each $\epsilon > 0$,
\[
d\rho_\epsilon = \frac{F + \epsilon}{1 + \epsilon} \, d\beta \quad \text{and} \quad d\nu_\epsilon = \frac{G + \epsilon}{1 + \epsilon} \, d\beta.
\]
Let $(\varphi, \psi)$ and $(\varphi_\epsilon, \psi_\epsilon)$ be Monge potentials associated with the Monge–Kantorovich problems $(\rho, \nu)$ and $(\rho_\epsilon, \nu_\epsilon)$ respectively with quadratic cost. Then, as $\epsilon$ goes to zero, $(\varphi_\epsilon, \epsilon > 0)$ converges to $\varphi$ in $\mathbb{D}_{2,1}(\rho)$, $(\psi_\epsilon, \epsilon > 0)$ converges to $\psi$ in $L^1(\nu)$ and $(\nabla \psi_\epsilon, \epsilon > 0)$ converges to $\nabla \psi$ in $L^2(\nu)$.

**Proof.** Let $\gamma$ and $\gamma_\epsilon$ be the optimal transport plans of the Monge–Kantorovich problems for $(\rho, \nu)$ and $(\rho_\epsilon, \nu_\epsilon)$, respectively. If we replace $\varphi_\epsilon$ with $\varphi_\epsilon - \mathbb{E}_{\rho_\epsilon}[\varphi_\epsilon]$ and $\psi_\epsilon$ with $\psi_\epsilon + \mathbb{E}_{\rho_\epsilon}[\varphi_\epsilon]$, we have
\[
F_\epsilon(x, y) = \varphi_\epsilon(x) + \psi_\epsilon(y) + \frac{1}{2} |x - y|^2 \geq 0 \quad \text{for any } x, y \in \mathbb{R}^n,
\]
\[
F_\epsilon(x, y) = \varphi_\epsilon(x) + \psi_\epsilon(y) + \frac{1}{2} |x - y|^2 = 0 \quad \gamma_\epsilon\text{-a.s.}
\]
for all $x, y \in \mathbb{R}^n$. Observe that
\[
0 \leq \int F_\epsilon(x, y) \, d\gamma + \epsilon \int \varphi_\epsilon(x) \, d\beta(x) + \epsilon \int \psi_\epsilon(y) \, d\beta(y)
\]
\[
= (1 + \epsilon) \int \varphi_\epsilon(x) \, d\rho_\epsilon(x) + (1 + \epsilon) \int \psi_\epsilon(y) \, d\nu_\epsilon(y) + \frac{1}{2} \int |x - y|^2 \, d\gamma
\]
\[
= (1 + \epsilon) \int \psi_\epsilon(y) \, d\nu_\epsilon(y) + \frac{1}{2} \int |x - y|^2 \, d\gamma.
\]
Note that the last inequality follows from the fact that $\mathbb{E}_{\rho_\epsilon}[\varphi_\epsilon] = 0$. Using Young’s inequality and [4, Lem. 8.3], we get
\[
\lim_{\epsilon \to 0} (1 + \epsilon) \int \psi_\epsilon \, d\nu_\epsilon = \lim_{\epsilon \to 0} \frac{1 + \epsilon}{2} \int |x - y|^2 \, d\gamma
\]
\[
= \lim_{\epsilon \to 0} -d_2(\rho_\epsilon, \nu_\epsilon)^2 = -d_2(\rho, \nu)^2 = -\frac{1}{2} \int |x - y|^2 \, d\gamma.
\]
Therefore, we have
\[
\lim_{\epsilon \to 0} \left( \int F_\epsilon(x, y) \, d\gamma + \epsilon \int \varphi_\epsilon(x) \, d\beta(x) + \epsilon \int \psi_\epsilon(y) \, d\beta(y) \right) = 0.
\]
On the other hand, \( F_\epsilon(x, x) = \varphi_\epsilon(x) + \psi_\epsilon(x) \geq 0 \). Hence, \( \epsilon \int \varphi_\epsilon \beta + \epsilon \int \psi_\epsilon \beta \geq 0 \), which implies
\[
\lim_{\epsilon \to 0} \int F_\epsilon(x, y) \, d\gamma = 0
\]
and we conclude that \((F_\epsilon, \epsilon > 0)\) is uniformly integrable with respect to \( \gamma \). Observe that
\[
\int |\nabla \varphi_\epsilon|^2 \, d\rho_\epsilon = d_2(\rho_\epsilon, \nu_\epsilon)^2 \leq 4H(\rho_\epsilon | \beta) + H(\nu_\epsilon | \beta)).
\]
Moreover, \( V_\epsilon = \frac{V}{1+\epsilon} + \frac{\epsilon}{1+\epsilon} \) and the convexity of the function \( x \mapsto x \log x \) implies
\[
H(\rho_\epsilon | \beta) = \int F_\epsilon \log F_\epsilon \, d\beta \leq \frac{1}{1+\epsilon} H(\rho | \beta).
\]
Similarly, \( H(\nu_\epsilon | \beta) \leq \frac{1}{1+\epsilon} H(\nu | \beta) \) and
\[
\sup_{\epsilon > 0} \int |\nabla \varphi_\epsilon|^2 \, d\rho_\epsilon < \infty.
\]
Combining the above result with \( \int |\nabla \varphi_\epsilon|^2 \, d\rho_\epsilon \geq \frac{1}{1+\epsilon} \int |\nabla \varphi_\epsilon|^2 \, d\rho \), we get
\[
\sup_{\epsilon > 0} \int |\nabla \varphi_\epsilon|^2 \, d\rho < \infty.
\]
Applying the Poincaré inequality yields \( \sup_{\epsilon > 0} \|\varphi_\epsilon\|_{\mathbb{D}_{2,1}(\rho)} < \infty \). Since the sequence \((\varphi_\epsilon, \epsilon > 0)\) is bounded in \( \mathbb{D}_{2,1}(\rho) \), it is also uniformly integrable with respect to \( \rho \) and \( \gamma \). We also know that the sequence \((F_\epsilon, \epsilon > 0)\) is uniformly integrable. Hence, \((\psi_\epsilon, \epsilon > 0)\) is uniformly integrable with respect to \( \gamma \). The rest of the proof goes along the proof of the previous lemmas. \( \blacksquare \)

4.2. Regularity of a backward Monge potential. Our approach for proving regularity of a backward Monge potential in a more general setting will be to approximate the functions \( f \) and \( g \) of (4.1) by appropriate sequences that will enable the use of Proposition 3.7. Explicitly, we first consider \( f_n \) on \( \mathbb{R}^n \) defined by \( e^{-f_n} = \mathbb{E}[e^{-f} | V_n] \), where \( V_n \) is generated by \( \{\delta e_1, \ldots, \delta e_n\} \) for an orthogonal basis \( \{e_i, i \geq 1\} \) of \( H \). In the second step, \( f_{nm} \) are chosen as smooth functions using the Ornstein–Uhlenbeck semiflow \((P_{1/m}) \) by \( e^{-f_{nm}} = P_{1/m}(e^{-f_n}) \), for each \( n \). Then \( F_{nmk} \) are continuous and compactly supported functions given by \( F_{nmk} = \frac{\theta_k e^{-f_{nmk}}}{\mathbb{E}[\theta_k e^{-f_{nmk}}]} \) for fixed \( m \) and \( k \). Finally, a strictly positive sequence given by \( e^{-f_{nmkl}} = \frac{F_{nmk+1/l}}{1+1/l} \) is formed to be compatible with the density \( e^{-f} > 0 \). Similarly, a sequence \((g_{nmkl})\) is also defined. As a result, the forward Monge potential \( \varphi_{nmkl} \) of the Monge–Kantorovich problem with respect to \((\rho_{nmkl}, \nu_{nmkl})\) is smooth and
Proposition 3.7 is applicable. We will work with a subsequence of \((\varphi_{nmkl})\), which has the form \((\varphi_{n,k_n,l_k,m_k})\) and can be extracted by applying the diagonal method three times. We will denote the corresponding sequences by \((\varphi_n), (f_n)\) and \((\rho_n)\). Similarly, \((\psi_n)\) will be formed by the diagonal method and its indices can be matched with those of \((\varphi_n)\). We will work with this subsequence by relabeling as \((\varphi_n, \psi_n)\). After this simplification, \((\varphi_n)\) converges to \(\varphi\) in \(\mathbb{D}_{2,1}(\rho)\), \((\theta_n P_1/\rho \psi_n)\) converges to \(\psi\) in \(L^1(\nu)\) and \((\sqrt{\theta_n} P_1/\nu \nabla \psi_n)\) converges to \(\nabla \psi\) in \(L^2(\nu)\).

**Lemma 4.5.** Let \((\varphi, \psi)\) be Monge potentials associated to the Monge–Kantorovich problem \((\rho, \nu)\), where \(d\rho = e^{-f} d\mu\) and \(d\nu = e^{-g} d\mu\), let \(f, g \in \mathbb{D}_{2,1}\) satisfy (4.1) and let the measure \(\rho\) satisfy the Poincaré inequality (2.1). Then

\[
\lim_n \nabla f_n = \nabla f, \quad \lim_n \nabla g_n \circ T_n = \nabla g \circ T
\]

in \(L^2(\rho, H)\).

**Proof.** It will be convenient to use the multi-index sequence defined above as

\[
e^{-f_{nmkl}} = \frac{\theta_k e^{-f_{nm}} + 1/l}{1 + 1/l}.
\]

We have

\[
\mathbb{E}[|\nabla f_{nmkl}|^2 e^{-f_{nmkl}}] = 4 \mathbb{E}[|\nabla e^{-f_{nmkl}/2}|^2] = \mathbb{E}\left[\frac{\nabla e^{-f_{nmkl}}}{e^{-f_{nmkl}/2}}\right]^2
\]

\[
= \mathbb{E}\left[1\{\theta_k=0\} \left|\frac{\nabla e^{-f_{nmkl}}}{e^{-f_{nmkl}/2}}\right|^2\right] + \mathbb{E}\left[1\{0<\theta_k<1\} \left|\frac{\nabla e^{-f_{nmkl}}}{e^{-f_{nmkl}/2}}\right|^2\right]
\]

\[
+ \mathbb{E}\left[1\{\theta_k=1\} \left|\frac{\nabla e^{-f_{nmkl}}}{e^{-f_{nmkl}/2}}\right|^2\right]
\]

\[
= II_{nmkl} + III_{nmkl} + II_{nmkl}.
\]

On the set \(\{\theta_k = 1\}\), we have

\[
III_{nmkl} = \frac{l}{l+1} \mathbb{E}\left[1\{\theta_k=1\} \left|\frac{\nabla e^{-f_{nm}}}{(e^{-f_{nm}} + 1/l)^{1/2}}\right|^2\right]
\]

\[
\leq \mathbb{E}[1\{\theta_k=1\}|\nabla P_{1/m}(e^{-f_n})|^2(P_{1/m}(e^{-f_n})^{-1})
\]

\[
\leq e^{-2/m} \mathbb{E}[1\{\theta_k=1\}P_{1/m}(|\nabla e^{-f_n}|^2)(P_{1/m}(e^{-f_n})^{-1})
\]

\[
= \mathbb{E}[1\{\theta_k=1\}P_{1/m}(|\nabla f_n e^{-f_n}|^2(P_{1/m}(e^{-f_n})^{-1})
\]

\[
= \mathbb{E}\left[1\{\theta_k=1\}P_{1/m}(|\nabla f_n|^2 e^{-f_n}) P_{1/m}(e^{-f_n})\right]
\]

\[
\leq \mathbb{E}\left[1\{\theta_k=1\} \left|\frac{\nabla e^{-f_n}}{e^{-f_n}}\right|^2\right]
\]
\[
\begin{align*}
\leq & \mathbb{E}\left[1_{\theta_k=1} \frac{\nabla \mathbb{E}[e^{-f}|V_n]]^2}{\mathbb{E}[e^{-f}|V_n]}\right] \leq \mathbb{E}\left[1_{\theta_k=1} \frac{\mathbb{E}[\nabla e^{-f}|V_n]]^2}{\mathbb{E}[e^{-f}|V_n]}\right] \\
\leq & \mathbb{E}\left[1_{\theta_k=1} \mathbb{E}[|\nabla f|^2 e^{-f}|V_n]]\mathbb{E}[e^{-f}|V_n]\right] \leq \mathbb{E}[|\nabla f|^2 e^{-f}].
\end{align*}
\]

On the set \(\{0 < \theta_k < 1\}\), we get
\[
II_{nmkl} = \frac{l}{1 + l} \mathbb{E}\left[1_{0 < \theta_k < 1} \frac{(\nabla \theta_k e^{-f_{nm}} + \theta_k \nabla e^{-f_{nm}})^2}{\theta_k e^{-f_{nm}} + 1/m}\right]
\leq 2 \mathbb{E}\left[1_{0 < \theta_k < 1} \frac{|\nabla \theta_k|^2}{\theta_k} e^{-f_{nm}}\right] + 2 \mathbb{E}\left[|\nabla e^{-f_{nm}}|^2\right]
\leq 2 \mathbb{E}\left[1_{0 < \theta_k < 1} \frac{|\nabla \theta_k|^2}{\theta_k} e^{-f}\right] + 2 \mathbb{E}\left[|\nabla e^{-f}|^2\right].
\]
Therefore, \(II_{nmkl}\) is bounded uniformly. Moreover, \(\lim_{\theta_k \to 0} \mu(\{0 < \theta_k < 1\}) = 0\), which implies \(\lim II_{nmkl} = 0\). When \(\theta_k = 0\), as the function \(e^{-f_{nmkl}}\) is constant, its derivative is zero and \(\lim I_{nmkl} = 0\) as well. As a result, \((\nabla e^{-f_{nmkl}/2})\) is uniformly bounded in \(L^2(\mu, H)\) and
\[
\limsup \mathbb{E}[|\nabla e^{-f_{nmkl}/2}|^2] \leq \mathbb{E}[|\nabla e^{-f/2}|^2].
\]

On the other hand, \((f_{nmkl})\) converges to \(f\) in \(L^0(\mu)\), that is, in probability, \(\mathbb{E}[|e^{-f_{nmkl}/2}|^2]\) converges to \(\mathbb{E}[|e^{-f/2}|^2]\), which implies that \((e^{-f_{n/2}})\) converges to \(e^{-f/2}\) in \(L^2(\mu)\) and \((\nabla e^{-f_{nmkl}/2})\) converges weakly to \(\nabla e^{-f/2}\) in \(L^2(\mu, H)\). By the weak lower semicontinuity of the norm, we get
\[
\mathbb{E}[|\nabla e^{-f/2}|^2] \leq \liminf \mathbb{E}[|\nabla e^{-f_{nmkl}/2}|^2]
\]
which implies that
\[
\lim_{n,l,k,m} \mathbb{E}[|\nabla e^{-f_{nmkl}/2}|^2] = \mathbb{E}[|\nabla e^{-f/2}|^2],
\]
and also \((\nabla e^{-f_{nmkl}/2})\) converges to \(\nabla e^{-f/2}\) in \(L^2(\mu, H)\) since it converges weakly. Hence,
\[
\lim_{n,l,k,m} \mathbb{E}[|\nabla f_{nmkl}|^2 e^{-f_{nmkl}}] = \lim_{n,l,k,m} 4 \mathbb{E}[|\nabla e^{-f_{nmkl}/2}|^2]
= 4 \mathbb{E}[|\nabla e^{-f/2}|^2] = \mathbb{E}[|\nabla f|^2 e^{-f}]
\]
and we get \(\lim_{n,l,k,m} \mathbb{E}[|\nabla f_{nmkl}|^2 e^{-f}] = \mathbb{E}[|\nabla f|^2 e^{-f}]\). Moreover, \((\nabla f_{nmkl})\) converges to \(\nabla f\) in \(L^0(\rho)\). Hence, \((\nabla f_{nmkl})\) converges to \(\nabla f\) in \(L^2(\rho)\). Similar calculations show that \((\nabla f_{nmkl} \circ T_{nmkl})\) converges to \(\nabla g \circ T\) in \(L^2(\rho)\).

We are ready to prove the regularity of a backward Monge potential. In the next theorem, note that we do not assume the Poincaré inequality since it is implied by \((1 - c)\)-convexity \([13\text{ Thm. 6.2}]\).
Theorem 4.6. Let \((W, H, \mu)\) be an abstract Wiener space, \(g \in D_{2,1}\) and \(f \in D_{2,1}\) such that
\begin{equation}
\int_{W} |\nabla f|^2 e^{-f} \, d\mu < \infty \quad \text{and} \quad \int_{W} |\nabla g|^2 e^{-g} \, d\mu < \infty
\end{equation}
and the function \(f\) is \((1 - c)\)-convex for some \(c \in [0, 1)\). Let \((\varphi, \psi)\) be forward and backward potentials to the Monge–Kantorovich problem with initial measure and target measure
\[ d\rho = e^{-f} d\mu, \quad d\nu = e^{-g} d\mu \]
and quadratic cost given by
\[ C(x, y) = \begin{cases} |x - y|_H^2 & \text{if } x - y \in H, \\ \infty & \text{if } x - y \notin H. \end{cases} \]
Then \(\nabla^2 \psi \in L^2(\nu, H \otimes H)\) and can be estimated by
\[ \mathbb{E}_\nu[|\nabla^2 \psi|^2] \leq \frac{3}{c} \left( \mathbb{E}_\rho[|\nabla \varphi|^2_H] + \mathbb{E}_\nu[|\nabla g|^2_H] + \mathbb{E}_\rho[|\nabla f|^2_H] \right), \]
where \(H \otimes H\) denotes the space of Hilbert–Schmidt operators on \(H\).

Proof. Thanks to condition (4.16), the Sobolev derivative is closable in \(L^2(\rho)\) and \(L^2(\nu)\). Define
\[ d\rho_n = e^{-f_n} \, d\mu, \quad d\nu_n = e^{-g_n} \, d\mu \]
with \(f_n\) and \(g_n\) as described before Lemma 4.5. The new probability measures \(\rho_n\) and \(\nu_n\) satisfy the sufficient conditions for the forward Monge potential to be smooth. Hence, if we apply Proposition 3.7, we have
\[ c\mathbb{E}_{\nu_n}[|\nabla^2 \psi_n|^2] \leq 3 \left( \mathbb{E}_{\rho_n}[|\nabla \varphi|^2_H] + \mathbb{E}_{\nu_n}[|\nabla g_n|^2_H] + \mathbb{E}_{\rho_n}[|\nabla f_n|^2_H] \right). \]

Lemmas 4.1–4.5 imply that the limits of the quantities on the right hand side exist and
\[ \mathbb{E}_{\nu_n}[|\nabla^2 \psi_n|^2] \leq \frac{3}{c} \left( \mathbb{E}_\rho[|\nabla \varphi|^2_H] + \mathbb{E}_\nu[|\nabla g|^2_H] + \mathbb{E}_\rho[|\nabla f|^2_H] \right). \]
Those lemmas also imply that \((\sqrt{\theta_n} P_{1/n} \nabla \psi_n)\) converges weakly to \(\nabla^2 \psi\). In view of the weak lower semicontinuity of the norms, if we take a weak limit with respect to \(n\), we obtain
\[ \mathbb{E}_\nu[|\nabla^2 \psi|^2] \leq \liminf_n \mathbb{E}[|\sqrt{\theta_n} \nabla^2 P_{1/n} \psi_n|^2 e^{-g}] \]
\[ \leq \sup_n e^{-2/n} \mathbb{E}[\theta_n P_{1/n} |\nabla^2 \psi_n|^2 e^{-g_n}] \]
\[ \leq \sup_n \mathbb{E}[|\nabla^2 \psi_n|^2 e^{-g_n}] \]
\[ \leq \frac{3}{c} \left( \mathbb{E}_\rho[|\nabla \varphi|^2_H] + \mathbb{E}_\nu[|\nabla g|^2_H] + \mathbb{E}_\rho[|\nabla f|^2_H] \right) \]
and the result follows. □
Remark 4.7. The result of Lemma 4.5 is more than what is needed in the proof of Theorem 4.6. Indeed, it can be shown that
\[ E_{\nu_n} [ |\nabla g_n|^2_H ] \leq 8 E_{\nu} [ |\nabla g|^2_H ] \quad \text{and} \quad E_{\rho_n} [ |\nabla f_n|^2_H ] \leq 8 E_{\rho} [ |\nabla f|^2_H ]. \]

4.3. Monge–Ampère equation. We will prove that a backward potential \( \psi \) solves the Monge–Ampère equation. The idea of the proof is to start with finite dimensions and take limits, as accomplished through the following lemmas.

Lemma 4.8. Under the assumptions of Theorem 4.6, \( (L\psi_n) \) converges to \( L\psi \) in the sense of distributions and \( L\psi \in L^1(\nu) \).

Proof. The convergence in the sense of distributions follows from duality. Once we show that \( (L\psi_n) \) is uniformly integrable, we are done. We will use an idea from [6]. First we will show
\[
\sup_n \int \frac{(L\psi_n)^2}{1 + |\nabla \psi_n|^2} \, d\rho < M < \infty;
\]
then we will show the uniform integrability of \( (L\psi_n) \) by using the convergence of \( (\nabla \psi_n) \) in \( L^2(\nu) \).

Let \( u \) be a decreasing function on \([0, +\infty]\). Observe that
\[
\int (L\psi_n)^2 u(|\nabla \psi_n|^2)e^{-g} \, d\mu = \int \langle \nabla \psi_n, \nabla (L\psi_n u(|\nabla \psi_n|^2)e^{-g}) \rangle_H \, d\mu \\
= \int \langle \nabla \psi_n, \nabla (L\psi_n u(|\nabla \psi_n|^2)) \rangle_H e^{-g} \, d\mu \\
- \int \langle \nabla \psi_n, \nabla g \rangle_H L\psi_n u(|\nabla \psi_n|^2)e^{-g} \, d\mu \\
= \int \langle \nabla \psi_n, \nabla (L\psi_n) \rangle_H u(|\nabla \psi_n|^2)e^{-g} \, d\mu \\
+ 2 \int \langle \nabla \psi_n, \nabla^2 \psi (\nabla \psi) \rangle_H u'(|\nabla \psi_n|^2) L\psi_n e^{-g} \, d\mu \\
- \int \langle \nabla \psi_n, \nabla g \rangle_H L\psi_n u(|\nabla \psi_n|^2)e^{-g} \, d\mu \\
= \int \langle \nabla \psi_n, L\nabla \psi_n \rangle_H u(|\nabla \psi_n|^2)e^{-g} \, d\mu \\
+ \int \langle \nabla \psi_n, \nabla \psi_n \rangle_H u(|\nabla \psi_n|^2)e^{-g} \, d\mu \\
+ 2 \int \langle \nabla \psi_n, \nabla^2 \psi (\nabla \psi) \rangle_H u'(|\nabla \psi_n|^2) L\psi_n e^{-g} \, d\mu \\
- \int \langle \nabla \psi_n, \nabla g \rangle_H L\psi_n u(|\nabla \psi_n|^2)e^{-g} \, d\mu.
\]

Finally, if we write the first line of the last term, we see that
\[
\left(\mathcal{L}\psi_n\right)^2 u(|\nabla\psi_n|^2) e^{-g} \, d\mu =
\]

(I) \[\int |\nabla^2 \psi_n|^2 u(|\nabla\psi_n|^2) e^{-g} \, d\mu\]

(II) \[-\int \langle \nabla^2 \psi_n (\nabla \psi_n), \nabla g \rangle_H u(|\nabla\psi_n|^2) e^{-g} \, d\mu\]

(III) \[+ 2\int |\nabla^2 \psi_n (\nabla \psi_n)|^2_H u'(|\nabla\psi_n|^2) e^{-g} \, d\mu\]

(IV) \[+ \int \langle \nabla \psi_n, \nabla \psi_n \rangle_H u(|\nabla\psi_n|^2) e^{-g} \, d\mu\]

(V) \[+ 2\int \langle \nabla \psi_n, \nabla^2 \psi (\nabla \psi) \rangle_H u'(|\nabla\psi_n|^2) \mathcal{L}\psi_n e^{-g} \, d\mu\]

(VI) \[-\int \langle \nabla \psi_n, \nabla g \rangle_H \mathcal{L}\psi_n u(|\nabla\psi_n|^2) e^{-g} \, d\mu.\]

We know that (I), (IV) are bounded and (II) is negative since \(u\) is decreasing. We will show that the other integrals are also bounded when we choose \(u\) properly. Observe that for all \(\epsilon > 0\) the Cauchy–Schwarz inequality implies

\[
|II| \leq \sqrt{\frac{1}{4\epsilon}} \left[ \int |\nabla g|^2 e^{-g} \, d\mu \right] \sqrt{4\epsilon \int |\nabla^2 \psi (\nabla \psi)|^2 u^2(|\nabla\psi_n|^2) e^{-g} \, d\mu}
\]

\[
\leq \frac{1}{4\epsilon} \int |\nabla g|^2 e^{-g} \, d\mu + \epsilon \int |\nabla^2 \psi (\nabla \psi)|^2 u^2(|\nabla\psi_n|^2) e^{-g} \, d\mu,
\]

and

\[
|VI| \leq \sqrt{\frac{1}{4\epsilon}} \left[ \int |\nabla g|^2 e^{-g} \, d\mu \right] \sqrt{4\epsilon \int |\nabla\psi_n|^2 \mathcal{L}\psi_n)^2 u^2(|\nabla\psi_n|^2) e^{-g} \, d\mu}
\]

\[
\leq \frac{1}{4\epsilon} \int |\nabla g|^2 e^{-g} \, d\mu + \epsilon \int |\nabla\psi_n|^2 \mathcal{L}\psi_n)^2 u^2(|\nabla\psi_n|^2) e^{-g} \, d\mu
\]

and

\[
|V| \leq \sqrt{\frac{1}{\epsilon}} \int \left(\mathcal{L}\psi_n\right)^2 u(|\nabla\psi_n|^2) e^{-g} \, d\mu
\]

\[
\times \sqrt{4\epsilon \int \left(\frac{u'}{u}\right)^2(|\nabla\psi_n|^2)|\nabla^2 \psi (\nabla \psi)|^2 |\nabla \psi|^2 e^{-g} \, d\mu}
\]

\[
\leq \epsilon \int \left(\mathcal{L}\psi_n\right)^2 u(|\nabla\psi_n|^2) e^{-g} \, d\mu
\]

\[
+ \frac{1}{\epsilon} \int \left(\frac{u'}{u}\right)^2(|\nabla\psi_n|^2)|\nabla^2 \psi (\nabla \psi)|^2 |\nabla \psi|^2 e^{-g} \, d\mu.
\]

If we take \(u(t) = \frac{1}{1+t}\) and \(\epsilon = \frac{1}{4}\), we get

\[
\left(\frac{u'}{u}\right)^2(|\nabla\psi_n|^2)|\nabla\psi_n|^4 \leq 1,
\]

\[
|\nabla\psi_n|^2 u(|\nabla\psi_n|^2) \leq 1
\]

and

\[
\epsilon u^2(|\nabla\psi_n|^2) + 2u'(|\nabla\psi_n|^2) \leq 0.
\]
Hence, we have
\[
\sup_n \left\{ \frac{(\mathcal{L}\psi_n)^2}{1 + |\nabla\psi_n|^2} \right\} d\rho < M < \infty
\]
where
\[
M = \int |\nabla g|^2 e^{-g} \, d\mu + 2 \sup_n \int |\nabla\psi_n|^2 e^{-g} \, d\mu + 10 \sup_n \int |\nabla^2\psi_n|^2 e^{-g} \, d\mu < \infty.
\]
Note that \((\nabla\psi_n)\) is uniformly integrable with respect to \(\nu\). So for every \(\epsilon > 0\) there exists \(\delta > 0\) such that \(\rho(E) < \delta\) implies \(\mathbb{E}_\nu[|\nabla\psi|^2 1_E] < \frac{\epsilon^2}{1 + 4M}\). Define
\[
E_1 = \left\{ |\mathcal{L}\psi_n| \leq \frac{\epsilon}{2M} \frac{|\mathcal{L}\psi_n|^2}{1 + |\nabla\psi_n|^2} \right\},
\]
\[
E_2 = \left\{ |\mathcal{L}\psi_n| > \frac{\epsilon}{2M} \frac{|\mathcal{L}\psi_n|^2}{1 + |\nabla\psi_n|^2} \right\}.
\]
We have \(\mathbb{E}_\nu[|\mathcal{L}\psi_n| 1_E] = \mathbb{E}_\nu[|\mathcal{L}\psi_n| 1_{E_1}] + \mathbb{E}_\nu[|\mathcal{L}\psi_n| 1_{E_2}] < \epsilon/2 + \epsilon/2\). Hence \((\mathcal{L}\psi_n)\) is uniformly integrable with respect to \(\nu\). Therefore, \((\mathcal{L}\psi_n)\) converges weakly to \(\mathcal{L}\psi\) in \(L^1(\nu)\).

**Lemma 4.9.** Under the assumptions of Theorem 4.6, \((f_n \circ S_n)\) converges to \(f \circ S\) in \(L^1(\nu)\), where \(S = T^{-1} = I_W + \nabla\psi\).

**Proof.** We have
\[
\int |f_n \circ S_n - f \circ S| \, d\nu \leq \int |f_n \circ S_n - f \circ S_n|_H \, d\nu + \int |f \circ S_n - f \circ S|_H \, d\nu.
\]
By Fatou’s lemma
\[
\lim \int |f_n \circ S_n - f \circ S_n| \, d\nu \leq \lim \int |f_n - f| e^{-f_n} \, d\mu = 0.
\]
Let \(\hat{f} \in C_b(W)\) be such that \(|f - \hat{f}|_{L^1(\rho)} < \epsilon\). Then, again using Fatou’s lemma,
\[
\lim \int |f \circ S_n - f \circ S| \, d\nu \leq \lim \int |f \circ S_n - \hat{f} \circ S_n| e^{-g_n} \, d\mu + \lim \int |\hat{f} \circ S_n - \hat{f} \circ S| e^{-g} \, d\mu
\]
\[
+ \int |\hat{f} \circ S - f \circ S| e^{-g} \, d\mu
\]
\[
\leq \lim \int |f - \hat{f}| e^{-f_n} \, d\mu + \lim \int |\hat{f} \circ S_n - \hat{f} \circ S| e^{-g} \, d\mu + \int |\hat{f} - f| e^{-f} \, d\mu
\]
\[
\leq 2\epsilon.
\]
Therefore, \(\lim \int |f_n \circ S_n - f \circ S| \, d\rho = 0\).

**Lemma 4.10.** Under the assumptions of Theorem 4.6,
\[
\lim_n \mathbb{E}_\nu[|\nabla^2\psi - \nabla^2\psi_n|^2] = 0
\]
where \(|\cdot|_2\) denotes the Hilbert–Schmidt operator norm.
Proof. Let $m, n$ be integers such that $m > n$ and $\pi_m^n$ be the orthogonal projection from $H_m$ onto $H_n$. Then [12, Thm. 2.5] implies

$$|\nabla^2 (\psi_n \circ \pi_m^n) - \nabla^2 \psi_m|_{L^1(\rho_m, H_m \otimes H_m)}^2 \leq C_1 \int_{H_m} (f_m - f_n \circ \pi_m^n) \, d\rho_m$$

$$+ \frac{C_2}{\epsilon} \int_{H_m} |\nabla g_m - \nabla (g_n \circ \pi_m^n)|^2 \, d\nu_m$$

for some constants $C_1, C_2 > 0$. If we take the limit with respect to $m$ first, and then with respect to $n$, we get the result.

**Theorem 4.11.** Let $(W, H, \mu)$ be an abstract Wiener space, $f \in D_{2, 1}$ and $g \in D_{2, 1}$ such that

$$\int |\nabla f|^2 e^{-f} \, d\mu < \infty \quad \text{and} \quad \int |\nabla g|^2 e^{-g} \, d\mu < \infty$$

and the function $f$ is $(1 - c)$-convex for some $c \in [0, 1)$. Let $\psi$ be a backward potential of the Monge–Kantorovich problem with initial and target measures

$$d\rho = e^{-f} \, d\mu, \quad d\nu = e^{-g} \, d\mu$$

and quadratic cost. Then the backward potential $\psi$ solves the Monge–Ampère equation

$$e^{-g} = e^{-f \circ S} \det (I_H + \nabla^2 \psi) \exp \left[-L \psi - \frac{1}{2}|\nabla \psi|_H^2\right]$$

$\nu$-a.s., where $S = T^{-1} = I_W + \nabla \psi$.

Proof. By the finite-dimensional result in [25], we have

$$e^{-g_n} = e^{-f_n \circ S_n} \det (I_H + \nabla^2 \psi_n) \exp \left[-L \psi_n - \frac{1}{2}|\nabla \psi_n|_H^2\right].$$

We already know that $e^{-g_n} \to e^{-g}$. From the approximation lemmas in Subsection 4.1 as $n \to \infty$, $\sqrt{\theta_n} P_{1/n} \nabla \psi_n \to \nabla \psi$ $\nu$-a.s. Due to the continuity of $P_t$ in $t$ [21, Prop. 2.1] and since $\theta_n \to 1$, it follows that $\nabla \psi_n \to \nabla \psi$ $\nu$-a.s. By Lemmas 4.9 and 4.10, we also know that $e^{-f_n \circ S_n} \to e^{-f \circ S}$, and $\nabla^2 \psi_n \to \nabla^2 \psi$ and the sequence $(-L \psi_n)$ has a limit, say $A$, $\nu$-a.s. Therefore, we have

$$e^{-g} = e^{-f \circ S} \det (I_H + \nabla^2 \psi) \exp \left[A - \frac{1}{2}|\nabla \psi|_H^2\right].$$

Once we show that $-L \psi = A$ $\nu$-a.s., we are done. Indeed, since $(-L \psi_n)$ is uniformly integrable, for every cylindrical function $\xi$, we have

$$\int A \xi e^{-g} \, d\mu = \lim_n \int -L \psi_n \xi e^{-g} \, d\mu = \lim_n \int -\langle \nabla \psi_n, \nabla \xi - \xi \nabla g \rangle_H \psi_n e^{-g} \, d\mu$$

$$= \lim_n \int -\langle \nabla \psi, \nabla \xi - \xi \nabla g \rangle_H \psi_n e^{-g} \, d\mu = \int -L \psi \xi e^{-g} \, d\mu$$

so $-L \psi = A$ $\nu$-a.s. ☐
Acknowledgements. The authors are grateful to A. Süleyman Üstünel for his helpful guidance and instructive comments on the manuscript. They would also like to thank the anonymous reviewer for several valuable comments including the simplification of the proof of Proposition 3.2.

This work is supported by TUBITAK Project 118F403.

References


Backward potential and the Monge–Ampère equation


Mine Çağlar, İhsan Demirel
Department of Mathematics
Koç University
Istanbul, Turkey
E-mail: mcaglar@ku.edu.tr
idemirel17@ku.edu.tr