CONVEXITY INEQUALITIES AND OPTIMAL TRANSPORT OF INFINITE DIMENSIONAL MEASURES

ALEXANDER V. KOLESNIKOV

ABSTRACT. We generalize Talagrand's inequality for transportation cost of Gaussian measures and give applications of our result. In particular, we establish an estimate that relates two different transporation mappings. In the finite-dimensional case we obtain a new log-Sobolev type inequality. In the infinite-dimensional case we consider transformations of measures absolutely continuous with respect to a given Gaussian measure into this Gaussian measure.

1. Introduction

According to a well-known result of Talagrand [30] the square of the quadratic transportation cost $W_2(\gamma, \mu)$ between the standard Gaussian measure γ on $\mathbb{R}^d$ and a probability measure $\mu = f \cdot \gamma$ can be estimated by the double entropy $\int_{\mathbb{R}^d} f \log f \, d\gamma$, i.e.,

$$W_2^2(\gamma, \mu) \le 2 \int_{\mathbb{R}^d} f \log f \, d\gamma. \tag{1.1}$$

The transportation cost $W_L(\nu_1, \nu_2)$ between two probability measures ν_1, ν_2 on a measurable space X can be defined as the minimum of the following functional (the Kantorovich functional):

$$\nu \to \Bigl[\int_{X \times X} L(x_1, x_2)_X \, d\nu(x_1, x_2)\Bigr]^{\frac{1}{2}}, \quad \nu \in \mathcal{P}(\nu_1, \nu_2), \tag{1.2}$$

where $\mathcal{P}(\nu_1, \nu_2)$ is the set of all probability measures on $X \times X$ with the marginals ν_1 and ν_2 and L is a positive function called "the cost function". The problem of minimizing (1.2) is called the mass transportation problem. This formulation was given by L. Kantorovich in [23]. If X is a Hilbert space and L is given by the square of the Hilbert norm $L(x_1, x_2) = (x_1 - x_2)_X^2$, the minimum of

$$\nu \to \Bigl[\int_{X \times X} (x_1 - x_2)_X^2 \, d\nu(x_1, x_2)\Bigr]^{\frac{1}{2}}, \quad \nu \in \mathcal{P}(\nu_1, \nu_2), \tag{1.3}$$

is called the quadratic transportation cost (the Kantorovich-Rubinstein or, sometimes, the Wasserstein distance). A detailed review of results on the mass transportation problem can be found in [29].

In many partial cases there exists a mapping $T : X \to X$ called an optimal transport or an optimal transfer plan such that $\nu_2 = \nu_1 \circ T^{-1}$ and

$$W_2^2(\nu_1, \nu_2) = \int_X (x - T(x))^2 \, d\nu_1.$$

In this case the measure $\nu_1 \circ (x, T(x))^{-1}$ on $X \times X$ minimizes the Kantorovich functional. The finite dimensional theory of optimal transport was studied by many authors. The first results on existence of optimal mappings were obtained in the classical work

Supported in part by RFBR Grant 01–01–00858, DFG Grant 436 RUS 113/343/0(R), RFBR MAS Grant 03–01–06270.

of A.D. Alexandroff [1]. The existence, uniqueness and regularity of finite-dimensional optimal transports were established under broad conditions by Brenier [11]. McCann [27] developed an elementary approach based on convex analysis (see also Gangbo [21]). In particular, it was shown in [27] that given two probability densities ϱ_1, ϱ_2, there exists an optimal transport T which pushes forward $\varrho_1\,dx$ to $\varrho_2\,dx$. Moreover, this mapping is $\varrho_1\,dx$-unique and has the form $T = \nabla\Psi$, where Ψ is a convex function that solves the following non-linear PDE (the so-called Monge–Ampère equation):

$$\varrho_2(\nabla\Psi)\det D^2\Psi = \varrho_1. \tag{1.4}$$

Recent investigations revealed many interesting links between the mass transportation problem and some well-known geometric and analytic inequalities, e.g., the Brunn-Minkowski, logarithmic Sobolev, Poincaré and concentration inequalities (see [26], [33]). Some interesting properties of the optimal transport for Gaussian measures have been established in [14]. In [14] and [22], applications of the optimal transport method to the well-known Gaussian correlation conjecture and FKG inequalities were given.

Recall that a probability measure μ on $\mathbb{R}^d$ is said to satisfy the logarithmic Sobolev inequality with a constant $C > 0$ if for every nice function φ one has

$$\mathrm{Ent}_\mu(\varphi^2) \le 2C \int_{\mathbb{R}^d} (\nabla\varphi)^2\,d\mu, \tag{1.5}$$

where $\mathrm{Ent}_\mu\psi$ is called the entropy of ψ with respect to μ and defined by

$$\mathrm{Ent}_\mu\psi := \int_{\mathbb{R}^d} \psi\log\psi\,d\mu - \Bigl(\int_{\mathbb{R}^d}\psi\,d\mu\Bigr)\log\int_{\mathbb{R}^d}\psi\,d\mu.$$

Recall that the standard Gaussian measure on $\mathbb{R}^d$ satisfies (1.5) with $C = 1$. If we apply (1.5) to $\sqrt{g}$, where g is a nonnegative function such that $\int_{\mathbb{R}^d} g\,d\mu = 1$, we get another equivalent form of (1.5):

$$\int_{\mathbb{R}^d} g\log g\,d\mu \le \frac{1}{2}I_\mu g,$$

where

$$\mathrm{I}_\mu\varphi := \int_{\mathbb{R}^d} \frac{\langle\nabla\varphi,\nabla\varphi\rangle}{\varphi}\,d\mu.$$

$\mathrm{I}_\mu\varphi$ is called the Fisher information.

Otto and Villani [28] have extended inequality (1.1) to the measures satisfying the logarithmic Sobolev inequality. More precisely, if μ satisfies (1.5), then

$$W_2^2(\mu,\nu) \le 2C\mathrm{Ent}_\mu\Bigl(\frac{d\nu}{d\mu}\Bigr).$$

A short proof was given by Bobkov, Gentil and Ledoux [4]. Another interesting generalization of (1.1) and (1.5) is due to Cordero-Erausquin [15] (see also [16]).

A probability measure μ satisfies the Poincaré inequality (or the spectral gap inequality) with a constant $C > 0$ if for every smooth φ one has

$$\mathrm{Var}_\mu\varphi \le C\int_{\mathbb{R}^d}\langle\nabla\varphi,\nabla\varphi\rangle\,d\mu, \tag{1.6}$$

where

$$\mathrm{Var}_\mu\varphi := \int_{\mathbb{R}^d}\varphi^2\,d\mu - \Bigl(\int_{\mathbb{R}^d}\varphi\,d\mu\Bigr)^2.$$

It is easy to verify that every measure satisfying (1.5) satisfies (1.6). Indeed, if we apply (1.5) to the function $1+\varepsilon\varphi$, we arrive at (1.6) in the limit as $\varepsilon \to 0$.

The existence of an optimal mapping T was studied in the infinite dimensional case too. Some interesting results were obtained in [17], [29], [31]. In this paper we develop another point of view on this problem. We emphasize that in the infinite dimensional case the standard choice of the minimized functional in (1.3) is not connected with ν_1, ν_2. On the other hand, it is well known (see [6]) that for every Radon Gaussian measure γ on a locally convex space X, there exists a continuously embedded Hilbert space $H \subset X$ of vectors of quasi-invariance of γ. This space H is endowed with its natural norm $\|\cdot\|_H$ (see [6]) and is called the Cameron–Martin space of γ. Many known mappings that transport γ to an equivalent measure have the form $x \to x + F(x)$, where $F(x)$ takes values in H. The simplest mappings of such a type are the shifts $x \to x+h$ along vectors from H. The density $\frac{d\gamma(\cdot+h)}{d\gamma}$ is given by the Cameron–Martin formula. Other examples are given by the Girsanov type transformations and measurable linear mappings of the type $F = E + K$, where K is a measurable extension of a Hilbert–Schmidt operator in H. A number of examples can be found in the recent books [6] and [32]. Fernique [18] has shown that for every probability measure $g \cdot \gamma$, there exists a mapping of the form $T = U + F$, where U is an automorphism of γ and F takes values in H, such that $g \cdot \gamma = \gamma \circ T^{-1}$. Therefore, given a Gaussian measure γ on X and a probability measure $\mu = g \cdot \gamma$, it is natural to consider the problem of minimizing the following modified Monge–Kantorovich functional

$$\nu \to \Bigl[\int_{X\times X} (x_1 - x_2)^2_H \, d\nu(x_1, x_2)\Bigr]^{\frac{1}{2}},$$

where the marginals of ν are, accordingly, γ and μ. We denote by $W_{H,2}(\gamma, \mu)$ the corresponding modified transportation cost.

Feyel and Üstünel [20] obtained a remarkable result that for every Gaussian measure γ and any probability measure $\mu \ll \gamma$ satisfying the additional assumption $W_{H,2}(\gamma, \mu) < \infty$, there exists a mapping T sending γ to μ such that $T(x) = x + F(x)$, $F\colon\ X \to H$ and F is the H-gradient of some 1-convex function φ (see the definition below). The mapping T is H-monotone, i.e. $\bigl\langle T(x+h) - T(x), h\bigr\rangle_H \geq 0$ for γ-almost all x and every $h \in H$, and admits an inverse mapping T^{-1}, i.e. $T\circ T^{-1} = T^{-1}\circ T = E$ γ-a.e. Another class of mappings, the so called triangular mappings, have been recently investigated in Bogachev, Kolesnikov, Medvedev [8]. Although such mappings are typical not optimal, they provide an efficient construction of transformation of measures. In particular, it has been shown in [8] that given a centered Radon Gaussian measure γ and a probability measure $\nu \ll \gamma$, there is a mapping F with values in the Cameron–Martin space of γ such that $\nu = \gamma \circ (I + F)^{-1}$. This shows that ν is representable in the sense of [32]. The approach in [8] develops a result from Kolesnikov [25] that extends Talagrand's inequality in the form of equality (contained in the one dimensional case already in Talagrand's paper).

In this work we prove the results announced in [25] and give several other results. Our principal results are based on the above mentioned equality which generalizes Talagrand's inequality and is proved in Section 2, where we also give a generalization of (1.1) to the case of non-Gaussian (e.g. convex) measures and a couple of optimal transfer plans. A partial case of our results was established for uniformly convex measures in [15] (see Remark 2.3 below). It is worth noting that these results also apply to mappings of other types.

Another result of the paper is a new log-Sobolev type inequality for strictly convex measures, i.e., measures with densities $e^{-\Phi}$, where $D^2\Phi > 0$ (see Section 3). We emphasize that no uniform bound $D^2\Phi > C > 0$ is required. This log-Sobolev type inequality implies a number of other well-known related inequalities (e.g., Brascamp–Lieb, HWI).

In particular, the Brascamp-Lieb inequality follows directly from our result exactly in the same way as the standard Poincaré inequality follows from (1.5).

In Section 4 we extend our finite-dimensional results to the infinite-dimensional case. In particular, we prove the generalized Talagrand inequality. As a consequence we easily get an important partial case of the general result from [20] on existence of an H-monotone mapping T that transports a probability measure $g \cdot \gamma$ to γ (under some integrability restrictions on g). Let us call mappings of such a type "backward" mappings, in contrast to "forward" mappings that are optimal transfer plans sending γ to $g \cdot \gamma$. It is worth noting that the construction of "backward" mappings is in some respects simpler as compared to "forward" mappings. This is due to the fact that the finite-dimensional Talagrand type estimates generalized to the case of a couple of optimal transports retain its simple and applicable form only for "backward" mappings. In particular, the proof of existence of a "backward" mappings becomes very easy. Moreover, we obtain useful estimates of the L_2-distance between two different optimal mappings via entropy-like functions of the densities. The author does not know whether "forward" mappings can be constructed in a similar way. Such "backward" mappings are useful for applications. For example, they appear in the log-Sobolev inequality (see 3.9).

Note that the problem of almost sure convergence of optimal transports in the infinite dimensional case is rather delicate. This problem is connected with a generalization of the classical Skorokhod representation theorem due to Blackwell and Dubins (see [2], [7]). For some partial results on the Skorokhod representation based on W_2-distance, see [17], [31].

In Section 5 we show under the additional assumption $0 < c < g < C$ that $-T$ is the logarithmic gradient of a convex probability measure ν that is absolutely continuous with respect to γ, and, moreover, its density can be obtained as the $L^1(\gamma)$-limit of the corresponding finite-dimensional approximations. The logarithmic gradients of measures are very important in the infinite-dimensional analysis (see Bogachev [6], Bogachev, Röckner [9]). In the proof we establish some useful estimates.

In Section 6 we give some estimates of the second derivatives of T. The following problem remains open: when does T admit a regular weak derivative? We show that this is the case if the second derivative of $-\log g$ is uniformly bounded from below. We also show that if $g > c > 0$, then the weak derivative of T can be estimated from below by some integrable operator-valued mapping with values in the set of invertible positive symmetric operators.

We assume throughout that X is a locally convex Suslin space and γ is a centered Radon Gaussian measure on X with the covariance operator R_γ. Recall that a Borel probability measure μ on a locally convex Suslin space is called convex if for every Borel sets A, B the following inequality holds:

$$\mu(\lambda A + (1-\lambda)B) \geq \mu(A)^{\lambda}\mu(B)^{1-\lambda}.$$

The general theory of such measures was developed by Borell [10]. The non-degenerate finite dimensional convex measures can be characterized as measures with log-concave densities, i.e., densities of the form e^{-V}, where V is a convex function on $\mathbb{R}^d$.

2. L^2-distance between transfer plans

Let $\mu = e^{-\Phi}\, dx$ be a probability measure such that $\Phi : \mathbb{R}^d \to \mathbb{R}$ is twice continuously differentiable and let f, g be smooth non-negative functions on $\mathbb{R}^d$. We assume that $f \cdot \mu$ and $g \cdot \mu$ are probability measures such that $\int_{\mathbb{R}^d} \|x\|^2 f\, d\mu < \infty$, $\int_{\mathbb{R}^d} \|x\|^2 g\, d\mu < \infty$,

$\int_{\mathbb{R}^d} \|x\|^2 \, d\mu < \infty$ and $g > 0$ everywhere. The identity mapping will be denoted by E or Id.

Let V and W be convex functions such that ∇V and ∇W transform μ to $f \cdot \mu$ and $g \cdot \mu$, respectively. The fact that g is smooth and positive yields some regularity properties of W (see [13] for details). Changing variables we have

$$g(\nabla W) e^{-\Phi(\nabla W)} \det D^2 W = e^{-\Phi}.$$

This identity implies that $\det D^2 W$ is positive, hence $D^2 W$ is non-degenerate everywhere. We obtain from the relation

$$\big\langle \nabla W(a + tv) - \nabla W(a), v \big\rangle = \int_0^t \big\langle D^2 W(a + sv) v, v \big\rangle \, ds$$

that the mapping $\nabla W(x) = y$ is injective. Since $g \cdot \mu$ has the full support, there exists a mapping $\omega(y)$ such that

$$\nabla W(\omega(y)) = y$$

for almost all y. It is well-known that $\omega = \nabla W^*$ almost everywhere, where W^* is the convex conjugated function. Note that ∇W^* is an optimal transfer plan between $g \cdot \mu$ and μ. This follows from the uniqueness of an optimal transfer plan and the monotonicity of ∇W^*. By the same reasoning ∇W^* is also injective and regular, hence $\nabla W(\nabla W^*(y)) = y$ everywhere.

Throughout the paper ∇W denotes the optimal transfer plan taking μ to $g \cdot \mu$ (the "forward" optimal transfer plan) and ∇W^* denotes its inverse, i.e., the optimal transfer plan taking $g \cdot \mu$ to μ (the "backward" optimal transfer plan). Recall, that given a probability measure $\varrho \, dx$ on $\mathbb{R}^d$ its logarithmic gradient is defined by $\frac{\nabla \varrho}{\varrho}$.

Let γ_d be the standard Gaussian measure on $\mathbb{R}^d$. Then the following remarkable relation holds:

$$\int_{\mathbb{R}^d} f \log f \, d\gamma_d = \frac{1}{2} \int_{\mathbb{R}^d} (x - \nabla V)^2 \, d\gamma_d - \int_{\mathbb{R}^d} \log \det{}_2 (D^2 V) \, d\gamma_d, \tag{2.7}$$

where $\det_2 A = \det A \exp(\mathrm{Tr}(E - A))$ is the so-called Fredholm–Carleman determinant of A. It is easy to verify that $\det_2 A \le 1$ for every nonnegative A. This formula can be easily deduced from the change of variables formula for Gaussian measures (see [6]), but can also verified directly as we do below in the general case. The main result of this section is the following generalization of (2.7).

Theorem 2.1. *The following identity holds:*

$$\begin{aligned} Ent_{g \cdot \mu}\Big(\frac{f}{g}\Big) &= \int_{\mathbb{R}^d} \big\langle \Lambda\big[g e^{-\Phi}, \nabla V, \nabla W\big] \big(\nabla V - \nabla W\big), \nabla V - \nabla W \big\rangle \, d\mu \\ &\quad - \int_{\mathbb{R}^d} \log \det{}_2 \sqrt{(D^2 W)^{-1}} D^2 V \sqrt{(D^2 W)^{-1}} \, d\mu, \end{aligned}$$

where $\Lambda\big[\psi, v_1, v_2\big]$ *is the operator-valued mapping on* $\mathbb{R}^d$ *defined as follows:*

$$\Lambda\big[\psi, v_1, v_2\big] = \int_0^1 s D^2\big[-\log \psi\big]\big((1 - s) v_1 + s v_2\big) \, ds, \quad v_1, v_2 \in \mathbb{R}^d.$$

Proof. Let $T = \nabla V \circ \nabla W^{-1} = \nabla V \circ \nabla W^*$. Then T maps $g \cdot \mu$ to $f \cdot \mu$, hence satisfies the following equation:

$$f(T(y)) e^{-\Phi(T(y))} |\det DT(y)| = g(y) e^{-\Phi(y)}.$$

Note that $E = D(\nabla W \circ \nabla W^*(y)) = D^2 W(\nabla W^*(y)) D^2 W^*(y)$. Hence

$$DT = D^2 V \circ D^2 W^* = D^2 V (D^2 W)^{-1} (\nabla W^*(y)).$$

Taking logarithm of both sides we get

$$\log f(T(y)) - \Phi(T(y)) + \log\det D^2V(D^2W)^{-1}(\nabla W^*(y)) = \log g(y) - \Phi(y).$$

Adding $(\Phi - \log g)(T)$ to both sides we have

$$\begin{aligned}\big[\log f(T(y)) - \log g(T(y))\big] = \big[\Phi(T(y)) - \log g(T(y)) - \Phi(y) + \log g(y)\big] \\ - \log\det D^2V(D^2W)^{-1}(\nabla W^*(y)).\end{aligned}$$

Let us apply the identity

$$\Psi(a) - \Psi(b) = \big\langle \nabla\Psi(b), a-b\big\rangle + \int_0^1 s\big\langle D^2\Psi((b-a)s+a)(b-a), b-a\big\rangle\, ds \tag{2.8}$$

to $\Psi = \Phi - \log g$ and $a = T(y)$, $b = y$.

Integrating the obtained formula with respect to $g\cdot\mu$ and using that $(g\cdot\mu)\circ(\nabla W^*)^{-1} = \mu$, we obtain

$$\begin{aligned}&\int_{\mathbb{R}^d} \log\frac{f(\nabla V)}{g(\nabla V)}\, d\mu = \int_{\mathbb{R}^d} \langle \nabla(\Phi - \log g), T - y\rangle g\, d\mu + \\ &\int_{\mathbb{R}^d}\Big[\int_0^1 s\big\langle D^2(\Phi - \log g)(sy + (1-s)T)\big(y - T\big), y - T\big\rangle\, ds\Big] g\, d\mu \\ &- \int_{\mathbb{R}^d} \log\det D^2V(D^2W)^{-1}(\nabla W^*) g\, d\mu.\end{aligned}$$

Note that $-\nabla(\Phi - \log g)$ is the logarithmic gradient of $ge^{-\Phi}$. Integrating by parts and changing variables we get

$$\begin{aligned}&\int_{\mathbb{R}^d} \log\frac{f}{g} f\, d\mu = \int_{\mathbb{R}^d} \langle \Lambda\big[ge^{-\Phi}, \nabla V, \nabla W\big]\big(\nabla V - \nabla W\big), \nabla V - \nabla W\rangle\, d\mu \\ &+ \int_{\mathbb{R}^d}\Big(\mathrm{Tr} DT - d - \log\det DT\Big)(\nabla W^*) g\, d\mu.\end{aligned}$$

Finally, we arrive at

$$\begin{aligned}&\int_{\mathbb{R}^d}\Big(\mathrm{Tr}DT - d - \log\det DT\Big)(\nabla W^*) g\, d\mu = \int_{\mathbb{R}^d}\Big(\mathrm{Tr}DT - d - \log\det DT\Big)\, d\mu \\ &= \int_{\mathbb{R}^d}\Big(\mathrm{Tr}D^2V(D^2W)^{-1} - d - \log\det D^2V(D^2W)^{-1}\Big)\, d\mu \\ &= \int_{\mathbb{R}^d}\Big[\mathrm{Tr}(D^2W)^{-\frac{1}{2}}D^2V(D^2W)^{-\frac{1}{2}} - d - \log\det(D^2W)^{-\frac{1}{2}}D^2V(D^2W)^{-\frac{1}{2}}\Big]\, d\mu = \\ &- \int_{\mathbb{R}^d} \log\det{}_2(D^2W)^{-\frac{1}{2}}D^2V(D^2W)^{-\frac{1}{2}}\, d\mu.\end{aligned}$$

The proof is complete. □

Theorem 2.2. *The following identity holds for the "backward" optimal transfer plans ∇V^* and ∇W^*:*

$$\begin{aligned}Ent_{f\cdot\mu}\Big(\frac{g}{f}\Big) &= \int_{\mathbb{R}^d} \langle \Lambda\big[e^{-\Phi}, \nabla V^*, \nabla W^*\big]\big(\nabla V^* - \nabla W^*\big), \nabla V^* - \nabla W^*\rangle g\, d\mu \\ &- \int_{\mathbb{R}^d} \log\det{}_2\big(\sqrt{(D^2W^*)^{-1}}D^2V^*\sqrt{(D^2W^*)^{-1}}\big) g\, d\mu.\end{aligned}$$

Moreover, for every nice function $\varphi\colon \mathbb{R}^d \to \mathbb{R}$

$$\int_{\mathbb{R}^d} \varphi(\nabla W^*) g \log \frac{g}{f}\, d\mu = \int_{\mathbb{R}^d} \langle \nabla V^* - \nabla W^*, \nabla \varphi(\nabla W^*) \rangle g \, d\mu$$

$$+ \int_{\mathbb{R}^d} \varphi(\nabla W^*) \langle \Lambda \big[e^{-\Phi}, \nabla V^*, \nabla W^* \big] \big(\nabla V^* - \nabla W^* \big), \nabla V^* - \nabla W^* \rangle g \, d\mu$$

$$- \int_{\mathbb{R}^d} \varphi(\nabla W^*) \log \det{}_2 \big(\sqrt{(D^2 W^*)^{-1}} D^2 V^* \sqrt{(D^2 W^*)^{-1}} \big) g \, d\mu.$$

The proof of Theorem 2.2 follows the same lines as that of Theorem 2.1, but working with the "backward" instead of "forward" mappings. The first identity can be also established by a direct application of Theorem 2.1 to the measure $g \cdot \mu$ and to the optimal transfer plans $\nabla V^*, \nabla W^*$ (i.e., we replace μ by $g \cdot \mu$ and V, W by V^*, W^* in the formulation of the theorem). Note that the "backward" identities do not involve the derivatives of g in contrast to the "forward" ones. This will be the key observation needed for the infinite-dimensional results in the following sections.

Yet another useful result can be obtained if we repeat the proof of Theorem 2.1 and apply (2.8) to Φ in place of $\Phi - \log f$. The partial case of this estimate (for $\mu = e^{-\Phi}\, dx$ satisfying $D^2\Phi > C > 0$) has been obtained in [15].

Proposition 2.3. *The following inequality holds:*

$$\int_{\mathbb{R}^d} (f \log f - g \log g)\, d\mu \geq \int_{\mathbb{R}^d} \Big\langle \nabla g(\nabla W), \nabla V - \nabla W \Big\rangle \frac{1}{g(\nabla W)}\, d\mu +$$
$$\int_{\mathbb{R}^d} \langle \Lambda [e^{-\Phi}, \nabla V, \nabla W] \big(\nabla V - \nabla W \big), \nabla V - \nabla W \rangle \, d\mu.$$

3. Log-Sobolev, Brascamp–Lieb and HWI inequalities

In this section we assume that Φ satisfies the assumptions of Section 2.

Corollary 3.1. *If Φ is strictly convex (i.e., $D^2\Phi > 0$), the following inequalities hold for every smooth positive function φ:*

1) *Log-Sobolev Inequality*

$$Ent_{\mu} \varphi^2 \leq \int_{\mathbb{R}^d} \Big\langle \Lambda^{-1} [e^{-\Phi}, \nabla W^*(y), y] \nabla \varphi, \nabla \varphi \Big\rangle \, d\mu, \tag{3.9}$$

where ∇W^* is the optimal transport of $\frac{\varphi^2}{\int_{\mathbb{R}^d} \varphi^2 \, d\mu} \cdot \mu$ to μ.

2) *the Brascamp–Lieb inequality*

$$Var_{\mu} \varphi \leq \int_{\mathbb{R}^d} \big\langle (D^2 \Phi)^{-1} \nabla \varphi, \nabla \varphi \big\rangle \, d\mu \tag{3.10}$$

3) *Suppose that for every $h \in \mathbb{R}^d$, the function $y \to \big\langle D^2\Phi(y) h, h \big\rangle$ is concave on a convex set $\Omega \subset \mathbb{R}^d$. Then*

$$Ent_{\mu_\Omega} \varphi^2 \leq 3 \int_{\mathbb{R}^d} \Big\langle (D^2 \Phi)^{-1} \nabla \varphi, \nabla \varphi \Big\rangle \, d\mu_\Omega, \tag{3.11}$$

where $\mu_\Omega := \frac{1}{\mu(\Omega)} I_\Omega(y) \mu(\, dy)$.

Suppose that $D^2\Phi \geq K \cdot \mathrm{Id}$, where $K \in \mathbb{R}$ is not necessarily positive. Then for every positive g such that $g \cdot \mu$ is a probability measure, the following inequality holds:

4) *HWI-inequality*

$$Ent_\mu g \le W_2(g\cdot\mu,\mu)\sqrt{I_\mu(g)} - \frac{K}{2}W_2^2(g\cdot\mu,\mu). \tag{3.12}$$

Proof. By the standard argument we reduce the proof to the case where $\varphi=\sqrt{g}$ and $g\cdot\mu$ is a probability measure. Let us apply Proposition 2.3 to $f=1$. Note that $\nabla V(x)=x$. Changing variables $y=\nabla W(x)$ and taking into account that ∇W sends μ to $g\cdot\mu$, we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^d} g\log g\,d\mu \le \\
&-\int_{\mathbb{R}^d}\bigl\langle\nabla g,\nabla W^*-y\bigr\rangle\,d\mu - \int_{\mathbb{R}^d}\bigl\langle\Lambda[e^{-\Phi},\nabla W^*,y]\bigl(\nabla W^*-y\bigr),\nabla W^*-y\bigr\rangle g\,d\mu\\
&=-\int_{\mathbb{R}^d}\|\sqrt{\Lambda}[e^{-\Phi},\nabla W^*,y]\bigl(\nabla W^*-y\bigr)+\frac{1}{2}\sqrt{\Lambda^{-1}}[e^{-\Phi},\nabla W^*,y]\Bigl(\frac{\nabla g}{g}\Bigr)\|^2 g\,d\mu\\
&+\frac{1}{4}\int_{\mathbb{R}^d}\bigl\langle\Lambda^{-1}[e^{-\Phi},\nabla W^*,y]\nabla g,\frac{\nabla g}{g}\bigr\rangle\,d\mu \le \frac{1}{4}\int_{\mathbb{R}^d}\bigl\langle\Lambda^{-1}[e^{-\Phi},\nabla W^*,y]\nabla g,\frac{\nabla g}{g}\bigr\rangle\,d\mu.
\end{aligned}$$

Thus we have obtained a log-Sobolev type inequality for convex measures.

Let us prove (3.10). To this end, we apply (3.9) to $\sqrt{\frac{1+\varepsilon\varphi}{1+\varepsilon\int_{\mathbb{R}^d}\varphi\,d\mu}}$ and let $\varepsilon\to 0$.

Now let us perturb μ by a log-concave Radon-Nikodym density, i.e., consider the probability measure $\tilde{\mu}=e^{-\phi}\mu$, where ϕ is convex. The inequality

$$\Lambda\Bigl[\frac{d\tilde{\mu}}{dx},v_1,v_2\Bigr] = \int_0^1 sD^2(\Phi+\phi)((1-s)v_1+sv_2) \ge \Lambda\Bigl[\frac{d\mu}{dx},v_1,v_2\Bigr]$$

yields

$$\mathrm{Ent}_{\tilde{\mu}}\varphi^2 \le \int_{\mathbb{R}^d}\bigl\langle\Lambda^{-1}[e^{-\Phi},\nabla W^*(y),y]\nabla\varphi,\nabla\varphi\bigr\rangle\,d\tilde{\mu}. \tag{3.13}$$

Let us take a convex set Ω and approximate I_Ω by smooth log-concave functions $e^{-\Phi_n}\to I_\Omega$. We obtain that (3.13) holds for the conditional measure $\tilde{\mu}=\mu_\Omega=\frac{1}{\mu(\Omega)}I_\Omega(y)\mu(\,dy)$. Suppose that Φ satisfies the additional assumptions of item 3). Note that $\nabla W^*(y)\in\Omega$ since μ_Ω is supported by Ω, hence $s\nabla W^*(y)+(1-s)y\in\Omega$ and

$$\begin{aligned}
&\bigl\langle\Lambda[e^{-\Phi},\nabla W^*,y]h,h\bigr\rangle = \int_0^1 s\bigl\langle D^2\Phi((1-s)\nabla W^*+sy)h,h\bigr\rangle\,ds\\
&\ge\int_0^1\Bigl(s(1-s)\bigl\langle D^2\Phi(\nabla W^*)h,h\bigr\rangle + s^2\bigl\langle D^2\Phi(y)h,h\bigr\rangle\Bigr)\,ds\\
&\ge\int_0^1 s^2\bigl\langle D^2\Phi(y)h,h\bigr\rangle\,ds = \frac{1}{3}\bigl\langle D^2\Phi(y)h,h\bigr\rangle.
\end{aligned}$$

The proof of (3.11) is complete.

The HWI-inequality (3.12) can be easily deduced from the inequality

$$\begin{aligned}
&\int_{\mathbb{R}^d} g\log g\,d\mu \le -\int_{\mathbb{R}^d}\bigl\langle\nabla g,\nabla W^*-y\bigr\rangle\,d\mu\\
&-\int_{\mathbb{R}^d}\bigl\langle\Lambda[e^{-\Phi},\nabla W^*,y]\bigl(\nabla W^*-y\bigr),\nabla W^*-y\bigr\rangle g\,d\mu
\end{aligned}$$

obtained in the proof of (3.9) and the Cauchy inequality. □

Let us comment on the obtained results.

Inequality (3.9) can be considered as a natural generalization of the classical log-Sobolev inequality for arbitrary convex measures. In particular, it implies the classical result that a probability convex measure $e^{-\Phi}\,dx$ with $D^2\Phi \geq \frac{1}{C}\cdot \mathrm{Id}$, $C>0$, satisfies the log-Sobolev inequality with C.

Inequality (3.10) was obtained by Brascamp and Lieb in [12]. One can ask whether the analogous generalization of the log-Sobolev inequality holds, i.e., whether the estimate

$$\frac{1}{2}\mathrm{Ent}_\mu(\varphi^2) \leq \int_{\mathbb{R}^d} \big\langle (D^2\Phi)^{-1}\nabla\varphi, \nabla\varphi\big\rangle\, d\mu \tag{3.14}$$

holds for every nice $\varphi : \mathbb{R}^d \to \mathbb{R}$. Surprisingly, the answer is negative. Inequality (3.11) was proved by Bobkov and Ledoux in [5], where they also gave a counterexample to (3.14). Note that (3.9) and the Brascamp–Lieb inequality are connected in the same way as the classical log-Sobolev and Poincaré inequalities.

The HWI inequality (3.12) was obtained in [28] (see [4] for a short proof).

Example 3.2. *Let $V\colon \mathbb{R}^+ \to \mathbb{R}$ be a convex increasing function and let $\mu = e^{-V\left(\frac{r^2}{2}\right)}\,dx$ be a probability measure on $\mathbb{R}^d$. Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be an even (or odd) smooth function. Then the following inequality holds:*

$$Ent_\mu f^2 \leq \int_{\mathbb{R}^d} \frac{1}{v\left(\frac{r^2}{2}\right)} \big\langle \nabla f, \nabla f\big\rangle\, d\mu,$$

where $v(\alpha) = \int_0^1 sV'(s^2\alpha)\,ds$.

Proof. Note that $\frac{\partial^2}{\partial_{x_i}\partial_{x_j}}V(\frac{r^2}{2}) = V''(\frac{r^2}{2})x_i x_j + V'\delta_{ij}$. Hence $D^2\big(V(\frac{r^2}{2})\big) \geq V'(\frac{r^2}{2})I$. The optimal transport ∇W between μ and $\frac{f^2}{\int_{\mathbb{R}^d} f^2\,d\mu}\cdot\mu$ is antisymmetric since f^2 is even. Since W^* is convex, we have $\big\langle \nabla W^*(x), x\big\rangle \geq 0$ and

$$\big\langle (1-s)\nabla W^*(x) + sx\big\rangle^2 = (1-s)^2\|\nabla W^*(x)\|^2 + 2s(1-s)\big\langle \nabla W^*(x), x\big\rangle + s^2x^2 \geq s^2x^2.$$

The function V' is increasing, hence $\Lambda(e^{-\Phi}, \nabla W^*, x) \geq \Big(\int_0^1 sV'\big(\frac{s^2}{2}r^2\big)\,ds\Big)I$. Applying (3.9) we complete the proof. ☐

Now we give a very simple estimate of the transportation cost for measures satisfying the Poincaré inequality.

Proposition 3.3. *Let μ have an a.e. positive density with respect to Lebesgue measure and satisfy (1.6), let f be an a.e. positive function such that $f\cdot\mu$ is a probability measure and let $f \in L^2(\mu)$. Suppose that $\int_{\mathbb{R}^d}\|x\|^2\,d\mu < \infty$ and $\int_{\mathbb{R}^d}\|x\|^2 f\,d\mu < \infty$. Then*

$$\frac{1}{4C}W_2^2(\mu, f\cdot\mu) \leq \int_{\mathbb{R}^d} f^2\,d\mu - 1.$$

Proof. Let ∇V be the optimal transfer plan between $f\cdot\mu$ and μ. Note that $V^*\in L^1(\mu)$. Indeed,

$$\int_{\mathbb{R}^d} V^*\,d\mu \le \int_{\mathbb{R}^d} V^*(0)\,d\mu + \int_{\mathbb{R}^d} \langle \nabla V^*, x\rangle\,d\mu \le V^*(0)$$
$$+\int_{\mathbb{R}^d} \Big\langle \nabla V^* - x, x\Big\rangle\,d\mu + \int_{\mathbb{R}^d} \|x\|^2\,d\mu \le V^*(0) + \int_{\mathbb{R}^d} \|x\|^2\,d\mu$$
$$+\sqrt{\int_{\mathbb{R}^d} \|x\|^2\,d\mu}\sqrt{\int_{\mathbb{R}^d} (\nabla V^* - x)^2\,d\mu} = V^*(0) + \int_{\mathbb{R}^d} \|x\|^2\,d\mu$$
$$+ W_2(\mu, f\,d\mu)\sqrt{\int_{\mathbb{R}^d} \|x\|^2\,d\mu} < \infty.$$

The function V is defined up to a constant C. Choose C in such a way that $\int_{\mathbb{R}^d} V^*\,d\mu = \frac{1}{2}\int_{\mathbb{R}} x^2\,d\mu$. Applying the well known identities $V(x) + V^*(\nabla V) = \langle x, \nabla V\rangle$ and $V(x) + V^*(y) \ge \langle x, y\rangle$ we obtain

$$\frac{1}{2}W_2^2(\mu, f\cdot\mu) = \frac{1}{2}\int_{\mathbb{R}^d} (x - \nabla V)^2 f\,d\mu = \int_{\mathbb{R}^d} \Big(\frac{x^2}{2} - V(x) - V^*(\nabla V) + \frac{\nabla V^2}{2}\Big) f\,d\mu$$
$$= \int_{\mathbb{R}^d} \Big(\frac{x^2}{2} - V^*\Big)\,d\mu - \int_{\mathbb{R}^d} \Big(V - \frac{x^2}{2}\Big)\,f\,d\mu$$
$$\le \int_{\mathbb{R}^d} \Big(\frac{x^2}{2} - V^*\Big)\,d\mu - \int_{\mathbb{R}^d} \Big(\frac{x^2}{2} - V^*\Big)\,f\,d\mu$$
$$= \int_{\mathbb{R}^d} \Big(\frac{x^2}{2} - V^*\Big)(1-f)\,d\mu \le \Big[\int_{\mathbb{R}^d} \Big(\frac{x^2}{2} - V^*\Big)^2\,d\mu \int_{\mathbb{R}^d} (1-f)^2\,d\mu\Big]^{\frac{1}{2}}$$
$$\le \Big[C\int_{\mathbb{R}^d} \Big(x - \nabla V^*\Big)^2\,d\mu \int_{\mathbb{R}^d} (1-f)^2\,d\mu\Big]^{\frac{1}{2}}$$
$$= \Big[C\int_{\mathbb{R}^d} \Big(x - \nabla V^{-1}\Big)^2\,d\mu \int_{\mathbb{R}^d} (1-f)^2\,d\mu\Big]^{\frac{1}{2}}$$
$$= \sqrt{C}\Big[\int_{\mathbb{R}^d} f^2\,d\mu - 1\Big]^{\frac{1}{2}} W_2(\mu, f\,d\mu).$$

Hence $\frac{1}{2}W_2(\mu, f\cdot\mu) \le \sqrt{C}\big[\int_{\mathbb{R}^d} f^2\,d\mu - 1\big]^{\frac{1}{2}}$. The proof is complete. □

Note that there exist estimates of the type $W_L(\mu, f\cdot\mu) \le C\int_X f\log f\,d\mu$ for measures satisfying Poincaré inequalities (e.g., the exponential measures) and an appropriate cost function $L(x_1 - x_2)$ (see [4], [30]). Typically, the cost function $L(x_1 - x_2)$ is quadratic for $x\in[0, A]$ and some $A > 0$ and linear for $x > A$.

4. **Infinite-dimensional transfer plans for Gaussian measures**

In this section, we consider an infinite dimensional centered Radon Gaussian measure γ on a locally convex Suslin space X and H denotes its Cameron–Martin space.

We denote by $\mathcal{L}$ the Hilbert space of all square integrable vector fields, i.e., the space of square integrable mappings $v : X \to H$ endowed with the norm $\int_X \|v\|_H^2\,d\gamma$. We denote by X^*_γ the completion of the dual space X^* in the $L^2(\gamma)$-norm. As usual we denote by $\mathcal{F}C_0^\infty$ the space of smooth cylindrical functions, i.e., the functions of the form $\varphi(l_1(x), \cdots, l_k(x))$, where $\varphi\in C_0^\infty(\mathbb{R}^k)$, $l_1, \ldots, l_k \in X^*$

Let us fix an orthonormal basis $\{\xi_n\}$ in X^*_γ. Then $\{e_n = R_\gamma(\xi_n)\}$ is an orthonormal basis in H. It is well known (see [6]) that to every $h \in H$ one can associate an element $\hat{h} \in X^*_\gamma$ such that $h = R_\gamma \hat{h}$. Let P_n denote the projection of X on the subspace generated by $\{e_1, \ldots, e_n\}$, i.e., $P_n(x) = \sum_{i=1}^n \xi_i(x) e_i$. The conditional expectation f_n of $f \in L^2(\gamma)$ with respect to the sigma-algebra $\mathcal{F}_n$ generated by $\{e_1, \cdots, e_n\}$ has the form

$$f_n(x) = \int_X f(P_n x + S_n y)\gamma(\,dy),$$

where $S_n y = \sum_{i=n+1}^\infty \xi_i(y) e_i$. It is well-known (see [6]) that H and X^*_γ are separable.

We say that a measure μ admits a logarithmic derivative along h if there exists a function $\beta^\mu_h(x) \in L^1(\mu)$ such that for every $\varphi \in \mathcal{F}C^\infty_0$

$$\int_X \frac{\partial \varphi}{\partial h} \mu(\,dx) = -\int_X \varphi \beta^\mu_h(x) \mu(\,dx).$$

It is well-known that $\beta^\gamma_h = -\hat{h}$. If μ is given by its smooth density ϱ with respect to γ, then $\beta^\mu_h = \frac{\partial_h \varrho}{\varrho} - \hat{h}(x)$. An integrable mapping $F : X \to \mathbb{R}$ is said to admit a weak derivative $F_h : X \to R$ along $h \in H$ if F_h is integrable and the following integration-by-parts formula holds:

$$\int_X F_h(x)\gamma(\,dx) = -\int_X F(x)\beta_h(x)\gamma(\,dx),$$

More generally, a weak derivative of F along h can be defined via the integration by parts formula for every measure μ that has the logarithmic derivative along h.

Recall that in the finite-dimensional case the logarithmic gradient of $\varrho\,dx$ is defined by $\frac{\nabla \varrho}{\varrho}$. In the infinite-dimensional case, there are no analogs of $\nabla \varrho$ and ϱ, but logarithmic gradients can by defined by the integration-by-parts formula. Let μ be absolutely continuous with respect to γ. If there exists a measurable mapping $\beta^\mu_H \colon X \to X$ such that $l(\beta^\mu_H) = \beta^\mu_{R_\gamma l}$ for every $l \in X^*$, then β^μ_H is called the logarithmic gradient of μ. For example, $\beta^\gamma_H(x) = -x$.

We show in this section that for every probability measure $\mu = g \cdot \gamma$ satisfying the conditions $\mathrm{Ent}_\gamma \frac{d\mu}{d\gamma}, \mathrm{Ent}_\mu \frac{d\gamma}{d\mu} < \infty$, there exists an H-monotone mapping T of the type $T(x) = x + F(x)$, where F takes values in H, such that T transforms $g \cdot \gamma$ to γ.

We give below a simple corollary of the finite-dimensional results obtained above. Though the corollary is a partial case of the general result from [20], we emphasize that the presented proof provides a very clear and simple analytic point of view on the infinite-dimensional optimal transfer problem.

Corollary 4.1. *Suppose that $\mu = g \cdot \gamma$ is a probability measure such that $Ent_\gamma \frac{d\mu}{d\gamma} = \int_X g \log g \, d\gamma < \infty$ and $Ent_\mu \frac{d\gamma}{d\mu} = -\int_X \log g \, d\gamma < \infty$. Then there exists a mapping $T = x + F(x) : X \to X$ such that $\gamma = (g \cdot \gamma) \circ T^{-1}$, T is H-monotone and $F \in \mathcal{L}$.*

Proof. First we consider the case where $0 < c \le g \le C$. Let us construct an approximating sequence of smooth functions g_n such that $g_n \to g$ for γ-a.e. x, $0 < c < g_n < C$ and every g_n has the form $g_n = \varphi_n(e_1, \cdots, e_n)$ for some $\varphi_n : \mathbb{R}^n \to \mathbb{R}^+$. For example, one can take $\mathbb{E}(g|\mathcal{F}_n)$ and then construct smooth approximations of $\mathbb{E}(g|\mathcal{F}_n)$ by the standard convolution procedure. We equip $\mathrm{span}\{e_1, \cdots, e_n\}$ with the Cameron–Martin norm. Then the measure $\gamma \circ P_n^{-1}$ is standard Gaussian. Consider the optimal transfer plan $T_n : \mathbb{R}^n \to \mathbb{R}^n$ between the finite dimensional measures $(g_n \cdot \gamma) \circ P_n^{-1}$ and $\gamma \circ P_n^{-1}$. Let us extend every T_n to all of X by setting $T_n(x) = T_n(P_n x) + x - P_n x$. Then $(g_n \cdot \gamma) \circ T_n^{-1} = \gamma$.

Applying Theorem 2.2 we get the following estimate

$$\frac{1}{2}\int_X (T_n - T_m)_H^2 g_n \, d\gamma \le \int_X \log \frac{g_n}{g_m} g_n \, d\gamma. \tag{4.15}$$

By the same reasoning $\frac{1}{2}\int_X (x - T_m)_H^2 \, d\gamma \le -\int_X \log g_m \, d\gamma$, hence $x - T_n$ is bounded in $\mathcal{L}$. We can extract a $\mathcal{L}$-weakly convergent subsequence $x - T_n \to x - T \in \mathcal{L}$. Let us show that $(g \cdot \gamma) \circ T^{-1} = \gamma$. Indeed, applying (4.15) we obtain by the properties of weak convergence and the Lebesgue dominated convergence theorem that

$$\begin{aligned} c \cdot \int_X (T - T_n)_H^2 \, d\gamma &\le c \cdot \underline{\lim}_m \int_X (T_m - T_n)_H^2 \, d\gamma \le \underline{\lim}_m \int_X (T_m - T_n)_H^2 g_n \, d\gamma \\ &\le \underline{\lim}_m \int_X \log \frac{g_n}{g_m} g_n \, d\gamma = \int_X \log \frac{g_n}{g} g_n \, d\gamma. \end{aligned}$$

Clearly, $T_n - x \to T - x$ in $\mathcal{L}$ and T is H-monotone. Passing to a subsequence, one can assume that $T_n - x \to T - x$ in H γ-a.e. Let us take a bounded Borel cylindrical function φ. Since $\int_X \varphi(T_n) g_n \, d\gamma = \int_X \varphi \, d\gamma$, we get by the dominated convergence theorem that $\int_X \varphi(T) g \, d\gamma = \lim_n \int_X \varphi(T_n) g_n \, d\gamma = \int_X \varphi \, d\gamma$, hence $(g \cdot \gamma) \circ T^{-1} = \gamma$. Note that estimate (4.15) remains true for the infinite dimensional case, more precisely, if $0 < c < f, g < C$ are γ-measurable functions on X, and T_f, T_g are the corresponding infinite dimensional transports, then

$$\frac{1}{2}\int_X (T_g - T_f)_H^2 g \, d\gamma \le \int_X \log \frac{g}{f} g \, d\gamma. \tag{4.16}$$

Now let $\mathrm{Ent}_\gamma(\frac{d\mu}{d\gamma}), \mathrm{Ent}_\mu \frac{d\gamma}{d\mu} < \infty$. Note that

$$\begin{aligned} \int_X |\log g| \, d\gamma &= -\int_X \log g \, d\gamma + 2\int_{g>1} \log g \, d\gamma = \mathrm{Ent}_\mu \frac{d\gamma}{d\mu} - 2\int_{g>1} \log\Bigl(\frac{1}{g}\Bigr) \frac{1}{g} g \, d\gamma \\ &\le \mathrm{Ent}_\mu \frac{d\gamma}{d\mu} + 2e^{-1} \int_{g>1} g \, d\gamma < \infty. \end{aligned}$$

Hence $\log g \in L^1(\gamma)$. Let $g_n = \frac{g \vee \frac{1}{n} \wedge n}{\int_X g \vee \frac{1}{n} \wedge n \, d\gamma}$. It follows from (4.16) that

$$\frac{1}{2}\int_X (T_{g_n} - T_{g_m})_H^2 g_n \, d\gamma \le \int_X \log \frac{g_n}{g_m} g_n \, d\gamma.$$

As above we extract an $\mathcal{L}$-weakly convergent subsequence $T_{g_n} - x \to T - x$ (denoted by $\{T_n - x\}$). $\mathcal{L}$-weak convergence $T_n - x \to T - x$ and $L_1(\gamma)$-convergence $\log g_n \to \log g$ yield that

$$\frac{1}{2}\int_X (T_n - T)_H^2 g_n \, d\gamma \le \lim_m \int_X \log \frac{g_n}{g_m} g_n \, d\gamma = \int_X \log \frac{g_n}{g} g_n \, d\gamma.$$

The obvious relations

$$\int_X g_n \log g_n \, d\gamma \to \int_X g \log g \, d\gamma \quad \text{and} \quad \int_X g_n \log g \, d\gamma \to \int_X g \log g \, d\gamma$$

imply that $\int_X (T_n - T)_H^2 g_n \, d\gamma \to 0$. Passing to a subsequence one can assume that $T_n - x \to T - x$ in H γ-a.e. This yields that $\gamma = (g \cdot \gamma) \circ T^{-1}$ and T is H-monotone. □

The proof of the following theorem is similar. It gives a natural generalization of Talagrand's inequality for two different transfer plans in the infinite-dimensional case.

Theorem 4.2. *Let $f\cdot\gamma$, $g\cdot\gamma$ be probability measures such that*

$$Ent_\gamma f\cdot\gamma<\infty,\quad Ent_\gamma g\cdot\gamma<\infty,\quad Ent_{f\cdot\gamma}\gamma<\infty,\quad Ent_{g\cdot\gamma}\gamma<\infty$$

and let T_f, T_g be the mappings constructed by Corollary 4.1. Then

$$\frac{1}{2}\int_X (T_g-T_f)^2_H g\,d\gamma \le Ent_{f\cdot\gamma}\frac{g}{f}.$$

5. Optimal transfer plans as logarithmic gradients of convex measures

In this section we show that the mapping $-T$ constructed in the previous section is the logarithmic gradient of a convex probability measure μ.

As above we consider a Suslin locally convex space X with a centered Radon Gaussian measure γ. We choose an orthonormal basis $\{e_i\}$ in H and consider the sequence of the corresponding projections P_n defined as above. The measure γ can be represented as the product measure $\gamma=\prod_{i=1}^n\gamma^i\bigoplus\tilde{\gamma}_n$, where every γ^i is the image of γ under projection $x\to\xi_i(x)e_i=x_ie_i$. Denote $\gamma_n:=\prod_{i=1}^n\gamma^i$. Let $C>g>c>0$ be the Radon-Nikodym density of some probability measure $g\cdot\gamma$ and let $T_n=\nabla W_n^*$ be the sequence of the optimal transfer plans constructed in Corollary 4.1. Let us chose the corresponding convex functions $W_n^*:\mathbb{R}^n\to\mathbb{R}$ in such a way that $\int_{\mathbb{R}^n}W_n^*\,d\gamma_n=\int_{\mathbb{R}^n}\frac{x^2}{2}\,d\gamma_n=\frac{n}{2}$ (the integrability of W_n^* can be shown in the same way as in Proposition 3.3). We define $\mu_n=\Big(\frac{1}{\sqrt{2\pi}}\Big)^n e^{-W_n^*}\,dx\bigoplus\tilde{\gamma}_n$. Clearly, $\{\mu_n\}$ is a sequence of measures equivalent to γ. It is convenient to identify μ_n with $\mu_n\circ P_n^{-1}$ and write $\int_{\mathbb{R}^n}\varphi\,d\mu_n$ in place of $\int_X\varphi\,d\mu_n$ for cylindrical functions φ depending only on P_nx. The subspace P_nX is equipped with the H-norm and all the finite-dimensional inequalities obtained below are considered in this norm.

Theorem 5.1. *Let $g\cdot\gamma$ be a probability measure such that $C>g>c>0$, where C and c are constants, and let T be the mapping constructed in Corollary 4.1. Then there exists a convex probability measure ν such that $-T$ is the logarithmic gradient of ν.*

Proof. First of all we estimate the total variation of μ_n. By Jensen's inequality

$$\mu_n(X)=\Big(\frac{1}{\sqrt{2\pi}}\Big)^n\int_{\mathbb{R}^n}e^{-W_n^*}\,dx=\int_{\mathbb{R}^n}e^{\frac{x^2}{2}-W_n^*}\,d\gamma_n\ge e^{\int_{\mathbb{R}^n}\left(\frac{x^2}{2}-W_n^*\right)d\gamma_n}=1$$

we obtain an estimate from below. For an estimate from above we use the following infimum-convolution inequality that is equivalent to the transportation cost inequality (see [4], [26]):

$$\int_{\mathbb{R}^n}e^{Qf}\,d\gamma_n\le e^{\int_{\mathbb{R}^n}f\,d\gamma_n},$$

where $Qf=\inf_{y\in\mathbb{R}^n}\Big[f(y)+\frac{1}{2}|x-y|^2\Big]$. In particular $Q\big(W_n-\frac{x^2}{2}\big)=\frac{x^2}{2}-W_n^*$. Hence

$$\mu_n(X)=\int_{\mathbb{R}^n}e^{\frac{x^2}{2}-W_n^*}\,d\gamma_n\le\exp\int_{\mathbb{R}^n}\Big(W_n-\frac{x^2}{2}\Big)\,d\gamma_n=\exp\int_{\mathbb{R}^n}\Big(W_n-\frac{x^2}{2}\Big)(\nabla W_n^*)g_n\,d\gamma_n$$

By using that $W_n(\nabla W_n^*)+W_n^*=\langle x,\nabla W_n^*\rangle$ we get:

$$\mu_n(X)\le\exp\int_{\mathbb{R}^n}\Big(\langle x,\nabla W_n^*\rangle-W_n^*-\frac{\nabla(W_n^*)^2}{2}\Big)g_n\,d\gamma_n\le\exp\int_{\mathbb{R}^n}\Big(\frac{x^2}{2}-W_n^*\Big)g_n\,d\gamma_n$$
$$\le e^{C\int_{\mathbb{R}^n}\left|\frac{x^2}{2}-W_n^*\right|d\gamma_n}.$$

Applying the classical Poincaré inequality for Gaussian measures, Theorem 2.2 and taking into account that

$$\int_{\mathbb{R}^n} \Bigl(\frac{x^2}{2} - W_n^*\Bigr)\, d\gamma_n = 0,$$

we obtain

$$\mu_n(X) \le \exp\sqrt{C^2 \int_{\mathbb{R}^n} (x - \nabla W_n^*)^2\, d\gamma_n} \le \exp\sqrt{\frac{C^2}{c}\int_{\mathbb{R}^n} 2g_n \log g_n\, d\gamma_n} \le e^{C\sqrt{\frac{2C}{c}\log C}}.$$

Let us set $A := \sup_n \mu_n(X)$.

Let us establish the uniform tightness of $\{\mu_n\}$. We apply Theorem 2.2 to the case $f = 1$, $g = g_n$, $\mu = \gamma_n$ and $\varphi_n = e^{\frac{\nabla W_n{}^2}{2} - W_n^*(\nabla W_n)}$. Taking into account that $\nabla W_n(\nabla W_n^*) = x$ and

$$\frac{\nabla \varphi_n(\nabla W_n^*)}{\varphi_n(\nabla W_n^*)} = (D^2 W_n^*)^{-1}\Bigl(x - \nabla W_n^*\Bigr)$$

we obtain

$$\Bigl(\frac{1}{\sqrt{2\pi}}\Bigr)^n \int_{\mathbb{R}^n} g_n \log g_n e^{-W_n^*}\, dx \ge \Bigl(\frac{1}{\sqrt{2\pi}}\Bigr)^n \frac{1}{2}\int_{\mathbb{R}^n} (x - \nabla W_n^*)^2 g_n e^{-W_n^*}\, dx$$
$$+ \Bigl(\frac{1}{\sqrt{2\pi}}\Bigr)^n \int_{\mathbb{R}^n} \langle (D^2W_n^*)^{-1}\bigl[x - \nabla W_n^*\bigr], x - \nabla W_n^* \rangle g_n e^{-W_n^*}\, dx.$$

Hence

$$\frac{AC}{c}\log C \ge \frac{1}{2}\int_{\mathbb{R}^n} (x - \nabla W_n^*)^2\, d\mu_n + \int_{\mathbb{R}^n} \langle (D^2W_n^*)^{-1}\bigl[x - \nabla W_n^*\bigr], x - \nabla W_n^* \rangle\, d\mu_n. \quad (5.17)$$

Then for every fixed $i \le n$ we have

$$\int_{\mathbb{R}^n} \Bigl[x_i^2 + \Bigl(\frac{\partial W_n^*}{\partial x_i}\Bigr)^2\Bigr]\, d\mu_n = \int_{\mathbb{R}^n} \Bigl(x_i - \frac{\partial W_n^*}{\partial x_i}\Bigr)^2\, d\mu_n$$
$$+ 2\int_{\mathbb{R}^n} x_i \frac{\partial W_n^*}{\partial x_i}\, d\mu_n = \int_{\mathbb{R}^n} \Bigl(x_i - \frac{\partial W_n^*}{\partial x_i}\Bigr)^2\, d\mu_n + 2\mu_n(X).$$

Hence

$$\sum_{i=1}^k \int_{\mathbb{R}^n} x_i^2\, d\mu_n < 2Ak + \int_{\mathbb{R}^n} (x - \nabla W_n^*)^2\, d\mu_n \le 2A\Bigl(k + \frac{C\log C}{c}\Bigr), \quad k \le n.$$

By the Chebyshev inequality

$$\mu_n\Bigl(\sum_{i=1}^k x_i^2 > N\Bigr) \le \frac{2A\Bigl(k + \frac{C\log C}{c}\Bigr)}{N}.$$

This implies that the measurable linear mapping $X \ni x \to (\xi_i(x))_i \in \mathbb{R}^\infty$ pushes forward the sequence of measures $\{\mu_n\}$ to a uniformly tight sequence of measures on $\mathbb{R}^\infty$. Moreover, let us fix some $h \in H$. Then $h = \sum_{i=1}^\infty h_i e_i = v_n + w$ where $v_n = \sum_{i=1}^n h_i e_i$. Hence

$$\int_X \hat{h}^2(x)\, d\mu_n = \int_X \hat{v}_n^2(x)\, d\mu_n + 2\int_X \hat{v}_n(x)\hat{w}(x)\, d\mu_n + \int_X \hat{w}^2(x)\, d\mu_n = \int_X \hat{v}_n^2(x)\, d\mu_n$$
$$+ \int_X \hat{w}^2(x)\, d\mu_n$$

In the same way as above one can easily show that

$$\int_X \hat{h}^2(x)\,d\mu_n \le 2\|v_n\|_H^2 A\Big(1+\frac{C\log C}{c}\Big) + \|w\|_H^2. \tag{5.18}$$

Hence, if $h \in H$, then $\sup_n \|\hat{h}\|_{L^2(\mu_n)} < \infty$.

Now let us prove that the sequence of the Radon–Nikodym densities

$$\varrho_n = \varrho(x_1,\cdots,x_n) := \frac{d\mu_n}{d\gamma_n} = e^{\frac{1}{2}\sum_{i=1}^n x_i^2 - W_n^*}$$

has an $L^1(\gamma)$-convergent subsequence. Indeed, we obtain by the log-Sobolev inequality for Gaussian measures that

$$\int_{\mathbb{R}^n} \varrho_n \log \varrho_n\,d\gamma_n \le \int_{\mathbb{R}^n} \varrho_n\,d\gamma_n\Big(\log \int_{\mathbb{R}^n} \varrho_n\,d\gamma_n\Big) + \frac{1}{2}\int_{\mathbb{R}^n} \frac{\big\langle \nabla \varrho_n, \nabla \varrho_n \big\rangle_H}{\varrho_n}\,d\gamma_n \le A\log A$$
$$+\frac{1}{2}\int_{\mathbb{R}^n} (x-\nabla W_n^*)^2\,d\mu_n \le A\log A + \frac{AC\log C}{c}.$$

This yields the γ-uniform integrability of $\{\varrho_n\}$. Now let us apply the Poincaré inequality for γ to the function $\log \varrho_n - \log \varrho_m$, where $m \ge n$. Taking into account that

$$\int_X \big(\log \varrho_n - \log \varrho_m\big)\,d\gamma = \int_X \Big(W_n^* - W_m^* + \sum_{i=n+1}^m x_i^2\Big)\,d\gamma = 0,$$

we get

$$\int_X \big(\log \varrho_n - \log \varrho_m\big)^2\,d\gamma = \mathrm{Var}_\gamma\Big(W_n^* - W_m^* + \sum_{i=n+1}^m x_i^2\Big) \le \int_X (T_m - T_n)_H^2\,d\gamma.$$

We know that $\{T_m - x\}$ is fundamental in $\mathcal{L}$, hence one can extract a γ-a.e. convergent subsequence from $\{W_n^* - \frac{1}{2}\sum_{i=1}^n x_i^2\}$. Since $\{\varrho_n\}$ is γ-uniformly integrable, we obtain that some subsequence of $\{\varrho_n\}$ (denoted again by $\{\varrho_n\}$) converges in $L_1(\gamma)$ to some function ϱ, hence $A \ge \mu(X) \ge 1$, where $\mu := \varrho\cdot\gamma$.

Let us prove that μ satisfies the following integration by parts formula:

$$\int_X \frac{\partial \varphi}{\partial h}\,d\mu = \int_X \varphi \hat{h}(T)\,d\mu$$

for every $\varphi \in \mathcal{FC}_0^\infty$, $h \in H, \|h\| = 1$. From $L_1(\gamma)$-convergence $\varrho_n \to \varrho$ and estimate (5.18) we obtain $\int_X \varphi\hat{h}(x)\,d\mu_n \to \int_X \varphi\hat{h}(x)\,d\mu$. Hence by the integration-by-parts formula it suffices to prove that

$$\lim_n \int_X \varphi \hat{h}(T_n - x)\,d\mu_n = \lim_n \int_X \varphi\big\langle T_n - x, h\big\rangle_H\,d\mu_n = \int_X \varphi\big\langle T - x, h\big\rangle_H\,d\mu.$$

Since

$$\varphi\big\langle T_n - x, h\big\rangle_H \frac{\partial \mu_n}{\partial \gamma} \to \varphi\big\langle T - x, h\big\rangle_H \frac{\partial \mu}{\partial \gamma} \quad \gamma\text{-a.e.},$$

it is enough to show that $f_n = |\big\langle T_n - x, h\big\rangle_H| \frac{\partial \mu_n}{\partial \gamma_n}$ is γ-uniformly integrable, hence converges in $L^1(\,d\gamma)$. To this end we show that

$$f_n|\log f_n| = \Big|\log\big|\big\langle T_n - x, h\big\rangle_H\big| - W_n^* + \frac{1}{2}\sum_{i=1}^n x_i^2\Big| \cdot \big|\big\langle T_n - x, h\big\rangle_H\big| \cdot e^{-W_n^* + \frac{1}{2}\sum_{i=1}^n x_i^2}$$

is an $L^1(\gamma)$-bounded sequence. By the Cauchy inequality it suffices to prove that the sequences $\int_{\mathbb{R}^n} (W_n^* - \frac{x^2}{2})^2\,d\mu_n$, $\int_{\mathbb{R}^n} (\nabla W_n^* - x)^2\,d\mu_n$ are bounded.

Indeed, boundedness of $\int_{\mathbb{R}^n}(\nabla W_n^* - x)^2\, d\mu_n$ follows by (5.17). By the Brascamp–Lieb inequality (3.10) we have

$$\int_{\mathbb{R}^n}\Big(\frac{1}{2}\sum_{i=1}^n x_i^2 - W_n^*\Big)^2 d\mu_n \le \frac{\Big(\int_{\mathbb{R}^n}\Big[\frac{1}{2}\sum_{i=1}^n x_i^2 - W_n^*\Big]\, d\mu_n\Big)^2}{\mu_n(X)} + \int_{\mathbb{R}^n}\big\langle (D^2W_n^*)^{-1}\big[x-\nabla W_n^*\big], x-\nabla W_n^*\big\rangle\, d\mu_n.$$

Applying (5.17) and the above estimate for the entropy

$$\int_{\mathbb{R}^n}\varrho_n\log\varrho_n\, d\gamma = \int_{\mathbb{R}^n}\Big[\frac{1}{2}\sum_{i=1}^n x_i^2 - W_n^*\Big]\, d\mu_n,$$

we see that the sequence $\int_{\mathbb{R}^n}\Big(\frac{1}{2}\sum_{i=1}^n x_i^2 - W_n^*\Big)^2 d\mu_n$ is bounded. The desired probability measure is $\nu := \frac{\mu}{\mu(X)}$. Clearly, ν is convex as the limit of convex measures. The proof is complete. □

6. Estimates of the second derivatives

We denote by $\mathcal{H}^\gamma$ the space of measurable mappings $A : X \to \mathcal{H}$, where $\mathcal{H}$ is the space of all Hilbert–Schmidt operators on H, such that

$$\int_X \|A(x)\|^2_{\mathcal{H}}\, d\gamma < \infty,$$

where $\|\cdot\|_{\mathcal{H}}$ is the Hilbert-Schmidt norm. We endow $\mathcal{H}^\gamma$ with its natural inner product

$$\big\langle A, B\big\rangle_{H^\gamma} = \int_X \sum_{i=1}^\infty \big\langle A(x)e_i, B(x)e_i\big\rangle_H\, d\gamma.$$

We denote in the following $\mathrm{Tr}A = \sum_{i=1}^\infty \big\langle Ae_i, Ae_i\big\rangle_H$ for every non-negative operator A on H.

We prove in this section that the measure ν constructed in Section 5 is non-degenerate in some sense. More precisely, the weak derivative of T can be estimated from below (in the sense of distributions) by some mapping DT that has the form $DT = (E+K)^2$, where $K \in \mathcal{H}^\gamma$. Moreover, DT admits an inverse operator $(E+L)^2$ for γ-almost all $x \in X$ and $L \in \mathcal{H}^\gamma$. DT can be obtained as the limit of the derivatives of the finite-dimensional approximations T_n. Unfortunately, it is not clear whether DT coincides with the weak derivative of T, i.e., whether inequality (6.19) can be replaced by the equality. However, we show that this is the case under the additional assumption that the second derivatives of $-\log g$ are uniformly bounded from below.

Some results on the second derivatives of the infinite-dimensional transport are obtained in [20]. The following problem remains open: when does the infinite-dimensional analog of the Monge-Ampère equation (1.4) hold? It was shown in [20] that under broad assumptions the desired change-of-variables formula holds in the form of inequality.

Theorem 6.1. *Let g satisfy the assumptions of Corollary 4.1. Suppose, in addition, $0 < c < g$. Then there exists a measurable mapping DT with values in the space of symmetric nonnegative operators on H such that $K = \sqrt{DT} - E \in \mathcal{H}^\gamma$, $DT(x)$ has the inverse operator $DT^{-1}(x)$ for γ-almost every x and the following inequality holds:*

$$\int_X \Big\langle T - x, h\Big\rangle_H \beta_h(x)\, d\gamma \le -\int_X \Big\langle (DT-E)h, h\Big\rangle_H\, d\gamma, \qquad (6.19)$$

where $h \in H$ and DT^{-1} has the form $(E+L)^2$ with $L \in \mathcal{H}^\gamma$.

Proof. Let us consider the finite-dimensional approximations $T_n \to T$ constructed in Section 4. By Theorem 2.2 we have that

$$\begin{aligned}
&\int_X \log \frac{g_m}{g_n} g_m \, d\gamma \\
&= \frac{1}{2}\int_X (T_n - T_m)_H^2 g_m \, d\gamma + \int_X \bigl[\mathrm{Tr}(DT_n DT_m^{-1} - E) - (\log\det DT_n DT_m^{-1})\bigr] g_m \, d\gamma \\
&\geq c \int_X \bigl[\mathrm{Tr}(DT_n DT_m^{-1} - E) - (\log\det DT_n DT_m^{-1})\bigr] \, d\gamma.
\end{aligned}$$

Analogously

$$\int_X \log \frac{g_n}{g_m} g_n \, d\gamma \geq c \int_X \bigl[\mathrm{Tr}(DT_m DT_n^{-1} - E) - (\log\det DT_m DT_n^{-1})\bigr] \, d\gamma.$$

Summing these inequalities we obtain

$$\begin{aligned}
&\int_X \log \frac{g_n}{g_m}(g_n - g_m) \, d\gamma \geq c \int_X \mathrm{Tr}(DT_n DT_m^{-1} + DT_m DT_n^{-1} - 2E) \, d\gamma = \\
&c \int_X \mathrm{Tr}(DT_m^{-\frac{1}{2}} DT_n DT_m^{-\frac{1}{2}} + DT_m^{\frac{1}{2}} DT_n^{-1} DT_m^{\frac{1}{2}} - 2E) \, d\gamma.
\end{aligned}$$

Applying the inequality $A + A^{-1} - 2E \geq (\sqrt{A} - E)^2 + (\sqrt{A^{-1}} - E)^2$, which holds for every positive symmetric matrix A, we see that the right hand side is bigger than

$$c \int_X \mathrm{Tr}\Bigl[\Bigl(\sqrt{DT_m^{-\frac{1}{2}} DT_n DT_m^{-\frac{1}{2}}} - E\Bigr)^2 + \Bigl(\sqrt{DT_m^{\frac{1}{2}} DT_n^{-1} DT_m^{\frac{1}{2}}} - E\Bigr)^2\Bigr] \, d\gamma,$$

which equals

$$c\Bigl\|\sqrt{DT_m^{-\frac{1}{2}} DT_n DT_m^{-\frac{1}{2}}} - E\Bigr\|_{\mathcal{H}^\gamma}^2 + c\Bigl\|\sqrt{DT_m^{\frac{1}{2}} DT_n^{-1} DT_m^{\frac{1}{2}}} - E\Bigr\|_{\mathcal{H}^\gamma}^2.$$

Taking $g_m = 1$ we see that the norms $\|\sqrt{DT_n} - E\|_{\mathcal{H}^\gamma}^2$, $\|\sqrt{DT_n^{-1}} - E\|_{\mathcal{H}^\gamma}^2$ are uniformly bounded. Hence one can extract an $\mathcal{H}^\gamma$-weakly convergent subsequence (denoted again by T_n) such that $\sqrt{DT_n} - E = K_n \to K$, $\sqrt{DT_n^{-1}} - E = L_n \to L$. By the Banach–Saks theorem, there exists a subsequence (denoted again by T_n) such that $\frac{1}{n}\sum_{i=1}^n \sqrt{DT_i} - E \to K$ and $\frac{1}{n}\sum_{i=1}^n \sqrt{DT_i^{-1}} - E \to L$ in the norm of the space $\mathcal{H}^\gamma$. Passing to a subsequence in this sequence of arithmetic means we have convergence in the Hilbert–Schmidt norm for γ-almost every x. To simplify notation we shall assume that this holds for the whole sequence of arithmetic means. Denote $\frac{1}{n}\sum_{i=1}^n \sqrt{DT_i}$ by $\overline{DT_n}^{\frac{1}{2}}$ and $\frac{1}{n}\sum_{i=1}^n \sqrt{DT_i^{-1}}$ by

$\overline{DT}_n^{-\frac{1}{2}}$. The above obtained inequality implies that for every $k \in \mathbb{N}$

$$\frac{1}{m}\sum_{i=1}^{m}\int_X \log\frac{g_n}{g_i}(g_n - g_i)\,d\gamma$$
$$\geq c\frac{1}{m}\sum_{i=1}^{m}\int_X \mathrm{Tr}(DT_i^{-\frac{1}{2}}DT_nDT_i^{-\frac{1}{2}} + DT_i^{\frac{1}{2}}DT_n^{-1}DT_i^{\frac{1}{2}} - 2E)\,d\gamma$$
$$\geq c\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{k}\int_X \big\langle DT_nDT_i^{-\frac{1}{2}}e_j, DT_i^{-\frac{1}{2}}e_j\big\rangle_H\,d\gamma$$
$$+ c\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{k}\int_X \big\langle DT_n^{-1}DT_i^{\frac{1}{2}}e_j, DT_i^{\frac{1}{2}}e_j\big\rangle_H\,d\gamma - 2ck.$$

Since $\int_X \log\frac{g_n}{g_m}(g_n - g_m)\,d\gamma \to \int_X \log\frac{g_n}{g}(g_n - g)\,d\gamma$, the averages

$$\frac{1}{m}\sum_{i=1}^{m}\int_X \log\frac{g_n}{g_i}(g_n - g_i)\,d\gamma$$

have the same limit. Then, by using the convexity properties, we obtain

$$\int_X \log\frac{g_n}{g}(g_n - g)\,d\gamma \geq \overline{\lim}_m \sum_{j=1}^{k}\Big[c\int_X \big\langle DT_n\overline{DT}_m^{-\frac{1}{2}}e_j, \overline{DT}_m^{-\frac{1}{2}}e_j\big\rangle_H\,d\gamma$$
$$+ c\int_X \big\langle DT_n^{-1}\overline{DT}_m^{\frac{1}{2}}e_j, \overline{DT}_m^{\frac{1}{2}}e_j\big\rangle_H\,d\gamma\Big] - 2ck.$$

Clearly, $\overline{DT}_n^{\frac{1}{2}} \to E + K, \overline{DT}_n^{-\frac{1}{2}} \to E + L$ in the operator norm for γ-almost all x. Then by Fatou's lemma

$$\int_X \log\frac{g_n}{g}(g_n - g)\,d\gamma \geq \sum_{j=1}^{k}\Big[c\int_X \big\langle DT_n(E+L)e_j, (E+L)e_j\big\rangle_H\,d\gamma$$
$$+ c\int_X \big\langle DT_n^{-1}(E+K)e_j, (E+K)e_j\big\rangle_H\,d\gamma\Big] - 2ck.$$

Applying the inequality $\frac{1}{n}\sum_{i=1}^{n} DT_i \geq (\frac{1}{n}\sum_{i=1}^{n}\sqrt{DT_i})^2$ one obtains as above that

$$0 \geq c\cdot\underline{\lim}_n \sum_{j=1}^{k}\Big[\int_X \big\langle(\overline{DT}_n^{\frac{1}{2}})^2(E+L)e_j, (E+L)e_j\big\rangle_H\,d\gamma$$
$$+ \int_X \big\langle(\overline{DT}_n^{-\frac{1}{2}})^2(E+K)e_j, (E+K)e_j\big\rangle_H\,d\gamma\Big] - 2ck$$
$$\geq c\int_X \sum_{j=1}^{k}\Big[\big\langle(E+K)^2(E+L)e_j, (E+L)e_j\big\rangle_H +$$
$$\big\langle(E+L)^2(E+K)e_j, (E+K)e_j\big\rangle_H - 2\big\langle e_j, e_j\big\rangle_H\Big]\,d\gamma.$$

Since the integrand on the right hand side of this formula is nonnegative γ-a.e. as the γ-a.e. limit of nonnegative functions, it vanishes. Similarly, one proves that

$$\big\langle(E+L)(E+K)^2(E+L)v, v\big\rangle_H + \big\langle(E+K)(E+L)^2(E+K)v, v\big\rangle_H = 2$$

for every measurable field $v : X \to H$ such that $\|v(x)\|_H = 1$ for almost every x. Therefore, $(E+L)(E+K)^2(E+L) + (E+K)(E+L)^2(E+K) = 2E$ γ-a.e. By Lemma 7.1 from Appendix we have $(E+K)(E+L) = (E+L)(E+K) = E$ γ-a.e. The inequalities obtained above and the trace property $\mathrm{Tr}AB = \mathrm{Tr}BA$ yield

$$\int_X \log \frac{g_n}{g}(g_n - g)\,d\gamma \geq \sup_k \sum_{j=1}^{k} \Bigl[c \int_X \bigl\langle DT_n(E+L)e_j, (E+L)e_j \bigr\rangle_H \,d\gamma$$
$$+ c \int_X \bigl\langle DT_n^{-1}(E+K)e_j, (E+K)e_j \bigr\rangle_H \,d\gamma \Bigr] - 2ck$$
$$\geq c \int_X \mathrm{Tr}\bigl[(E+L)DT_n(E+L) + (E+K)DT_n^{-1}(E+K) - 2E\bigr]\,d\gamma$$
$$= c \int_X \mathrm{Tr}\bigl[(E+K_n)(E+L)^2(E+K_n) + (E+L_n)(E+K)^2(E+L_n) - 2E\bigr]\,d\gamma.$$

Since $A + A^{-1} - 2E \geq (\sqrt{A} - E)^2 + (\sqrt{A^{-1}} - E)^2$ for every positive symmetric operator A, it follows that

$$\int_X \log \frac{g_n}{g}(g_n - g)\,d\gamma \geq$$
$$c \int_X \|\sqrt{(E+L)DT_n(E+L)} - E\|^2_{\mathcal{H}}\,d\gamma + c \int_X \|\sqrt{(E+K)DT_n^{-1}(E+K)} - E\|^2_{\mathcal{H}}\,d\gamma.$$

and

$$\int_X \log \frac{g_n}{g}(g_n - g)\,d\gamma \geq c \int_X \|\sqrt{(E+K_n)(E+L)^2(E+K_n)} - E\|^2_{\mathcal{H}}\,d\gamma$$
$$+ c \int_X \|\sqrt{(E+L_n)(E+K)^2(E+L_n)} - E\|^2_{\mathcal{H}}\,d\gamma.$$

Letting $n \to \infty$ and passing to a subsequence we obtain that the operators

$$\sqrt{(E+K)DT_n^{-1}(E+K)} - E \quad \text{and} \quad \sqrt{(E+L)DT_n(E+L)} - E$$

tend to the zero operator in the Hilbert–Schmidt norm γ-a.e., hence $K_n \to K$ and $L_n \to L$ in the operator norm γ-a.e.

Now let us prove that

$$\lim_n \int_X \bigl\langle (K_n - K)v_1, v_2 \bigr\rangle_H \,d\gamma = 0 \tag{6.20}$$

for every $v_1, v_2 \in H$. Indeed, one can easily construct a symmetric operator B such that $B(\mathrm{span}\{v_1, v_2\}) = \mathrm{span}\{v_1, v_2\}$, $B|_{\mathrm{span}\{v_1,v_2\}^\perp} = 0$ and $\bigl\langle A, B \bigr\rangle_{\mathcal{H}} = \bigl\langle Av_1, v_2 \bigr\rangle_H$ for every Hilbert–Schmidt operator A. We immediately obtain the claim from $\mathcal{H}$-weak convergence $K_n \to K$.

Now let us prove (6.19). By Corollary 4.1 $T_n \to T$ in $\mathcal{L}$. Then, by (6.20), Fatou's lemma and the fact that $K_n h \to Kh$ γ-a.e., we obtain

$$\int_X \bigl\langle T - x, h \bigr\rangle_H \beta_h(x)\,d\gamma = \lim_n \int_X \bigl\langle T_n - x, h \bigr\rangle_H \beta_h(x)\,d\gamma$$
$$= -\lim_n \int_X \frac{\partial}{\partial h} \bigl\langle T_n - x, h \bigr\rangle_H \,d\gamma = -\lim_n \int_X \bigl\langle (DT_n - E)h, h \bigr\rangle_H \,d\gamma$$
$$= -\lim_n \int_X \bigl\langle (2K_n + K_n^2)h, h \bigr\rangle_H \,d\gamma \leq -\int_X \bigl\langle ((E+K)^2 - E)h, h \bigr\rangle_H \,d\gamma.$$

The proof is complete. □

Lemma 6.2. *Let $\nabla\Phi\colon \mathbb{R}^d \to \mathbb{R}^d$ be the optimal transfer plan sending a probability measure $e^{\psi}\cdot\gamma_d$ to γ_d, where ψ is smooth. Suppose that $\frac{\partial^2\psi}{\partial h^2} \geq -M$ almost everywhere for some $h\in\mathbb{R}^d$, $\|h\|=1$ and $M\geq 0$. Then $\left(\frac{\partial^2\Phi}{\partial h^2}\right)^2 \leq 4(1+M)$.*

Proof. We follow an idea of Caffarelli [14]. By approximation we can replace the measure γ_d by $\frac{\chi(B_R)\cdot\gamma_d}{\gamma_d(B_R)}$, where $R>0$ and $B_R=\{x : \|x\|\leq R\}$. Let us fix some $\varepsilon>0$ and show that the incremental quotient

$$\delta\Phi_\varepsilon = \Phi(x+\varepsilon h) + \Phi(x-\varepsilon h) - 2\Phi(x)$$

satisfies the maximum principle. It can be shown exactly in the same way as in [14] (see Lemma 4) that $\delta\Phi_\varepsilon \to 0$ as $|x|\to\infty$.

Let x_0 be the point where $\delta\Phi_\varepsilon$ attains its maximum. Then

$$\nabla\Phi(x_0+\varepsilon h) + \nabla\Phi(x_0-\varepsilon h) = 2\nabla\Phi(x_0) \tag{6.21}$$

and

$$D^2\Phi(x_0+\varepsilon h) + D^2\Phi(x_0-\varepsilon h) \leq 2D^2\Phi(x_0).$$

By the concavity of $\log\det$ we obtain

$$\log\det D^2\Phi(x_0) \geq \log\det\Bigl(\frac{D^2\Phi(x_0+\varepsilon h) + D^2\Phi(x_0-\varepsilon h)}{2}\Bigr)$$
$$\geq \frac{1}{2}\Bigl(\log\det D^2\Phi(x_0+\varepsilon h) + \log\det D^2\Phi(x_0-\varepsilon h)\Bigr).$$

Hence, by the change-of-variables formula

$$2\Bigl(\psi(x_0) + \frac{\bigl(\nabla\Phi(x_0)\bigr)^2}{2} - \frac{x_0^2}{2}\Bigr) \geq \psi(x_0+\varepsilon h) + \frac{\bigl(\nabla\Phi(x_0+\varepsilon h)\bigr)^2}{2} - \frac{(x_0+\varepsilon h)^2}{2}$$
$$+\psi(x_0-\varepsilon h) + \frac{\bigl(\nabla\Phi(x_0-\varepsilon h)\bigr)^2}{2} - \frac{(x_0-\varepsilon h)^2}{2}$$

and we obtain from (6.21) that

$$\varepsilon^2 \geq \delta_{\varepsilon h}\psi(x_0) + \frac{1}{4}\Bigl(\nabla\Phi(x_0+\varepsilon h) - \nabla\Phi(x_0-\varepsilon h)\Bigr)^2.$$

By the convexity of Φ

$$\Phi(x_0) \geq \Phi(x_0-\varepsilon h) + \bigl\langle\nabla\Phi(x_0-\varepsilon h), \varepsilon h\bigr\rangle, \quad \Phi(x_0) \geq \Phi(x_0+\varepsilon h) - \bigl\langle\nabla\Phi(x_0+\varepsilon h), \varepsilon h\bigr\rangle,$$

hence

$$\bigl\langle\nabla\Phi(x_0+\varepsilon h) - \nabla\Phi(x_0-\varepsilon h), \varepsilon h\bigr\rangle \geq \delta_{\varepsilon h}\Phi(x_0).$$

Consequently

$$\bigl\langle\nabla\Phi(x_0+\varepsilon h) - \nabla\Phi(x_0-\varepsilon h)\bigr\rangle^2\varepsilon^2 \geq \bigl(\delta_{\varepsilon h}\Phi(x_0)\bigr)^2$$

and

$$\varepsilon^2 \geq \delta_{\varepsilon h}\psi(x_0) + \Bigl(\frac{\delta_{\varepsilon h}\Phi(x_0)}{2\varepsilon}\Bigr)^2.$$

Hence for every x one has $(\delta_{\varepsilon h}\Phi(x))^2 \leq (\delta_{\varepsilon h}\Phi(x_0))^2 \leq 4(1+M)\varepsilon^4$. Therefore, $\left(\frac{\partial^2\Phi}{\partial h^2}\right)^2 \leq 4(1+M)$. □

Remark 6.3. *In fact a stronger estimate holds: under the assumptions of Lemma 6.2 one has* $\Bigl(\frac{\partial^2\Phi}{\partial h^2}\Bigr)^2 \le 1+M$. *This can be readily shown by a heuristic application of the maximum principle. If* $\frac{\partial^2\Phi}{\partial h^2}(x)$ *attains its maximum at some point* x_0, *we obtain by differentiating the identity*

$$\psi - \frac{x^2}{2} = \log\det D^2\Phi - \frac{(\nabla\Phi)^2}{2}$$

along h *that*

$$\psi_h - (x,h) = \frac{(\det D^2\Phi)_h}{\det D^2\Phi} - \Bigl\langle D^2\Phi h, \nabla\Phi\Bigr\rangle$$

$$\psi_{hh} - 1 = \frac{(\det D^2\Phi)_{hh}}{\det D^2\Phi} - \frac{(\det D^2\Phi)^2_h}{(\det D^2\Phi)^2} - \Bigl\langle \nabla\frac{\partial^2\Phi}{\partial h^2}, \nabla\Phi\Bigr\rangle - \Bigl(\frac{\partial^2\Phi}{\partial h^2}\Bigr)^2.$$

Since x_0 *is a point of maximum, we obtain* $\nabla\frac{\partial^2\Phi}{\partial h^2}(x_0) = 0$ *and* $(\det D^2\Phi)_{hh} \le 0$. *Hence* $\Bigl(\frac{\partial^2\Phi}{\partial h^2}\Bigr)^2 \le 1 - \psi_{hh} \le 1+M$.

Theorem 6.4. *Suppose that* g *satisfies the conditions of Theorem 6.1 and, in addition,* $-\frac{\partial^2\log g}{\partial h^2} \le M$. *Then* T *has a weak derivative along* h, *i.e.*

$$\int_X \Bigl\langle T - x, h\Bigr\rangle_H \beta_h(x)\, d\gamma = -\int_X \Bigl\langle (DT-E)h, h\Bigr\rangle_H d\gamma.$$

Moreover, D_hT *is uniformly bounded.*

Proof. It is easy to show that the finite dimensional approximations g_n can be chosen in such a way that $-\frac{\partial^2\log g_n}{\partial h^2} \le M$. Indeed, it follows from Prékopa's theorem (see [26]) that the conditional expectation $\mathbb{E}(g|\mathcal{F}_n)$ preserves this property. By Lemma 6.2 one has $D_hT_n\Bigr)^2 \le 4(1+M)$. By Theorem 6.1 we obtain that a subsequence of $\{D_hT_n(x)\}$ tends to $(E+K)^2(x)h$ for γ-almost every x. Applying Lebesgue's domination convergence theorem and integration by parts we readily obtain our claim. □

7. Appendix

Lemma 7.1. *Let* K, L *be compact symmetric nonnegative operators on a separable Hilbert space* H. *Suppose that* $A = E+K$ *and* $B = E+L$ *satisfy the following relation:*

$$AB^2A + BA^2B = 2E. \tag{7.22}$$

Then $AB = BA = E$.

Proof. It is well-known that AB can be written as $AB = UC$, where $C = |AB|$ is symmetric and nonnegative, U is a partially isometric operator, i.e., $U : \mathrm{Ker}U^{\perp} \to \mathrm{Ran}U$ is an isometric mapping and $\mathrm{Ker}U = \mathrm{Ker}C$. Clearly, $C - E$ is compact. Hence $\mathrm{Ker}C$ is finite dimensional and there exists an orthonormal basis $\{e_i\}$ in H such that $Ce_i = \lambda_i e_i$. For every λ_i we denote by H_{λ_i} the corresponding space of eigenvectors of C with the eigenvalue λ_i. Clearly, H_{λ_i} is finite dimensional if $\lambda_i \ne 1$. We obtain from (7.22) that $UC^2U^* + C^2 = 2E$. Hence

$$\Bigl\langle CU^*e_i, CU^*e_j\Bigr\rangle = \Bigl\langle UC^2U^*e_i, e_j\Bigr\rangle = 2\Bigl\langle e_i, e_j\Bigr\rangle - \Bigl\langle Ce_i, Ce_j\Bigr\rangle = 0$$

if $i \ne j$ and $\Bigl\langle CU^*e_i, CU^*e_i\Bigr\rangle = 2 - \lambda_i^2$. This implies that $U^*(H_{\lambda_i}) \subset H_{\mu_i}$, where $\mu_i = \sqrt{2-\lambda_i^2}$. Hence UC is an isometry on H_1 and for every λ_i the spaces $H_{\lambda_i} \oplus H_{\mu_i}$ and $(H_{\lambda_i} \oplus H_{\mu_i})^{\perp}$ are invariant with respect to C and U^*, hence with respect to UC. One can easily prove that there exists an invertible symmetric operator $\tilde{B}$ such that $B\tilde{B}$ is the orthogonal projection P on $\mathrm{Ker}B^{\perp}$. Then $AP = UC\tilde{B}$. Suppose for simplicity that the first m vectors $\{e_1, \cdots, e_m\}$ of the basis constitute a basis in H_0 and the vectors $\{e_{m+1}, \cdots, e_{m+k}\}$ constitute a basis in H_2. Note that $\mathrm{Ker}B \subset \mathrm{Ker}C$. By the invariance of $H_0 \oplus H_2$ and $(H_0 \oplus H_2)^{\perp}$ with respect to UC, we see that the $m+k$-dimensional matrices $A_{m+k} = \langle Ae_i, e_j \rangle$, $\tilde{B}_{m+k} = \langle \tilde{B}e_i, e_j \rangle$, where $1 \le i, j \le m+k$, satisfy the following relation: $A_{m+k}P_{m,k} = 2U_{m,k}\tilde{B}_{m+k}$, where $P_{m,k}$ is the projection of $\mathbb{R}^{m+k}$ on $\mathrm{Ker}B^{\perp} \supset \mathrm{span}\{e_{m+1}, \cdots, e_{m+k}\}$ and $U_{m,k}$ is an operator of the following type:

$$U_{m,k}e_i = 0 \quad \text{if} \quad 1 \le i \le m,$$

and $U_{m,k}$ is an isometric embedding of $\mathrm{span}\{e_{m+1}, \cdots, e_{m+k}\}$ into $\mathrm{span}\{e_1, \cdots, e_m\}$. This implies that the $k \times k$ matrix

$$\langle Ae_i, e_j \rangle, \quad m+1 \le i, j \le m+k,$$

is zero. But A_{m+k} and $\tilde{B}_{m+k}$ are nonnegative, hence

$$\langle A_{m+k}e_i, e_j \rangle = 0, \quad \text{for} \quad 1 \le i \le m, \quad m+1 \le j \le m+k.$$

Since $U_{m,k}$ is an isometry on $\mathrm{span}\{e_{m+1}, \cdots, e_{m+k}\}$, we obtain that

$$\langle \tilde{B}e_i, e_j \rangle = 0, \quad m+1 \le i, j \le m+k.$$

Since $\tilde{B}$ is nonnegative, this contradicts the fact that $\tilde{B}$ admits an inverse operator. Hence $H_0 = H_2 = \emptyset$, $\mathrm{Ker}A = \mathrm{Ker}B = \mathrm{Ker}C = \mathrm{Ker}U = \emptyset$, by the Fredholm alternative, A, B, C are invertible and U is a unitary operator.

Now we prove that $H_{\lambda_i} = \emptyset$ for every $\lambda_i \ne 1$. Indeed, we have $A = UCB^{-1}$. Since A, B^{-1} are invertible, the matrices $\langle Ae_i, e_j \rangle$, $\langle B^{-1}e_i, e_j \rangle$ are non-degenerate for every finite set $K \subset \mathbb{N}$ of indices $i, j \in K$. The spaces $H_{\lambda_i} \oplus H_{\mu_i}$ and $(H_{\lambda_i} \oplus H_{\mu_i})^{\perp}$ are invariant with respect to UC. We obtain as above that $UC|_{H_{\lambda_i} \oplus H_{\mu_i}}$ is a composition of two strictly positive matrices E, F. Hence $\mathrm{Tr}UC|_{H_{\lambda_i} \oplus H_{\mu_i}} = \mathrm{Tr}EF = \mathrm{Tr}\sqrt{F}E\sqrt{F} > 0$. But $\mathrm{Tr}UC|_{H_{\lambda_i} \oplus H_{\mu_i}} = 0$, since $UC(H_{\lambda_i}) \subset H_{\mu_i}$ and $UC(H_{\mu_i}) \subset H_{\lambda_i}$. Thus we have $AB = U$, where U is a unitary operator. Clearly, $AB = E$. The proof is complete. □

Acknowledgment The author is grateful to Prof. Vladimir Bogachev for fruitful discussions and helpful comments.

References

[1] Alexandrov, A.D. (1942). Existence and uniqueness of the convex surface with a given integral curvature. *Dokl. Akad. Nauk USSR* **35**, 131–134.

[2] Blackwell, D. and Dubins, L.E. (1983). An extension of Skorohod's almost sure representation theorem. *Proc. Amer. Math. Soc.* **89**, no. 4, 691–692.

[3] Bobkov, S.G. Large deviations via transfer plans (to appear).

[4] Bobkov, S.G., Gentil, I. and Ledoux, M. (2001). Hypercontractivity of Hamilton-Jacobi equations. *J. Math. Pures Appl.* **80**, no. 7, 669–696.

[5] Bobkov, S.G. and Ledoux, M. (2000). From Brunn-Minkovsky to Brascamp-Lieb and to logarithmic Sobolev inequality. *Geom. and Funct. Anal.* **10**, 1028–1052.

[6] Bogachev, V.I. (1998). Gaussian Measures. Amer. Math. Soc., Rhode Island, 1998.

[7] Bogachev, V.I. and Kolesnikov, A.V. (2001). Open mappings of probability measures and the Skorokhod representation theorem. *Theory Probab. Appl.* **46**, no. 1, 20–38.

[8] Bogachev V.I., Kolesnikov A.V. and Medvedev K.V. (2003) On triangular transformations of measures. *Russian Math. Dokl.* (to appear).

[9] Bogachev V.I. and Röckner M. (1995) Regularity of invariant measures on finite- and infinite-dimensional spaces and applications. *J. Funct. Anal.* **133**, no. 1, 168–223.

[10] Borell, C. (1974) Convex measures on locally convex spaces. *Ark Mat.* **12**, no. 2, 239–252.

[11] Brenier, Y. (1991) Polar factorization and monotone rearrangement of vector valued functions. *Comm. Pure Appl. Math.* **44**, 375–417.

[12] Brascamp, H. and Lieb, E.H. (1976) On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Funct. Anal.* **22**, 366–389.

[13] Caffarelli, L.A. (1992) The regularity of mappings with a convex potential. *J. Amer. Math. Soc.* **5**, no. 1, 99–104.

[14] Caffarelli, L.A. (2000) Monotonicity properties of optimal transportation and the FKG and related inequalities. *Commun. Math. Phys.* **214**, no. 3, 547–563.

[15] Cordero-Erausquin D. (2002) Some applications of mass transport to Gaussian type inequalities. *Arch. Rat. Mech. Anal.* **161**, 257–269.

[16] Cordero-Erausquin D., Gangbo, W. and Houdre, C. (2001) Inequalities for generalized entropy and optimal transportation. D. *Procedings of the workshop: Mass transportation methods in kinetic theory and hydrodynamics* (to appear).

[17] Cuesta-Albertos, J.A. and Matrán-Bea, C. (1994) Stochastic convergence through Skorohod representation theorems and Wasserstein distances. In *First International Conference on Stochastic Geometry, Convex Bodies and Empirical Measures (Palermo, 1993). Rend. Circ. Mat. Palermo (2)* Suppl. No. **35**, 89–113.

[18] Fernique, X. (2003) Extension du théorème de Cameron-Martin aux translations aléatoires. *The Ann. of Probab.* **31**, no. 3, 1296–1304.

[19] Feyel, D. and Üstünel, A.S. (2000) The notions of convexity and concavity on Wiener space. *J. Funct. Anal.* **176**, 400–428.

[20] Feyel, D. and Üstünel, A.S. (2002) Transport of measures on Wiener space and the Girsanov theorem *C. R. A. S.* **334**, no. 1, 1025–1028.

[21] Gangbo, W. (1994) An elementary proof of the polar factorization of vector-valued functions. *Arch. Rational. Mech. Anal.* **128**, 381–399.

[22] Harge, G. Inequalities for Gaussian measure and an application to Wiener space (to appear).

[23] Kantorovich, L.V. (1942) On the translocation of masses. *C.R. (Doklady) Acad. Sci. Nauk URSS* **37**, no. 7-8, 199–201.

[24] Knote, H. (1957) Contributions to the theory of convex bodies. *Michigan Math. J.* **4**, 39–52.

[25] Kolesnikov, A.V. (2003) Convexity inequalities and nonlinear transformations of measures. *Russian Math. Dokl.* (to appear).

[26] Ledoux, M. (2001) The concentration of measure phenomenon. Amer. Math. Soc., Providence, Rhode Island.

[27] McCann, R.J. (1995) Existence and uniqueness of monotone meausure-preserving maps. *Duke Math. J.* **80**, 309–323.

[28] Otto, F. and Villani, C. (2000) Generalization of an inequality by Talagrand, and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173**, 361–400.

[29] Rachev, S.T. and Rüschendorf, L. (1998) Mass Transportation Problems. V I, II, Springer, New-York.

[30] Talagrand, M. (1996) Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.* **6**, 587–600.

[31] Tuero, A. (1993) On the stochastic converegence of representation based on Wasserstein metrics. *The Ann. of Probab.* **21**, no. 1, 72–85.

[32] Üstünel, A.S. and Zakai, M. (2000) Transformation of measure on Wiener space. Springer, Berlin.

[33] Villani, C. (2003) Topics in Optimal Transportation. Amer. Math. Soc. Providence, Rhode Island.