TRIANGULAR TRANSFORMATIONS OF MEASURES

V.I. Bogachev, A.V. Kolesnikov, K.V. Medvedev

We obtain a new identity for the entropy of a nonlinear image of a measure on $\mathbb{R}^n$, which yields the well-known inequality of Talagrand. We study triangular mappings on $\mathbb{R}^n$ and on $\mathbb{R}^\infty$, i.e., the mappings T such that the ith coordinate function T_i depends only on the variables $x_1, \ldots, x_i$. With the help of such mappings we give a positive solution to the well-known open problem on the representability of every probability measure ν that is absolutely continuous with respect to a Gaussian measure γ on an infinite dimensional space as the image of γ under a mapping of the form $T(x) = x + F(x)$, where F takes on values in the Cameron–Martin space of the measure γ. As an application we also prove a generalized logarithmic Sobolev inequality.

The bibliography 23 titles.

The AMS classification: 28C20, 46G12, 60B11

1. Introduction

This work is devoted to the study of the so called triangular mappings, i.e., mappings $T = (T_1, \ldots, T_n) \colon \mathbb{R}^n \to \mathbb{R}^n$ such that T_1 is a function of x_1, T_2 is a function of (x_1, x_2), T_3 is a function of (x_1, x_2, x_3) and so on: T_i is a function of $(x_1, x_2, \ldots, x_i)$. Analogously one defines triangular mappings on $\mathbb{R}^\infty$, the countable product of real lines. A triangular mapping is called increasing if every its component T_i is increasing with respect to the variable x_i. The same terminology is used for mappings defined on subsets of $\mathbb{R}^n$ or $\mathbb{R}^\infty$. Consideration of triangular mappings is very natural in the problems of probability theory connected with transformations of sequences of random variables. The choice of this term is explained by the fact that for a differentiable triangular transformation (for example, linear) the Jacobi matrix has a triangular form. In [15], triangular transformations of uniform distributions on convex sets were constructed (see also [2]), and in [21], there has been actually established the existence of a triangular mapping T of the standard Gaussian measure γ on $\mathbb{R}^n$ into an arbitrary absolutely continuous probability measure $\nu = f \cdot \gamma$. Talagrand's inequality established with the aid of such transformations estimates the L^2-norm of the difference $T - I$ via the integral of $f \log f$ against the measure γ, i.e., the entropy of the Radon–Nikodym density. We establish an exact equality for the entropy of a nonlinear image of a measure, which directly implies Talagrand's inequality. The obtained identity is used in our study of triangular mappings.

It is well known that every Radon probability measure on a metric space is the image of Lebesgue measure on an interval (or any other atomless probability measure) under certain Borel mapping. However, the problem often arises of transforming a given measure into another one by means of a mapping from more narrow classes. This subject, which has become very popular in the past decade, is connected with a whole series of classical problems in the theory of extremal problems, measure theory, nonlinear analysis, and the theory of nonlinear partial differential equations, in particular, with the well-known Monge–Kantorovich mass transportation problem. The interaction of all these directions has lead not only to exciting results on measure transformations, but also to discovery of interesting links between the diverse areas and to unexpected applications. In particular, certain new functional inequalities have been obtained. Among the diverse classes of mappings considered by many authors, one should distinguish the monotone type mappings (for example, gradients of convex functions) and the above defined triangular mappings.

Among recent bright results on monotone mappings we note the theorem from [8] and [18] (see a detailed discussion in [23]) according to which every absolutely continuous probability measure μ on $\mathbb{R}^n$ can be transformed into any probability measure ν on $\mathbb{R}^n$ by means of a transformation

T of the form $T(x) = \nabla\varphi(x)$, where φ is a convex function. Such transformations are natural analogs of increasing functions on the real line. This theorem plays an important role in obtaining a number of interesting inequalities in [19], [9]. In the papers [11], [12], an infinite dimensional analog of this theorem has been obtained in the partial case of a Gaussian measure, namely, it has been shown that a Radon Gaussian measure γ with the Cameron–Martin space H can be transformed into any probability measure $\nu = f \cdot \gamma$ that is absolutely continuous with respect to γ and satisfies some additional restriction by a monotone (in a natural sense) mapping of the form

$$T(x) = x + F(x), \quad F\colon X \to H, \tag{1.1}$$

where $F(x) = \nabla_H\varphi(x)$ for some function φ from the Sobolev class $W^{2,1}(\gamma)$. For example, the aforementioned restriction is satisfied if $f \log f \in L^1(\gamma)$. The transformations of the form (1.1) are abstract Girsanov transformations of the classic Wiener space; many works are devoted to their study (see references in [3] and [22]). In the case of the classic Wiener space $C[0,1]$, the Cameron–Martin space consists of the absolutely continuous functions x with $x(0) = 0$ and $x' \in L^2[0,1]$. It was made clear in the works of Cameron and Martin, Maruayama, Prokhorov, Skorokhod, and Girsanov that under broad assumptions a mapping of the form

$$T(w)(t) = w(t) + \int_0^t \sigma\big(s, w(\cdot)\big)\, ds$$

transforms the Wiener measure into an equivalent one, and many important in applications mappings that accomplish equivalent transformations of the Wiener measure have the indicated form. We remark that measures of the form $\gamma \circ T^{-1}$ with a mapping T of type (1.1) are called representable in Definition 2.7.1 of the book [22]. It is shown in Chapter 2 of that book that the set of representable measures is everywhere dense in the set of all probability measures absolutely continuous with respect to γ in the metric generated by the variation norm. In addition, it is proved there that any probability measure μ equivalent to γ can be transformed into γ by a mapping of the form (1.1). It has been shown in [10] that the measure γ can be transformed into every probability measure ν that is absolutely continuous with respect to it by a mapping of the form $T(x) = U(x) + F(x)$, where U preserves γ and $F\colon X \to H$. However, the problem of existence of a transformation $T = I + F$ with a mapping F taking on values in the Cameron–Martin space and transforming the measure γ into an arbitrary probability measure ν absolutely continuous with respect to γ was left open in the cited papers. Our main results (they have been announced in [7]; here some assertions are reinforced) are as follows. In §2 we introduce and study *canonical triangular mappings* of measures. In §3 we establish an exact equality for the entropy of a nonlinear image of a measure on $\mathbb{R}^n$, which yields Talagrand's inequality and reinforces a result of the work [16], obtained for gradient mappings. In §4 some applications of the obtained results are given. In the infinite dimensional case, we prove the existence of a triangular mapping T of a centered Radon Gaussian γ into an arbitrary probability measure ν that is absolutely continuous with respect to γ such that $T(x) = x + F(x)$, where F takes on values in the Cameron–Martin space H of the measure γ. A generalized logarithmic Sobolev inequality is obtained.

On Gaussian measures, see [3]. One can assume throughout that we deal with the countable power of the standard Gaussian measure on the real line. In this case, the Cameron–Martin space H is the classic Hilbert space l^2.

Let $\mu \circ T^{-1}$ denote the image of a measure μ under a measurable mapping T, i.e., the measure defined by the equality $\mu \circ T^{-1}(B) := \mu\big(T^{-1}(B)\big)$. Let $W_{loc}^{p,r}(\mathbb{R}^n)$, where $p \geq 1$ and $r \in \mathbb{N}$, denote the Sobolev class of functions on $\mathbb{R}^n$ which along with their generalized partial derivatives up to order r belong to $L^p(U)$ for every ball U.

2. Properties of triangular transformations of measures

We recall (details can be found in [4, Ch. 10]) that for every Borel probability measure μ on the space $\mathbb{R}^n \times \mathbb{R}^k$ having the projection μ_n on $\mathbb{R}^n$, there exist conditional probability measures

μ^x on the affine subspaces $x + \mathbb{R}^k$, where $x \in \mathbb{R}^n$, such that for every Borel set B in $\mathbb{R}^n \times \mathbb{R}^k$, the function $x \mapsto \mu^x(B)$ is Borel measurable and one has

$$\mu(B) = \int_{\mathbb{R}^n} \mu^x(B)\,\mu_n(dx). \tag{2.1}$$

One can also define conditional measures on one and the same space $\mathbb{R}^k$ (such conditional measures will be denoted by μ_x in order to distinguish them from the conditional measures μ^x on the fibers), but then (2.1) should be written in the form

$$\mu(B) = \int_{\mathbb{R}^n} \mu_x(B^x)\,\mu_n(dx).$$

where $B^x := \{y \in \mathbb{R}^k \colon (x, y) \in B\}$. The measure μ^x is the image of the measure μ_x under the shift by the vector x. For every bounded (or μ-integrable) Borel function φ on $\mathbb{R}^{n+k}$ there holds the equality

$$\int_{\mathbb{R}^{n+k}} \varphi(z)\,\mu(dz) = \int_{\mathbb{R}^n} \int_{x+\mathbb{R}^k} \varphi(z)\,\mu^x(dz)\,\mu_n(dx)$$
$$= \int_{\mathbb{R}^n} \int_{\mathbb{R}^k} \varphi(x, y)\,\mu_x(dy)\,\mu_n(dx),$$

where in the last integral the points of $\mathbb{R}^{n+k}$ are written in the form $z = (x, y)$.

If the measure μ is given by a density ϱ with respect to Lebesgue measure (by convention we take a Borel version of ϱ), then the measure μ_n is given by density

$$\varrho_n(x) = \int_{\mathbb{R}^k} \varrho(x, y)\,dy,$$

and the conditional measures are given by densities

$$\varrho^x(y) = \varrho(x, y)\Big(\int_{\mathbb{R}^k} \varrho(x, z)\,dz\Big)^{-1}$$

with respect to the k-dimensional Lebesgue measure. In this case, for the points x such that the integral of $\varrho(x, z)$ in z vanishes (the set of all such points is Borel measurable and has μ_n-measure zero), we take for ϱ^x some fixed probability density (for definiteness we take the standard Gaussian density).

For every pair of probability measures μ and ν on $\mathbb{R}^n$, where μ is absolutely continuous, there exists a Borel increasing triangular mapping $T_{\mu,\nu}$ that transforms μ into ν. This mapping is defined on some Borel set of full μ-measure, and every kth component of $T_{\mu,\nu}$, as a function of the variables $x_1, \ldots, x_k$, is defined on a Borel set in $\mathbb{R}^k$ whose intersections with the straight lines parallel to the kth coordinate line are intervals. As shown below, such a mapping is unique up to a redefinition on a set of μ-measure zero provided that ν possesses nonatomic conditional measures on the coordinate lines (e.g., is absolutely continuous). We shall call *canonical* the version of $T_{\mu,\nu}$ that is defined as follows by induction on n. For $n = 1$ we set

$$F_\mu(t) := \mu\big((-\infty, t)\big), \quad t \in \mathbb{R}^1, \quad G_\mu(u) := \inf\big\{s \colon F_\mu(s) \geq u\big\}, \quad u \in (0, 1),$$
$$T_{\mu,\nu} := G_\nu \circ F_\mu.$$

If the function G_ν has a finite limit as $u \to 0$ or $u \to 1$, then we define $G_\nu(0)$ or $G_\nu(1)$ as the corresponding limit. If the function F_μ assumes some of the values 0 and 1 (the sets $F_\mu^{-1}(0)$ and $F_\mu^{-1}(1)$ are either empty or rays) and the function G_ν has no finite limit at the corresponding point, then the mapping $F_{\mu,\nu}$ is defined on some interval (bounded or unbounded) of full μ-measure. The mapping F_μ takes μ to Lebesgue measure λ on $(0, 1)$, and G_ν takes λ to ν. This remains true for any probability measures provided that μ has no atoms. Note that the function G_ν is left continuous. Indeed, suppose that points u_i increase to u, but $G_\nu(u_i) < G_\nu(u) - \varepsilon$, where $\varepsilon > 0$. Then there exist points s_i with $F_\nu(s_i) \geq u_i$ and $s_i < G_\nu(u) - \varepsilon$. We may assume that the sequence $\{s_i\}$ increases to some $s \leq G_\nu(u) - \varepsilon$. This gives $F_\nu(s) \geq u$ by the left continuity of F_ν, which contradicts the definition of $G_\nu(u)$. Since the function F_μ is continuous,

the function $G_\nu \circ F_\mu$ is left continuous. If the measure ν is equivalent to Lebesgue measure (or at least is positive on all intervals), then this function is continuous. It is readily seen that the function G_ν is injective (unlike, generally speaking, the function F_ν) if ν has no atoms (for example, is absolutely continuous). In the case $n = 2$ we denote by μ_1 and ν_1 the projections of the measures μ and ν to the first coordinate line and take the canonical mapping T_1 that takes μ_1 to ν_1. Let μ_x and ν_x denote the conditional probability measures on the second coordinate line, $x \in \mathbb{R}^1$. For μ_1-almost every x_1, one has the conditional probability density

$$\varrho_\mu^{x_1}(x_2) := \varrho_\mu(x_1, x_2)\Big(\int_{\mathbb{R}} \varrho_\mu(x_1, u)\, du\Big)^{-1}.$$

The measure on the real line with density $\varrho_\mu^{x_1}$ can be transformed by a one dimensional canonical mapping to the conditional probability measure $\nu_{T_1(x_1)}$. The corresponding canonical mapping is denoted by $x_2 \mapsto T_2(x_1, x_2)$ (according to the one dimensional case, the domain of definition of this mapping may be a proper interval). If ν has a density ϱ_ν, then $\nu_{T_1(x_1)}$ is defined by density $\varrho_\nu^{T_1(x_1)}(\,\cdot\,)$, where

$$\varrho_\nu^{x_1}(x_2) := \varrho_\nu(x_1, x_2)\Big(\int_{\mathbb{R}} \varrho_\nu(x_1, u)\, du\Big)^{-1}.$$

We observe that $\int \varrho_\nu\big(T_1(x_1), u\big)\, du > 0$ for μ_1-almost all x_1 by the equality $\nu_1 = \mu_1 \circ T_1^{-1}$. It is clear that $T := (T_1, T_2)$ is an increasing triangular mapping and takes μ into ν. Indeed, for every bounded Borel function h on $\mathbb{R}^2$ we have

$$\int h(x, t)\, d\nu = \int \int h(x, t)\, \nu_x(dt)\nu_1(dx).$$

By using twice the change of variable, we obtain a chain of equalities

$$\int \int h(x, t)\, \nu_x(dt)\, \nu_1(dx) = \int \int h\big(T_1(x), t\big)\, \nu_{T_1(x)}(dt)\, \mu_1(dx)$$

$$= \int \int h\big(T_1(x), T_2(x, t)\big)\, \mu_x(dt)\, \mu_1(dx) = \int h \circ T\, d\mu.$$

The construction continues inductively by using one dimensional conditional densities on the last coordinate line. If for some $n \geq 1$ the existence of canonical triangular mappings is already established, then such a mapping for measures on $\mathbb{R}^{n+1}$ is constructed in the same way as in the two dimensional case considered above. Namely, the projections of the measures μ and ν on $\mathbb{R}^n$ are denoted by μ_n and ν_n. The corresponding conditional measures on the last coordinate line are denoted by μ_x and ν_x, $x \in \mathbb{R}^n$, and the density of μ_x is denoted by ϱ_μ^x. By the inductive assumption there exists a canonical Borel triangular mapping $T = (T_1, \ldots, T_n)\colon \mathbb{R}^n \to \mathbb{R}^n$ taking μ_n to ν_n (the domain of definition of T may be a proper Borel subset of $\mathbb{R}^n$ of full μ_n-measure). We take for $T_{\mu,\nu}$ the mapping $T_{\mu,\nu} = (T_1, \ldots, T_{n+1})$, where the last component is defined as follows: for fixed $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, the function $t \mapsto T_{n+1}(x_1, \ldots, x_n, t)$ is the canonical transformation of the measure with density ϱ_μ^x to the measure $\nu_{T(x)}$. The domain of definition of T_{n+1} is a Borel set of full μ-measure such that its intersections with the straight lines parallel to the last coordinate line are intervals (bounded or unbounded). Let us verify that the function T_{n+1} is Borel. Since it is increasing and left continuous in x_{n+1}, it suffices to verify that it is Borel with respect to $x = (x_1, \ldots, x_n)$ whenever x_{n+1} is fixed (see, e.g., [4, Lemma 6.4.6]). By construction we have

$$T_{n+1}(x, x_{n+1}) = G_{\nu_{T(x)}}\big(F_{\mu_x}(x_{n+1})\big).$$

The function $x \mapsto F_{\mu_x}(x_{n+1}) = \mu_x\big((-\infty, x_{n+1})\big)$ is Borel measurable. Hence it suffices to show that the function $(x, z) \mapsto G_{\nu_{T(x)}}(z)$ on $\mathbb{R}^n \times (0, 1)$ is Borel measurable. Since it is increasing and left continuous in z, one has only to verify that the function $x \mapsto G_{\nu_{T(x)}}(z)$ is Borel measurable for every fixed z, which by the Borel measurability of T reduces everything to the justification of the Borel measurability of the function $\psi\colon x \mapsto G_{\nu_x}(z)$. For every $c \in \mathbb{R}^1$ the set $\{x\colon \psi(x) < c\}$

is the projection of the set $E := \{(x,s)\colon \nu_x\big((-\infty,s)\big) \geq z,\, s < c\}$ on $\mathbb{R}^n$. The set E is Borel, because so is the function $(x,s) \mapsto \nu_x\big((-\infty,s)\big)$, which is increasing and left continuous in s and Borel measurable in x. Since the sections $E_x := \{s\colon (x,s) \in E\}$ are semiclosed intervals, hence are σ-compact, the Arsenin–Kunugui theorem (see [14, Theorem 35.46]) yields the Borel measurability of the projection of E. Thus, the Borel measurability of T is established. The chain of equalities indicated above remains valid if we replace μ_1 and ν_1 by μ_n and ν_n, which shows that we obtain a required mapping.

If the measure μ is equivalent to Lebesgue measure, then one can take a nonvanishing Borel version of its density. Then the mapping $T_{\mu,\nu}$ is defined on all of $\mathbb{R}^n$ (the same is true if the projection of μ on the first coordinate line and the conditional measures on the other coordinate lines are not concentrated on rays). In the general case, the domain of definition of $T_{\mu,\nu}$ is a Borel set of full μ-measure such that the kth component is defined on a Borel set in $\mathbb{R}^k$ whose intersections with the straight lines parallel to the kth coordinate line are intervals. The mapping $T_{\mu,\nu}$ is defined on the whole space also in the case where the measure ν is concentrated on a bounded set, because in such a case, if $n = 1$, the function $T_{\mu,\nu}$ is defined on the whole real line due to existence of finite limits of the function G_ν at the points 0 and 1. In general, one cannot always extend $T_{\mu,\nu}$ to an increasing mapping on the whole space. For example, if μ is Lebesgue measure on the interval $[0,1]$, considered as on the whole real line, and ν is the standard Gaussian measure on the real line, then the mapping $T_{\mu,\nu}$ is defined on $(0,1)$, but cannot be extended to an increasing function on the whole real line.

Finally, instead of transformations of measures on $\mathbb{R}^n$ one can deal with transformations of measures on the cube $[0,1]^n$. Then the mapping $T_{\mu,\nu}$ is defined on the whole cube, because in the one dimensional case it is defined on the whole closed interval $[0,1]$. In some respects, it is more convenient to consider mappings of the cube. We observe that the case $\mathbb{R}^n$ reduces to the case $[0,1]^n$. To this end, by using the mapping $(x_1,\ldots,x_n) \mapsto (\mathrm{arctg}\,x_1,\ldots,\mathrm{arctg}\,x_n)$ and its inverse we pass from $\mathbb{R}^n$ to $(0,1)^n$ (this preserves the class of increasing triangular Borel mappings). Given two measures μ and ν on $(0,1)^n$, we take the mapping $T_{\mu,\nu}$ on the cube $[0,1]^n$ corresponding to their extensions to this cube and let $\Omega = T_{\mu,\nu}^{-1}\big((0,1)^n\big)$.

It is clear that in the case where the measure μ is equivalent to Lebesgue measure, the mapping $T_{\mu,\nu}$ is injective, since its first component is injective on the real line, the second component $T_2(x_1,x_2)$ is injective as a function of x_2 with fixed x_1 and so on. Hence by Lusin's theorem $T_{\mu,\nu}$ takes all Borel sets to Borel ones. In the general case, if ν possesses atomless conditional measures on the coordinate lines (e.g., is absolutely continuous), then the mapping $T_{\mu,\nu}$ is injective on a Borel set of full μ-measure. Indeed, in the one dimensional case this is obvious (in this case ν is atomless). The multidimensional case is justified by induction. To this end, we take a set $E \subset \mathbb{R}^{n-1}$ with $\mu_{n-1}(E) = 1$ on which the mapping $(T_1,\ldots,T_{n-1})$ is injective. Then $E \times \mathbb{R}^1$ contains a set of full μ-measure on which T is injective, since for every $y = (x_1,\ldots,x_{n-1}) \in E$, the function $t \mapsto T_n(x_1,\ldots,x_{n-1},t)$ is injective on a set of full μ_y-measure.

We note that in a similar manner an increasing triangular Borel mapping $T_{\mu,\nu}$ is constructed in a more general case where ν is an arbitrary Borel probability measure on $\mathbb{R}^n$ and a Borel probability measure μ is such that its projections on $\mathbb{R}^k$, $k = 1,\ldots,n$, and the corresponding conditional measures have no atoms. A justification is given in [1]. With the help of finite dimensional canonical triangular mappings one constructs in an obvious way a canonical Borel triangular mapping $T_{\mu,\nu}$ on the space $\mathbb{R}^\infty$ transforming a Borel probability measure μ to a Borel probability measure ν, where it is assumed that the finite dimensional projections of μ satisfy the aforementioned condition, for example, are absolutely continuous. To this end, the components of $T_{\mu,\nu}$ are constructed inductively as the components of the canonical triangular mappings that transform the projections of μ on $\mathbb{R}^n$ to the projections of ν. According to our construction the first n components of the canonical mapping on $\mathbb{R}^{n+1}$ give the canonical mapping on $\mathbb{R}^n$.

It is clear from the construction that the mapping $T_{\mu,\nu}$ depends on our choice of conditional measures for μ and ν. However, there holds the following uniqueness property in the class of μ-equivalent increasing triangular mappings.

Lemma 2.1. *Let μ be a Borel probability measure on $\mathbb{R}^\infty$. Suppose that increasing triangular Borel mappings $T = (T_n)_{n=1}^\infty$ and $S = (S_n)_{n=1}^\infty$ are such that $\mu \circ T^{-1} = \mu \circ S^{-1}$ and that, for every n, the mapping $(T_1, \ldots, T_n)$ is injective on a Borel set of full measure with respect to the projection of μ on $\mathbb{R}^n$. Then $T(x) = S(x)$ μ-a.e.*

In particular, if the projections of the measures μ and ν on the spaces $\mathbb{R}^n$ are absolutely continuous, then there exists a canonical triangular mapping $T_{\mu,\nu}$ and it is unique up to μ-equivalence in the class of increasing Borel triangular mappings that transform μ into ν.

Proof. It is clear that our assertion reduces to the case of mappings on $\mathbb{R}^n$. We prove it by induction on n. Let $n = 1$. Suppose that a point x_0 belongs to the topological support of μ. If $T(x_0) < S(x_0)$, then x_0 cannot be an atom of μ, since $\mu(x: T(x) < t) = \mu(x: S(x) < t)$ for all t and one can take $t = (T(x_0) + S(x_0))/2$. Now we may assume that both functions T and S are continuous at x_0, since their discontinuity points form an at most countable set. By the continuity of both functions at x_0 there exists a point $x_1 > x_0$ that is not an atom of μ such that the functions T and S are continuous at x_1 and $T(x_1) < S(x_0)$. Taking $t = T(x_1)$ we obtain that there exists a point $y < x_0$ such that $\mu((y, x_1)) = 0$ contrary to the fact that x_0 belongs to the topological support of μ.

Suppose our assertion is proved for some $n \geq 1$. Let us consider the case of $\mathbb{R}^{n+1}$. Set $\nu := \mu \circ T^{-1} = \mu \circ S^{-1}$. Denote by μ_n and ν_n the projections of μ and ν on $\mathbb{R}^n$. On the last coordinate line we fix conditional measures μ_y and ν_y, $y \in \mathbb{R}^n$. By the inductive assumption, for any $i \leq n$ we have $T_i(x) = S_i(x)$ μ-a.e. Indeed, the images of the measure μ_n under the mappings $T_0 := (T_1, \ldots, T_n)$ and $S_0 := (S_1, \ldots, S_n)$ coincide (they coincide with ν_n). This gives $T_0 = S_0$ μ_n-a.e., which is equivalent to the coincidence of these mappings μ-a.e., because they depend only on $y := (x_1, \ldots, x_n)$. Let us show now that for μ_n-a.e. $y = (x_1, \ldots, x_n)$, there holds the equality $T_{n+1}(x_1, \ldots, x_n, x_{n+1}) = S_{n+1}(x_1, \ldots, x_n, x_{n+1})$ for μ_y-a.e. x_{n+1}. To this end, by the one dimensional case it suffices to verify the coincidence μ_n-a.e. of the measures $\mu_y \circ F_y^{-1}$ and $\mu_y \circ G_y^{-1}$, where

$$F_y(t) = T_{n+1}(x_1, \ldots, x_n, t), \quad G_y(t) = S_{n+1}(x_1, \ldots, x_n, t).$$

By hypothesis, there exists a Borel set $E \subset \mathbb{R}^n$ with $\mu_n(E) = 1$ such that the mapping $T_0 = S_0$ is Borel and injective on E. There is a Borel mapping J on $\mathbb{R}^n$ such that $J(T_0(y)) = J(S_0(y)) = y$ for all $y \in E$. Let us take a countable family of bounded Borel functions φ_i on $\mathbb{R}^n$ separating Borel measures and an analogous countable family of functions ψ_j on the real line. Let $\zeta_i = \varphi_i \circ J$. Then $\zeta_i(S_0(y)) = \zeta_i(T_0(y)) = \varphi_i(y)$ for all $y \in E$, i.e., μ_n-a.e. For all i and j, there holds the equality

$$\int_{\mathbb{R}^{n+1}} \zeta_i(y)\psi_j(t)\,\nu(dydt) = \int_{\mathbb{R}^{n+1}} \zeta_i\big(S_0(y)\big)\psi_j\big(S_{n+1}(y,t)\big)\,\mu(dydt)$$

$$= \int_{\mathbb{R}^n} \Big(\int_{\mathbb{R}^1} \psi_j\big(S_{n+1}(y,t)\big)\,\mu_y(dt)\Big)\varphi_i(y)\,\mu_n(dy) = \int_{\mathbb{R}^n} \Big(\int_{\mathbb{R}^1} \psi_j(t)\,\mu_y \circ G_y^{-1}(dt)\Big)\varphi_i(y)\,\mu_n(dy).$$

The same equality is fulfilled for the measures $\mu_y \circ F_y^{-1}$ in place of $\mu_y \circ G_y^{-1}$. Due to our choice of the functions φ_i and ψ_j we obtain the equality $\mu_y \circ G_y^{-1} = \mu_y \circ F_y^{-1}$ for μ_n-a.e. y. $\square$

The assumption that ν possesses atomless conditional measures on the coordinate lines is essential for the uniqueness statement. Indeed, let μ be Lebesgue on the square $[0,1]^2$ and let $T_1(x_1) = S_1(x_1) = 0$, $T_2(x_1, x_2) = x_2$, $S_2(x_1, x_2) = (x_2 + 1)/2$ if $0 \leq x_1 \leq 1/2$, and $S_2(x_1, x_2) = (x_2 - 1)/2$ if $1/2 < x_1 \leq 1$. Then T and S transform μ into Lebesgue measure on the unit interval of the second coordinate line.

It is interesting to compare the canonical triangular mapping $T_{\mu,\nu}$ transforming a probability measure μ on $\mathbb{R}^n$ to a probability measure ν with the mapping Ψ which also takes μ to ν and

satisfies the equality

$$\int_{\mathbb{R}^n} |\Psi(x) - x|^2 \, \mu(dx) = \inf_{\sigma \in M(\mu,\nu)} \left\{ \int_{\mathbb{R}^{2n}} |y - x|^2 \, \sigma(dxdy) \right\},$$

where $M(\mu, \nu)$ is the set of all probability measures on $\mathbb{R}^n \times \mathbb{R}^n$ having μ and ν as the projections to the factors. It is known (see [20], [23]) that if μ is absolutely continuous and both measures have finite second moments, then the minimum is attained at a measure σ which is the image of μ under the mapping $x \mapsto (\Psi(x), x)$ for some mapping Ψ that is the gradient of some convex function V. In the one dimensional case Ψ is an increasing function and hence coincides with $T_{\mu,\nu}$. However, in higher dimensions this is no longer true: typical triangular mappings are not gradients, because the derivative of a gradient is given by a symmetric matrix (such a matrix can be triangular only if it is diagonal). It would be interesting to clarify how big is the difference between the indicated minimum and the value given by the canonical triangular mapping.

Theorem 2.2. *Suppose a sequence of absolutely continuous probability measures ν_j on $\mathbb{R}^n$ converges in variation to a measure ν and let μ be a probability measure on $\mathbb{R}^n$ equivalent to Lebesgue measure. Then the sequence of canonical triangular mappings T_{μ,ν_j} converges in measure μ to the mapping $T_{\mu,\nu}$.*

Proof. Let $n = 1$. Then $\lim_{j\to\infty} T_{\mu,\nu_j}(t) = T_{\mu,\nu}(t)$ for almost every t, since the function F_μ is strictly increasing and $\lim_{j\to\infty} G_{\nu_j}(u) = G_\nu(u)$ for all points $u \in (0,1)$ at which the function G_ν is continuous, i.e., except for an at most countable set. Indeed, let $u_0 \in (0,1)$ be a continuity point of G_ν and $s_0 = G_\nu(u_0)$. It follows by the continuity of G_ν at u_0 that for an arbitrarily small $\varepsilon > 0$ there holds the inequality

$$F_\nu(s_0 - \varepsilon) < F_\nu(s_0) < F_\nu(s_0 + \varepsilon).$$

Convergence of measures ν_j to ν in variation yields uniform convergence of the functions F_{ν_j} to F_ν. Hence for $\delta := (F_\nu(s_0 + \varepsilon) - u_0)/2$ one can find a number N_1 such that for every $j > N_1$, there holds the inequality $|F_{\nu_j}(s_0 + \varepsilon) - F_\nu(s_0 + \varepsilon)| < \delta$. Therefore,

$$F_{\nu_j}(s_0 + \varepsilon) - u_0 = F_{\nu_j}(s_0 + \varepsilon) - F_\nu(s_0 + \varepsilon) + F_\nu(s_0 + \varepsilon) - u_0 \geq F_\nu(s_0 + \varepsilon) - u_0 - \delta = \delta > 0.$$

Hence for every $j > N_1$ we have $s_0 + \varepsilon \geq \inf\{t \colon F_{\nu_j}(t) \geq u_0\} = G_{\nu_j}(u_0)$. Similarly, there exists N_2 such that $s_0 - \varepsilon \leq G_{\nu_j}(u_0)$ for all $j \geq N_2$. Thus, $|G_{\nu_j}(u_0) - s_0| < \varepsilon$ for all sufficiently large j.

Suppose the theorem is proven for some $n \geq 1$ and we are given probability measures $\nu_j = f_j \, dx$ convergent in variation to a measure $\nu = f \, dx$ on $\mathbb{R}^{n+1}$. It suffices to verify that every subsequence in $\{T_{\mu,\nu_j}\}$ contains a further subsequence convergent μ-a.e. We recall the following fact (see Corollary 9.9.11 in [4]). Let μ be a Radon probability measure on a completely regular space X with the Borel σ-algebra $\mathcal{B}$, let $T_j \colon X \to X$ be measurable transformations convergent μ-a.e. to a transformation T, let the measures $\mu \circ T_j^{-1}$ and $\mu \circ T^{-1}$ be absolutely continuous with respect to μ, and let their densities $f_j := d(\mu \circ T_j^{-1})/d\mu$ be uniformly integrable. Suppose that $\mathcal{B}$-measurable functions φ_j converge in measure μ to a function φ. Then the functions $\varphi_j \circ T_j$ converge in measure to $\varphi \circ T$. Let P_n denote the projection from $\mathbb{R}^{n+1}$ on $\mathbb{R}^n$ and let $\mu^{(n)}$, $\nu^{(n)}$ and $\nu_j^{(n)}$ denote the projections of the measures μ, ν and ν_j on $\mathbb{R}^n$. It is clear that the measure $\mu^{(n)}$ is equivalent to Lebesgue measure. By the inductive assumption and the Riesz theorem one can assume, passing to a subsequence, that the mappings $P_n \circ T_{\mu,\nu_j}$ converge $\mu^{(n)}$-a.e. to the mapping $P_n \circ T_{\mu,\nu}$, since they are the natural extensions of the mappings $\Psi_j := T_{\mu^{(n)},\nu_j^{(n)}}$ and $\Psi := T_{\mu^{(n)},\nu^{(n)}}$, respectively, i.e., $P_n \circ T_{\mu,\nu_j} = \Psi_j \circ P_n$, $P_n \circ T_{\mu,\nu} = \Psi \circ P_n$. This is readily seen from the properties of canonical mappings. Since the measures ν_j converge in variation to the measure ν, passing to a subsequence once again, we may assume that $f_j \to f$ a.e. In addition, the densities f_j are uniformly integrable. Set

$$\Phi_j(x) := \big(\Psi_j(P_n x), x_{n+1}\big), \quad \Phi(x) := \big(\Psi(P_n x), x_{n+1}\big).$$

Let $\nu_{j,z}$, $z \in \mathbb{R}^n$, denote the conditional measures on the last coordinate line for the measure ν_j. It follows by the above described inductive construction of the components of canonical mappings and the considered one dimensional case that it suffices to obtain convergence of the one dimensional conditional measures $\nu_{j,\Psi_j(y)}$ to the conditional measure $\nu_{\Psi(y)}$ for almost all $y \in \mathbb{R}^n$. For every $z \in \mathbb{R}^n$ we let

$$\theta_j(z) := \int_{\mathbb{R}^1} f_j(z,s)\,ds, \quad \theta(z) := \int_{\mathbb{R}^1} f(z,s)\,ds.$$

The conditional measures $\nu_{j,z}$ and ν_z are given, respectively, by the densities

$$g_j(z,t) = f_j(z,t)/\theta_j(z), \quad g(z,t) = f(z,t)/\theta(z).$$

Here $g(z,t) = 0$ if $\theta(z) = 0$, and likewise for g_j. The functions θ_j converge to θ with respect to the norm of $L^1(\mathbb{R}^n)$, hence in measure $\mu^{(n)}$. According to the aforementioned assertion $\theta_j(\Psi_j(y)) \to \theta(\Psi(y))$ in measure $\mu^{(n)}$. Let us pass to an a.e. convergent subsequence. By the same assertion the functions $f_j \circ \Phi_j$ converge in measure μ to $f \circ \Phi$. Indeed, let us consider the measure $\mu' := \mu^{(n)} \otimes \gamma$, where γ is the standard Gaussian measure on the last coordinate line. This measure is equivalent to μ, and the measures $\mu' \circ \Phi_j^{-1} = \nu_j^{(n)} \otimes \gamma$ converge in variation to the measure $\mu' \circ \Phi^{-1} = \nu^{(n)} \otimes \gamma$. Passing to a subsequence we obtain that for $\mu^{(n)}$-almost every y, the functions $g_j(\Psi_j(y),t)$ of the real argument t converge a.e. to the function $g(\Psi(y),t)$. Since we deal with probability densities, we obtain convergence in $L^1(\mathbb{R}^1)$, i.e., convergence in variation of the corresponding measures, which gives convergence a.e. of the last components of T_{μ,ν_j}. □

There holds the following change of variables formula for increasing triangular mappings.

Lemma 2.3. *Let $T = (T_1, \ldots, T_n)\colon \mathbb{R}^n \to \mathbb{R}^n$ be an increasing Borel triangular mapping. Suppose that the functions*

$$x_i \mapsto T_i(x_1, \ldots, x_i)$$

are absolutely continuous on all compact intervals for a.e. $(x_1, \ldots, x_{i-1}) \in \mathbb{R}^{i-1}$. Set by definition $\det DT := \prod_{i=1}^{n} \partial_{x_i} T_i$. Then for every Borel function φ that is integrable on the set $T(\mathbb{R}^n)$, the function $\varphi \circ T \det DT$ is integrable over $\mathbb{R}^n$ and there holds the equality

$$\int_{T(\mathbb{R}^n)} \varphi(y)\,dy = \int_{\mathbb{R}^n} \varphi(T(x)) \det DT(x)\,dx. \tag{2.2}$$

If the mapping T is defined only on a Borel set $\Omega \subset \mathbb{R}^n$ and every function T_i is defined on a Borel set in $\mathbb{R}^i$ whose sections by the straight lines parallel to the ith coordinate line are intervals and the indicated condition is fulfilled for the compact intervals in those sections, then the same assertion is true with Ω in place of $\mathbb{R}^n$.

Proof. For $n = 1$ our assertion coincides with the classic change of variables formula for absolutely continuous functions. Next we apply induction on n and assume the assertion to be true in the case of dimension $n-1$. We make the function φ zero outside the Souslin set $T(\mathbb{R}^n)$. Let $S = (T_1, \ldots, T_{n-1})$. Then for almost every $y_n \in \mathbb{R}^1$, the function $(y_1, \ldots, y_{n-1}) \mapsto \varphi(y_1, \ldots, y_n)$ is integrable over $\mathbb{R}^{n-1}$, hence by the inductive assumption and the fact that the mapping S on $\mathbb{R}^{n-1}$ satisfies our hypotheses, we obtain

$$\int_{T(\mathbb{R}^n)} \varphi(y)\,dy = \int_{\mathbb{R}^n} \varphi(y)\,dy = \int_{-\infty}^{+\infty} \int_{\mathbb{R}^{n-1}} \varphi(S(z),y_n) \det DS(z)\,dz\,dy_n,$$

which after interchanging the limits of integration and the change of variable $y_n = T_n(z,x_n)$ for fixed $z \in \mathbb{R}^{n-1}$ leads to (2.2) by the equality $\det DT = (\det DS)\partial_{x_n} T_n$. A similar reasoning applies to the second case mentioned in the formulation, when T is defined on Ω. □

Let us give a simple sufficient condition on the measures μ and ν ensuring the absolute continuity of the ith component of $T_{\mu,\nu}$ with respect to the variable x_i.

Lemma 2.4. *A canonical triangular mapping $T_{\mu,\nu}$ on $\mathbb{R}^n$ that transforms an absolutely continuous probability measure μ to a probability measure ν satisfies the hypothesis of the preceding lemma if the measure ν is equivalent to Lebesgue measure.*

Proof. It suffices to observe that in the one dimensional case the function $T_{\mu,\nu}$ is absolutely continuous on the intervals, since $T_{\mu,\nu} = G_\nu \circ F_\mu$, where both functions are increasing and absolutely continuous on the intervals. The absolute continuity of F_μ is obvious and the absolute continuity (on every interval) of the function G_ν that is inverse to the absolutely continuous function F_ν follows by the fact that it is continuous, increasing and has Lusin's property (N) (see Exercise 5.8.48 in [4]). Property (N) follows by the condition $F_\nu' > 0$ a.e. (see Lemma 5.8.13 in [4]). $\qquad\square$

If the measure ν is not equivalent to Lebesgue measure, then the ith component of the canonical triangular mapping may be discontinuous. For example, the canonical mapping of Lebesgue measure on $[0,1]$ to the measure ν with density 2 on $[0,1/4] \cup [3/4,1]$ and 0 on $(1/4,3/4)$ has a jump. Nevertheless, the change of variables formula proven above remains valid without assumption on the absolute continuity made in the lemma if T is a canonical mapping of absolutely continuous measures (certainly, not every increasing Borel triangular mapping has this property).

Proposition 2.5. *Let μ and ν be probability measures on $\mathbb{R}^n$ with densities ϱ_μ and ϱ_ν with respect to Lebesgue measure. Then for the canonical triangular mapping $T_{\mu,\nu} = (T_1, \ldots, T_n)$ there holds the equality*

$$\varrho_\mu(x) = \varrho_\nu\big(T_{\mu,\nu}(x)\big) \det DT_{\mu,\nu}(x) \quad \text{for } \mu\text{-a.e. } x, \tag{2.3}$$

where $\det DT_{\mu,\nu} := \prod_{i=1}^n \partial_{x_i} T_i$ exists almost everywhere by the monotonicity of T_i in x_i.

Proof. Let us consider first the one dimensional case. Then $T_{\mu,\nu} = S \circ T$, where T is the canonical mapping of the measure μ to Lebesgue measure λ on $(0,1)$, i.e., the distribution function of the measure μ, and S is the canonical mapping of the measure λ to the measure ν, i.e., the inverse function to the distribution function F_ν of the measure ν. By differentiating the identity $F_\nu\big(S(y)\big) = y$ we obtain $\varrho_\nu\big(S(y)\big)S'(y) = 1$ a.e. Indeed, it suffices to observe that if Z is a Lebesgue measure zero set on which the derivative of F_ν does not exist or differs from ϱ_ν, then $S^{-1}(Z)$ has Lebesgue measure zero. This is a direct consequence of the equality $\lambda \circ S^{-1} = \nu$ and the absolute continuity of ν. Now we observe that

$$\varrho_\nu\Big(S\big(T(x)\big)\Big)S'\big(T(x)\big) = 1 \quad \text{for } \mu\text{-a.e. } x.$$

This is clear from the equality $\mu \circ T^{-1} = \lambda$. In a similar manner with the help of the equality $\mu \circ T^{-1} = \lambda$ we conclude that

$$T_{\mu,\nu}'(x) = S'\big(T(x)\big)T'(x) \quad \text{for } \mu\text{-a.e. } x.$$

Thus, for μ-a.e. x we obtain

$$\varrho_\nu\big(T_{\mu,\nu}(x)\big)T_{\mu,\nu}'(x) = \varrho_\nu\big(T_{\mu,\nu}(x)\big)S'\big(T(x)\big)T'(x) = T'(x) = \varrho_\mu(x).$$

Next we use induction on n and assume our assertion to be proven in dimension $n-1$. We write the points of $\mathbb{R}^n$ in the form (x, x_n), $x \in \mathbb{R}^{n-1}$. Set $\widetilde{T}(x) = \big(T_1(x), \ldots, T_{n-1}(x)\big)$. The projections of the measures μ and ν on $\mathbb{R}^{n-1}$ are denoted by μ' and ν', and their densities with respect to Lebesgue measure on $\mathbb{R}^{n-1}$ are denoted by $\varrho_{\mu'}$ and $\varrho_{\nu'}$, respectively. We observe that $\widetilde{T}$ coincides with $T_{\mu',\nu'}$. By the inductive assumption one has

$$\varrho_{\mu'}(x) = \varrho_{\nu'}\big(\widetilde{T}(x)\big) \det D\widetilde{T}(x) \quad \mu'\text{-a.e.} \tag{2.4}$$

For μ'-a.e. fixed $x \in \mathbb{R}^{n-1}$, the function $t \mapsto T_n(x,t)$ transforms the one dimensional conditional density $\varrho_\mu^x(x_n) = \varrho_\mu(x, x_n)/\varrho_{\mu'}(x)$ of the measure μ to the conditional density

$$\varrho_\nu^{\widetilde{T}(x)}(x_n) = \varrho_\nu\big(\widetilde{T}(x), x_n\big)/\varrho_{\nu'}\big(\widetilde{T}(x)\big)$$

of the measure ν. According to the one dimensional case we obtain

$$\frac{\varrho_\mu(x, x_n)}{\varrho_{\mu'}(x)} = \frac{\varrho_\nu\big(\widetilde{T}(x), T_n(x, x_n)\big)}{\varrho_{\nu'}\big(\widetilde{T}(x)\big)} \partial_{x_n} T_n(x, x_n) \quad \text{for } \mu_x\text{-a.e. } x_n.$$

Together with (2.4) and the equality $\det DT(x, x_n) = \partial_{x_n} T_n(x, x_n) \det D\widetilde{T}(x)$ this completes the proof. $\qquad\square$

We emphasize once again that the partial derivative in the formulation is an almost everywhere existing usual partial derivative, not the one in the sense of distributions (which has a singular component in the case of a function that is not absolutely continuous). This result substantially reinforces the one proved in [2] under additional conditions on the densities of the given measures. Let us give a sufficient condition for the continuous differentiability of a canonical mapping.

Lemma 2.6. *Suppose that probability measures μ and ν on $\mathbb{R}^n$ are given by continuous positive densities ϱ_μ and ϱ_ν whose Sobolev partial derivatives up to order $n + 1$ are integrable over $\mathbb{R}^n$. Then the canonical triangular mapping $T_{\mu,\nu}$ is continuously differentiable. The same is true if in place of the integrability of partial derivatives up to order $n + 1$ we require the continuity of the first order partial derivatives of the densities and the existence of nonnegative integrable functions $\theta_1,\ldots,\theta_n$ on the real line such that the functions ϱ_μ, ϱ_ν, $|\partial_{x_i}\varrho_\mu|$, $|\partial_{x_i}\varrho_\mu|$ are estimated by the function $\theta_1(x_1)\cdots\theta_n(x_n)$.*

Proof. First we consider the canonical mapping of the measure μ to Lebesgue measure λ on the open cube $(0, 1)^n$. In this case the last component T_n of the corresponding canonical mapping has the form

$$T_n(x, x_n) = \int_{-\infty}^{x_n} \varrho_\mu(x, s)\, ds \Big(\int_{-\infty}^{+\infty} \varrho_\mu(x, s)\, ds\Big)^{-1},$$

where the points of $\mathbb{R}^n$ are written as (x, x_n), $x \in \mathbb{R}^{n-1}$. We observe that the function $\partial_{x_n} T_n$ is continuous. To this end, it suffices to make sure of the continuity of the function

$$G(x) := \int_{-\infty}^{+\infty} \varrho_\mu(x, s)\, ds.$$

It is easily seen that this function has Sobolev partial derivatives up to order n that are integrable over $\mathbb{R}^{n-1}$, which by the Sobolev embedding theorem implies the continuity of G.

The functions $\partial_{x_i} T_n$, $i \le n-1$, are continuous as well. Indeed, by the positivity of ϱ_μ it suffices to verify the continuity of the functions

$$H_i(x, x_n) := \int_{-\infty}^{x_n} \partial_{x_i}\varrho_\mu(x, s)\, ds.$$

These functions have integrable Sobolev partial derivatives up to order $n + 1$, hence are continuous. Therefore, the mapping $T_{\mu,\lambda}$ is continuously differentiable.

In the general case, the mapping $T_{\mu,\nu}$ is the composition of the mappings $T_{\mu,\lambda}$ and $T_{\lambda,\nu} = T_{\nu,\lambda}^{-1}$. The continuous differentiability of $T_{\lambda,\nu}$ follows by the inverse mapping theorem due to the nondegeneracy of $DT_{\nu,\lambda}$. The second assertion of the lemma is proved similarly with the help of Lebesgue's dominated convergence theorem. $\qquad\square$

3. ESTIMATES OF THE ENTROPY OF THE RADON–NIKODYM DENSITIES

If a probability measure ν is given by a density f with respect to a probability measure μ and $f \log f \in L^1(\mu)$, where we set $f(x) \log f(x) = 0$ whenever $f(x) = 0$, then we let

$$\text{Ent}_\mu(f) := \int f \log f \, d\mu.$$

The Fredholm–Carleman determinant $\det_2 A$ of a matrix A is defined by the equality

$$\det{}_2 A = \exp\big(\text{Tr}(I - A)\big) \det A,$$

where I is the unit matrix and Tr is the matrix trace. Let D and D^2 denote the first and second derivatives. For a positive function ψ on $\mathbb{R}^n$ from the class C^2 and two vectors $v_1, v_2 \in \mathbb{R}^n$, we introduce the operator

$$\Lambda\big[\psi, v_1, v_2\big] = \int_0^1 s D^2\big[-\log \psi\big]\big((1-s)v_1 + s v_2\big)\, ds.$$

Theorem 3.1. *Suppose that we are given a probability measure $\mu = \exp(-\Phi)\, dx$ on $\mathbb{R}^n$ and two injective locally Lipschitzian mappings T_f and T_g with $\mu \circ T_f^{-1} = f \cdot \mu$ and $\mu \circ T_g^{-1} = g \cdot \mu = \exp(-\Theta)\, dx$ such that $\Phi, \Theta \in C^2(\mathbb{R}^n)$, $f > 0$ a.e. and $f \log(f/g) \in L^1(\mu)$. Let T_g have a locally Lipschitzian inverse T_g^{-1}, let $T = (T_1, \ldots, T_n) := T_f \circ T_g^{-1}$, and let $\det DT > 0$ a.e. Suppose that the functions*

$$\big\langle DT_f(x)\big(DT_g(x)\big)^{-1} e_i, e_i \big\rangle, \tag{3.1}$$

$$\log \det \Big[DT_f(x)\big(DT_g(x)\big)^{-1}\Big], \tag{3.2}$$

$$\partial_{x_i}\Theta(x)\big(T_i(x) - x_i\big)g(x) \tag{3.3}$$

are μ-integrable, or, which is equivalent, the functions

$$\partial_{x_i} T_i, \ \log \det DT, \ \partial_{x_i}\Theta(x)\big(T_i(x) - x_i\big)$$

are $g \cdot \mu$-integrable. Then there holds the equality

$$\mathrm{Ent}_{g\cdot\mu}\Big(\frac{f}{g}\Big) \tag{3.4}$$

$$= \int_{\mathbb{R}^n} \big\langle \Lambda\big[e^{-\Theta}, T_f, T_g\big]\big(T_f - T_g\big), T_f - T_g \big\rangle\, d\mu - \int_{\mathbb{R}^n} \log \det{}_2\big[DT_f(DT_g)^{-1}\big]\, d\mu$$

$$= \int_{\mathbb{R}^n} \Big\{ \big\langle \Lambda\big[e^{-\Theta}, T(x), x\big]\big(T(x) - x\big), T(x) - x \big\rangle - \log \det{}_2 DT(x) \Big\} g(x)\, \mu(dx).$$

In particular, this is true if $T_f = T_{\mu, f\cdot\mu}$ and $T_g = T_{\mu, g\cdot\mu}$ are canonical triangular mappings satisfying the indicated integrability conditions. In addition, in this case one has the inequality $\log \det{}_2 DT_f(DT_g)^{-1} \leq 0$.

Proof. The set $T_f(\mathbb{R}^n)$ is Borel measurable by the injectivity of T_f, and its complement has measure zero by the equivalence of the measure $f \cdot \mu$ and Lebesgue measure. Let $S_g := T_g^{-1}$. The mapping T is locally Lipschitzian as well. By the differentiation of composition formula one has

$$DT(x) = DT_f\big(S_g(x)\big) DS_g(x) = DT_f\big(S_g(x)\big)\big[DT_g\big(S_g(x)\big)\big]^{-1} \quad \text{a.e.}$$

By the change of variables formula for locally Lipschitzian mappings, for every Lebesgue integrable function φ on $\mathbb{R}^n$, the function $\varphi \circ T \det DT$ is also integrable and there holds the equality

$$\int_{\mathbb{R}^n} \varphi\big(T(x)\big)\det DT(x)\, dx = \int_{\mathbb{R}^n} \varphi(y)\, dy.$$

Hence for every bounded Borel function ψ on $\mathbb{R}^n$, on the account of the equality $f\cdot\mu = (g\cdot\mu)\circ T^{-1}$ we obtain

$$\int_{\mathbb{R}^n} \psi\big(T(x)\big) f\big(T(x)\big) \exp\big[-\Phi\big(T(x)\big)\big]\det DT(x)\, dx = \int_{\mathbb{R}^n} \psi(y) f(y)\exp[-\Phi(y)]\, dy$$

$$= \int_{\mathbb{R}^n} \psi\big(T(x)\big) g(x)\exp[-\Phi(x)]\, dx.$$

Since the mapping T is injective, we arrive at the equality

$$f\big(T(x)\big)\exp\big[-\Phi\big(T(x)\big)\big]\det DT(x) = g(x)\exp[-\Phi(x)] = \exp[-\Theta(x)] \quad \text{a.e.} \tag{3.5}$$

By taking the logarithm we obtain

$$\log f\big(T(x)\big) - \Phi\big(T(x)\big) + \log \det DT(x) = -\Theta(x) \quad \text{a.e.} \tag{3.6}$$

Adding $(\Phi - \log g)(T(x)) = \Theta(T(x))$ on both sides we obtain

$$\log f(T(x)) - \log g(T(x)) = \Theta(T(x)) - \Theta(x) - \log \det DT(x) \quad \text{a.e.}$$

Let us apply the identity

$$\Psi(a) - \Psi(b) = \langle \nabla \Psi(b), a - b \rangle + \int_0^1 s \langle D^2\Psi((b-a)s + a)(b-a), b - a \rangle \, ds \qquad (3.7)$$

to $\Psi = \Theta$ and $a = T(x)$, $b = y$. We obtain almost everywhere

$$\log f(T(x)) - \log g(T(x)) = \langle \nabla\Theta(x), T(x) - x \rangle$$
$$+ \int_0^1 s \langle D^2\Theta(sx + (1-s)T(x))(x - T(x)), x - T(x) \rangle \, ds - \log \det DT(x). \quad (3.8)$$

Note that the left-hand side of equality (3.8) is integrable with respect to the measure $g \cdot \mu$ by the integrability of the function $f \log(f/g)$ with respect to the measure μ and the equality $(g \cdot \mu) \circ T^{-1} = f \cdot \mu$. Hence

$$\int_{\mathbb{R}^n} \log \frac{f(T(x))}{g(T(x))} \, g(x) \, \mu(dx) = \int_{\mathbb{R}^n} f(x) \log \frac{f(x)}{g(x)} \, \mu(dx).$$

The first and third summands on the right in (3.8) are integrable with respect to the measure $g \cdot \mu$ by hypothesis. Hence the second summand on the right in (3.8) is also integrable with respect to $g \cdot \mu$. Integrating (3.8) against the measure $g \cdot \mu$, we arrive at the equality

$$\int_{\mathbb{R}^n} \log \frac{f(T(x))}{g(T(x))} \, g(x) \, \mu(dx) = \int_{\mathbb{R}^n} \langle \nabla\Theta(x), T(x) - x \rangle g(x) \, \mu(dx) \qquad (3.9)$$
$$+ \int_{\mathbb{R}^n} \left[\int_0^1 s \langle D^2\Theta(sx + (1-s)T(x))(x - T(x)), x - T(x) \rangle \, ds \right] g(x) \, \mu(dx)$$
$$- \int_{\mathbb{R}^n} \log \det \left\{ DT_f(S_g(x)) [DT_g(S_g(x))]^{-1} \right\} g(x) \, \mu(dx).$$

Since $(g \cdot \mu) \circ S_g^{-1} = \mu$, the last integral on the right-hand side is written as

$$\int_{\mathbb{R}^n} \log \det DT_f(x) (DT_g(x))^{-1} \, \mu(dx).$$

We observe that $-\nabla\Theta$ is the logarithmic gradient of the function $\varrho := \exp(-\Theta)$, i.e., $-\nabla\Theta = \nabla\varrho/\varrho$. By hypothesis, the functions $\dfrac{\partial_{x_i}\varrho(x)}{\varrho(x)}(T_i(x) - x_i)$ and $\partial_{x_i}T_i$ are integrable with respect to the measure $\varrho \, dx = g \cdot \mu$, because the integrability of the functions

$$\partial_{x_i}T_i(x) = \langle DT(x)e_i, e_i \rangle = \left\langle DT_f(S_g(x))[DT_g(S_g(x))]^{-1}e_i, e_i \right\rangle$$

with respect to the measure $g \cdot \mu$ follows, in view of the equality $(g \cdot \mu) \circ S_g^{-1} = \mu$, by the integrability of the functions $\left\langle DT_f(x)(DT_g(x))^{-1}e_i, e_i \right\rangle$ with respect to the measure μ. Thus, by means of the integration by parts formula the first integral on the right-hand side in (3.9) is transformed to

$$\int_{\mathbb{R}^n} \sum_{i=1}^n \partial_{x_i}(T_i(x) - x_i) \, g(x) \, \mu(dx) = \int_{\mathbb{R}^n} \sum_{i=1}^n \partial_{x_i}T_i(x) \, g(x) \, \mu(dx) - n$$
$$= \int_{\mathbb{R}^n} \sum_{i=1}^n \left\langle DT_f(x)(DT_g(x))^{-1}e_i, e_i \right\rangle \mu(dx) - n.$$

The justification of integration by parts under our assumptions follows by [3, Theorem 5.1.2]. By using the equality $x = T_g(S_g(x))$ and changing variables once again we arrive at the identity

$$\int_{\mathbb{R}^n} f \log \frac{f}{g} \, d\mu = \int_{\mathbb{R}^n} \big\langle \Lambda\big[e^{-\Theta}, T_f, T_g\big](T_f - T_g), T_f - T_g \big\rangle \, d\mu$$

$$+ \int_{\mathbb{R}^n} \left(\mathrm{Tr}\Big[DT_f(x)\big(DT_g(x)\big)^{-1}\Big] - n - \log\det\Big[DT_f(x)\big(DT_g(x)\big)^{-1}\Big] \right) \mu(dx).$$

$$= \int_{\mathbb{R}^n} \big\langle \Lambda\big[e^{-\Theta}, T_f, T_g\big](T_f - T_g), T_f - T_g \big\rangle \, d\mu - \int_{\mathbb{R}^n} \log\det\nolimits_2 \Big[DT_f(x)\big(DT_g(x)\big)^{-1}\Big] \, d\mu.$$

In the case where T_f and T_g are canonical triangular mappings, the matrices DT_f and DT_g are lower-triangular and their diagonal elements are positive. The matrix $DT_f(DT_g)^{-1}$ possesses these two properties as well. It remains to observe that if A is a triangular matrix with numbers $\alpha_i > 0$ at the diagonal, then $0 < \det_2 A \leq 1$, since $\alpha_i \exp(1 - \alpha_i) \leq 1$. $\qquad \square$

In the case where $T_g(x) = x$ and $g = 1$ we arrive at the following assertion.

Corollary 3.2. *Suppose we are given probability measures $\mu = \exp(-\Phi)\,dx$ and $f \cdot \mu$ on $\mathbb{R}^n$, where $\Phi \in C^2(\mathbb{R}^n)$ and $f > 0$ a.e. Let $T = (T_1, \ldots, T_n)$ be a locally Lipschitzian mapping such that $\mu \circ T^{-1} = f \cdot \mu$ and $\det DT > 0$ a.e. Suppose that the functions*

$$f \log f, \ \ \langle DT(x)e_i, e_i \rangle, \ \ \log \det DT(x), \ \ \partial_{x_i} \Phi(x)\big(T_i(x) - x_i\big)$$

are integrable with respect to the measure μ. Then one has

$$\mathrm{Ent}_\mu(f) = \int_{\mathbb{R}^n} \big\langle \Lambda\big[e^{-\Phi}, T(x), x\big](T(x) - x), T(x) - x \big\rangle \, \mu(dx) - \int_{\mathbb{R}^n} \log \det\nolimits_2 DT \, d\mu.$$

In particular, this is true if T is a canonical triangular mapping satisfying the indicated integrability conditions. In addition, in this case one has the inequality $\log \det_2 DT \leq 0$.

Remark 3.3. It is obvious from the proof of the theorem that the regularity conditions on the functions Φ, Θ and mappings T_f and T_g can be relaxed as follows: it suffices to have the inclusions $T_i, g \in W^{1,1}_{loc}(\mathbb{R}^n)$, $\Phi, \Theta \in W^{1,2}_{loc}(\mathbb{R}^n)$ and the validity of formula (3.5). This concerns also the corollary. Moreover, the same reasoning leads to the following general result.

Theorem 3.4. *Let measures $\mu = \exp(-\Phi)\,dx$, $f \cdot \mu$ and $g \cdot \mu = \exp(-\Theta)\,dx$ satisfy the hypotheses of the preceding theorem and let Borel mappings T_f and T_g on $\mathbb{R}^n$ be such that $\mu \circ T_f^{-1} = f \cdot \mu$, $\mu \circ T_g^{-1} = g \cdot \mu$. Suppose that there exist a Borel mapping $T = (T_1, \ldots, T_n)$ on $\mathbb{R}^n$ and an almost everywhere positive measurable function J_T such that $T_f = T \circ T_g$ and almost everywhere there holds the equality*

$$f\big(T(x)\big) \exp\big[-\Phi\big(T(x)\big)\big] J_T(x) = g(x) \exp[-\Phi(x)] = \exp[-\Theta(x)]. \tag{3.10}$$

In addition, suppose that the functions $x_i \mapsto T_i(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)$ are absolutely continuous on all intervals for almost all $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$, and one has

$$\frac{f}{g} \log \frac{f}{g}, \ \ \partial_{x_i} T_i, \ \ \log J_T, \ \ \partial_{x_i}\Theta(x)\big(T_i(x) - x_i\big) \in L^1(g \cdot \mu). \tag{3.11}$$

Then there holds the equality

$$\mathrm{Ent}_{g \cdot \mu}\left(\frac{f}{g}\right) = \int_{\mathbb{R}^n} \big\langle \Lambda\big[e^{-\Theta}, T(x), x\big](T(x) - x), T(x) - x \big\rangle$$

$$+ \int_{\mathbb{R}^n} \big[\mathrm{div}T(x) - n - \log J_T(x)\big] g(x)\, \mu(dx),$$

where $\mathrm{div}T(x) := \sum_{i=1}^{n} \partial_{x_i} T_i(x)$.

In particular, this equality holds true if $f > 0$ a.e., T_f and T_g are canonical triangular mappings and condition (3.11) is fulfilled.

Remark 3.5. The conditions imposed on the mappings T_f and T_g and the measures $f \cdot \mu$ and $g \cdot \mu$ are not independent; however, in the stated form they are more convenient for the subsequent applications. In the case where $g = 1$ and $T_g = I$, the following condition on the mapping T_f is sufficient for the validity of formula (3.5): T_f belongs to the Sobolev class $W_{loc}^{1,1}(\mathbb{R}^n, \mathbb{R}^n)$, is invertible, has Lusin's property (N), and $\det DT(x) > 0$ a.e. (see [13]). For example, if $T_f \in W_{loc}^{p,1}(\mathbb{R}^n, \mathbb{R}^n)$, where $p > n$, then it suffices that the continuous version of T_f be invertible and $\det DT(x) > 0$ a.e.

Remark 3.6. (i) All the integrability conditions imposed in Theorem 3.1 are fulfilled if the mappings T_f and T_g coincide with the identity mapping outside some compact set.

(ii) The integrability conditions on the functions (3.1)–(3.3) are fulfilled if we are given the integrability of the function $|\nabla\Theta|g$ with respect to the measure μ and the function $T_k(x) - x_k$ for every $k = 1, \ldots, n$ has the form $\theta_k(x_1, \ldots, x_k)$, where the function θ_k on $\mathbb{R}^k$ has compact support. Indeed, in this case the functions $\partial_{x_i} T_i$, $\log \det DT = \sum\limits_{i=1}^n \log \partial_{x_i} T_i$ and $T_i(x) - x_i = \theta_i$ are bounded, since the functions θ_i and $\partial_{x_i}\theta_i$ are bounded and the function $\partial_{x_i} T_i = 1 + \partial_{x_i}\theta_i$ is separated from zero due to the fact that it does not vanish and equals 1 at all points x whose projections on $\mathbb{R}^k$ do not belong to some compact set.

(iii) In the case where $T_g(x) = x$, $g = 1$ and $F(x) := T_f(x) - x$, a sufficient condition for the integrability of the functions (3.1)–(3.3) is μ-integrability of the function

$$|\nabla\Phi(x)||F(x)| + \|DT_f(x)\| + |\log \det DT_f(x)|.$$

For example, this condition is fulfilled if the measure μ has all moments, there holds the estimate

$$\|DT_f(x)\| + |\nabla\Phi(x)| \le c_1 + c_2|x|^k,$$

and the function $\det DT_f(x)$ is separated from zero.

Remark 3.7. If the function $|T(x) - x|g(x)$ is integrable with respect to μ, which is equivalent to μ-integrability of the function $|T_f(x) - T_g(x)|$, then in place of the μ-integrability of every function $\langle DT_f(x)(DT_g(x))^{-1} e_i, e_i \rangle$ in (3.1) it suffices to have μ-integrability of their sum that equals $\mathrm{Tr}\Big[DT_f(x)(DT_g(x))^{-1} \Big]$. In other words, in place of $g \cdot \mu$-integrability of the functions $\partial_{x_i} T_i$ it suffices to require the integrability of their sum $\mathrm{div} T$. To this end, in the justification of the equality

$$\int_{\mathbb{R}^n} \langle \nabla\varrho(x), T(x) - x \rangle \, dx = -\int_{\mathbb{R}^n} \mathrm{div} T(x)\, \varrho(x)\, dx + n$$

with $\varrho = \exp(-\Theta)$ in place of ϱ, we substitute $\zeta\varrho$, where $\zeta \in C_0^\infty(\mathbb{R}^n)$. We obtain

$$\int_{\mathbb{R}^n} \zeta(x)\langle \nabla\varrho(x), T(x) - x \rangle \, dx + \int \langle \nabla\zeta(x), T(x) - x \rangle \varrho(x)\, dx$$

$$= -\int_{\mathbb{R}^n} \mathrm{div} T(x)\, \zeta(x)\varrho(x)\, dx + n\int_{\mathbb{R}^n} \zeta(x)\varrho(x)\, dx.$$

Now it remains to take a sequence of functions $\zeta_k \in C_0(\mathbb{R}^n)$ such that $0 \le \zeta_k \le 1$, $\zeta_k(x) = 1$ if $|x| \le k$ and $|\nabla\zeta_k(x)| \le M < \infty$ for some constant M independent of k. Here the second integral on the left-hand side approaches zero by Lebesgue's dominated convergence theorem.

For $n = 1$ and $g = 1$ the established equality (naturally, without traces and determinants) is present in the reasoning in [21], and in [9] a close argumentation with the use of Taylor's formula is employed in the derivation of an inequality for some optimal mappings. In the multidimensional case an equality of such a type in a more special situation was obtained first in [16]. The exact equality found by us may be a source of diverse functional inequalities, which will be demonstrated below.

Let us give a modification of Theorem 3.1 imposing somewhat different conditions. Suppose that as above we are given three probability measures $\mu = \exp(-\Phi)\, dx$, $g \cdot \mu = \exp(-\Theta)\, dx$

and $f \cdot \mu$ on $\mathbb{R}^n$ and mappings $T_f, T_g \colon \mathbb{R}^n \to \mathbb{R}^n$ transforming the measure μ to $g \cdot \mu$ and $f \cdot \mu$, respectively. Suppose that there exists a Borel mapping $T = (T_1, \ldots, T_n)$ such that $T_f = T \circ T_g$.

Theorem 3.8. (i) *Let* $T \in W^{1,1}_{loc}(\mathbb{R}^n, \mathbb{R}^n)$ *and let formula* (3.5) *be valid. Suppose that the measures* $f \cdot \mu$ *and* $g \cdot \mu$ *have finite second moments,* $|\nabla\Theta| \in L^2_{loc}(g \cdot \mu)$ *and that the quantity* $\mathrm{Ent}_{g\cdot\mu}\left(\frac{f}{g}\right)$ *is finite. Let* $\det DT \geq 0$ *and* $D^2\Theta \geq C \cdot I$, $C \in \mathbb{R}$. *Assume that there exists a function* $w \in L^1(g \cdot \mu)$ *such that* $-\log\det_2 DT \geq w$. *Then*

$$\mathrm{Ent}_{g\cdot\mu}\left(\frac{f}{g}\right) = \int_{\mathbb{R}^n} \left\langle \Lambda\left[e^{-\Theta}, T_f, T_g\right](T_f - T_g), T_f - T_g \right\rangle d\mu - \int_{\mathbb{R}^n} \log\det_2 DT\, g\, d\mu.$$

(ii) *Let* T *be a canonical triangular mapping. Then for the validity of the foregoing equality, in which by definition we set*

$$\det_2 DT := \prod_{i=1}^{n} \exp(1 - \partial_{x_i} T_i)\partial_{x_i} T_i,$$

it suffices to have only the following conditions: $f > 0$ *a.e., the measures* $f \cdot \mu$ *and* $g \cdot \mu$ *possess finite second moments,* $|\nabla\Theta| \in L^2_{loc}(g \cdot \mu)$, $(f/g)\log(f/g) \in L^1(g \cdot \mu)$ *and* $D^2\Theta \geq C \cdot I$, *where* $C \in \mathbb{R}$.

Proof. (i) Since $\log\det DT$ can be expressed via the logarithm of the Fredholm–Carleman determinant of the matrix DT by means of the equality

$$\log\det DT = \log\det_2 DT - \mathrm{Tr}(I - DT),$$

as in the proof of the preceding theorem, we obtain the formula

$$\log f\big(T(x)\big) - \log g\big(T(x)\big) - \big\langle \nabla\Theta(x), T(x) - x \big\rangle + \mathrm{Tr}\big(DT(x) - I\big)$$
$$= \int_0^1 s\Big\langle D^2\Theta\big(sx + (1-s)T(x)\big)\big(x - T(x)\big), x - T(x)\Big\rangle ds - \log\det_2 DT(x).$$

We observe that the functions standing on the left-hand side are locally integrable with respect to the measure $g \cdot \mu$. Indeed, the integrability of the function $\log f\big(T(x)\big) - \log g\big(T(x)\big)$ follows by the change of variables formula from the integrability of the function $f\log\frac{f}{g}$ with respect to the measure μ. Let $F(x) := T(x) - x$. The existence of finite second moments of the measures $f \cdot \mu$ and $g \cdot \mu$ means μ-integrability of the functions $x^2 g$ and $|T|^2 g$, since $(g \cdot \mu) \circ T^{-1} = f \cdot \mu$. Thus, $|F| \in L^2(g \cdot \mu)$. Taking into account that $\nabla\Theta \in L^2_{loc}(g \cdot \mu)$, Hölder's inequality gives the local $g \cdot \mu$-integrability of the functions $\langle \nabla\Theta, F \rangle$. Let us fix a nonnegative function $\varphi \in C_0^\infty(\mathbb{R}^n)$. Integrating the indicated formula with respect to the measure $\varphi g \cdot \mu$, we obtain

$$\int_{\mathbb{R}^n} \log\frac{f(T)}{g(T)}\varphi g\, d\mu - \int_{\mathbb{R}^n} \langle\nabla\Theta, F\rangle\varphi g\, d\mu + \int_{\mathbb{R}^n} \mathrm{Tr}(DF)\varphi g\, d\mu$$
$$= \int_{\mathbb{R}^n} \left[\int_0^1 s\Big\langle D^2\Theta\big(x + (1-s)F(x)\big)\big(F(x)\big), F(x)\Big\rangle ds\right]\varphi(x)g(x)\, \mu(dx)$$
$$- \int_{\mathbb{R}^n} \log\det_2 DT\varphi g\, d\mu.$$

The splitting of the integral on the right-hand side in the sum of the two integrals is possible, because both integrands are estimated from below by integrable functions. Integrating by parts, we obtain

$$\int_{\mathbb{R}^n} \log\frac{f(T)}{g(T)}\varphi g\, d\mu - \int_{\mathbb{R}^n} \big\langle\nabla\varphi, F\big\rangle g\, d\mu$$
$$= \int_{\mathbb{R}^n} \left[\int_0^1 s\Big\langle D^2\Theta\big(x + (1-s)F(x)\big)\big(F(x)\big), F(x)\Big\rangle ds\right]\varphi(x)g(x)\, \mu(dx)$$
$$- \int_{\mathbb{R}^n} \log\det_2 DT\varphi g\, d\mu.$$

Let us take a sequence of functions $\varphi_j \in C_0^\infty(\mathbb{R}^n)$ with $0 \leq \varphi_j \leq 1$ and uniformly bounded gradients such that $\varphi_j(x) = 1$ whenever $|x| \leq j$. By using that

$$\left\langle D^2\Theta\big(x + (1-s)F(x)\big)\big(F(x)\big), F(x)\right\rangle \geq C|F(x)|^2$$

and $-\log\det_2 DT \geq w$, where the functions $|F|^2 g$ and wg are integrable with respect to μ, we conclude by Fatou's theorem that the functions $\log\det_2 DT$ and

$$\int_0^1 s\left\langle D^2\Theta\big(x + (1-s)F(x)\big)\big(F(x)\big), F(x)\right\rangle ds$$

are integrable with respect to the measure $g \cdot \mu$ and that in the last equality one can set $\varphi = 1$. After a change of variables this gives

$$\int_{\mathbb{R}^n} \log\frac{f}{g} f\, d\mu = \int_{\mathbb{R}^n} \left\langle \Lambda\big[e^{-\Theta}, T_f, T_g\big]\big(T_f - T_g\big), T_f - T_g\right\rangle d\mu$$

$$- \int_{\mathbb{R}^n} \log\det_2 DT g\, d\mu.$$

Assertion (i) is proven. Assertion (ii) is readily seen from the above reasoning, since formula (3.5) is valid according to what has been proven above, $\det DT \geq 0$, $\det_2 DT \leq 1$, and the functions $x_i \mapsto T_i(x_1, \ldots, x_{i-1}, x_i)$ are absolutely continuous on all intervals. $\qquad\square$

We shall say that a probability measure μ on $\mathbb{R}^n$ with a twice differentiable density $\exp(-\Phi)$ is uniformly convex with constant $C > 0$ if

$$D^2\Phi(x) \geq C \cdot I.$$

Corollary 3.9. *Suppose that the measure $g \cdot \mu$ in Theorem 3.8 is uniformly convex with constant C. Then for the canonical triangular mappings T_f and T_g we have*

$$\mathrm{Ent}_{g\cdot\mu}\left(\frac{f}{g}\right) \geq \frac{C}{2}\int_{\mathbb{R}^n} \big|T_f(x) - T_g(x)\big|^2 \mu(dx).$$

Proof. By the estimate $D^2\Theta \geq C \cdot I$ we obtain

$$\left\langle \int_0^1 sD^2\Theta\big((1-s)T_f + sT_g\big)(v), v\right\rangle ds \geq \frac{C}{2}|v|^2$$

for every vector v. $\qquad\square$

Corollary 3.10. *Suppose that a probability measure μ on the space $\mathbb{R}^n$ is uniformly convex with constant C (for example, let μ be the standard Gaussian measure). Let ν be an absolutely continuous probability measure on $\mathbb{R}^n$ such that letting $f := d\nu/d\mu$ we have $f \log f \in L^1(\mu)$. Then for the canonical triangular mapping $T_{\mu,\nu}$ we have*

$$\int |x - T_{\mu,\nu}(x)|^2 \mu(dx) \leq \frac{2}{C}\int f(x) \log f(x)\, \mu(dx). \tag{3.12}$$

In the case of the standard Gaussian measure one has $C = 1$.

Proof. Let us find a sequence of smooth functions f_j convergent a.e. to f such that the measures $\nu_j := f_j \cdot \mu$ are probabilistic, $f_j = 1$ outside of some cube K_j, and the quantities $\mathrm{Ent}_\mu(f_j)$ converge to $\mathrm{Ent}_\mu(f)$. To this end, we find first a positive convex function θ on $\mathbb{R}$ such that the function $\theta(|f \log f|)$ is integrable with respect to μ. It is readily verified that there exists a sequence of nonnegative smooth functions f_j convergent a.e. to f such that $f_j = 1$ outside of some cube K_j, $\int_{\mathbb{R}^n} f_j\, d\mu = 1$, and the quantities $\int \theta(|f_j \log f_j|)\, d\mu$ are uniformly bounded. By the uniform integrability of the functions $f_j \log f_j$ their integrals converge to the integral of $f \log f$. It follows by the definition of uniform convexity that the measure μ is equivalent to Lebesgue measure. By construction we have convergence of the probability densities f_j in $L^1(\mu)$, i.e., convergence in variation of the measures ν_j. Then by Theorem 2.2 the sequence of the canonical triangular mappings T_{μ,ν_j} converges in measure μ to the mapping $T_{\mu,\nu}$. Suppose

that for the mappings T_{μ,ν_j} estimate (3.12) is known. Then by Fatou's theorem it holds also for $T_{\mu,\nu}$. Thus, our assertion reduces to the case where the function f is smooth and coincides with 1 outside of some cube. Set $T := T_{\mu,\nu}$. By convexity the measure μ has finite second moment. Hence the measure $f \cdot \mu$ has finite second moment too, which gives μ-integrability of the function $|T|^2$. The function $|\nabla \Theta|$ is locally bounded by the convexity of Θ. Thus, assertion (ii) of the preceding theorem applies. $\qquad \square$

4. Some applications of triangular transformations

Here we discuss applications of the above results to measures on the countable product of real lines and to generalized logarithmic Sobolev inequalities.

Theorem 4.1. *Let μ be the product of countably many copies of a probability measure σ on the real line that is uniformly convex with constant C (for example, is a nondegenerate Gaussian measure) considered on the space $\mathbb{R}^\infty$ of all real sequences and let ν be a Borel probability measure on $\mathbb{R}^\infty$ that is absolutely continuous with respect to μ. Then there exists a Borel triangular mapping $T \colon \mathbb{R}^\infty \to \mathbb{R}^\infty$ such that $T(x) = x + F(x)$, where $F \colon \mathbb{R}^\infty \to l^2$ and $\nu = \mu \circ T^{-1}$. If, in addition, letting $f := d\nu/d\mu$, we have $f \log f \in L^1(\mu)$, then*

$$\int |F(x)|_{l^2}^2 \, \mu(dx) \leq \frac{2}{C} \int f(x) \log f(x) \, \mu(dx). \tag{4.1}$$

Proof. Suppose first that $f \log f \in L^1(\mu)$. Denote by ν_n the projection of ν on the space $\mathbb{R}^n$. It is clear that $\nu_n = f_n \cdot \mu_n$, where μ_n is the product of n copies of the measure σ and $f_n \in L^1(\mu_n)$. Indeed, for every Borel set A in $\mathbb{R}^n$ we have

$$\nu_n(A) = \int_{\mathbb{R}^\infty} I_A f \, \mu(dx) = \int_{\mathbb{R}^\infty} I_A \mathbb{E}_\mu^{\mathcal{B}_n} f \, \mu(dx) = \int_A \mathbb{E}_\mu^{\mathcal{B}_n} f \, \mu_n(dx) = \int_A f_n \, \mu_n(dx),$$

where $\mathbb{E}_\mu^{\mathcal{B}_n} f$ is the conditional expectation of f with respect to the σ-algebra $\mathcal{B}_n$ generated by the first n coordinate functions. Thus, $f_n = \mathbb{E}_\mu^{\mathcal{B}_n} f$. Since the function $x \log x$ is convex, Jensen's inequality for conditional expectations yields the uniform boundedness of the integrals $\int f_n \log f_n \, d\mu_n$. Let us take the canonical triangular mapping $T_n = (T_n^1, \ldots, T_n^n)$ on $\mathbb{R}^n$ transforming μ_n to ν_n. By the uniqueness of canonical mappings T_n coincides with $(T_{n+1}^1, \ldots, T_{n+1}^n)$, i.e., $T_n^k = T_m^k$ whenever $n, m \geq k$. Hence we obtain the functions $T^1 = T_1^1$, $T^2 = T_2^2, \ldots$, $T^n = T_n^n, \ldots$, defining a Borel triangular mapping T. For every cylindrical Borel set B with a base B_n in $\mathbb{R}^n$ we have

$$\mu \circ T^{-1}(B) = \mu_n\big(T_n^{-1}(B_n)\big) = \nu_n(B_n) = \nu(B).$$

Therefore, $\mu \circ T^{-1} = \nu$. The uniform boundedness of the integrals of $|x - T_n(x)|^2$ against the measure μ (which follows by Corollary 3.9) and Fatou's theorem yield that $F(x) := T(x) - x \in l^2$ for μ-a.e. x and estimate (4.1) holds true. We redefine F by zero on the Borel set $F^{-1}(\mathbb{R}^\infty \backslash l^2)$ of zero μ-measure.

In the general case we consider a partition of $\mathbb{R}^\infty$ into disjoint sets E_n of positive μ-measure on each of which the function f is bounded. To this end, we choose all sets of nonzero μ-measure among the sets $\{k \leq f < k+1\}$. Let us partition the real line into intervals D_n with $\mu_1(D_n) = \nu(E_n)$. The measure $I_{D_n} \cdot \mu_1$ can be transformed by means of an increasing mapping Ψ_n on the interval to the measure $\nu(E_n)\mu_1$. Let us represent μ in the form $\mu = \mu_1 \otimes \mu'$, where μ' is the product of countably many copies of σ corresponding to the coordinates $x_2, x_3, \ldots$. Then the mapping $x \mapsto \big(\Psi_n(x_1), x_2, x_3, \ldots\big)$ transforms the measure $(I_{D_n} \cdot \mu_1) \otimes \mu'$ to $\nu(E_n)\mu$. According to what has been proven above, there exists a triangular mapping Λ_n with $\Lambda_n(x) - x \in l^2$ transforming the measure $\nu(E_n)\mu$ to the measure $I_{E_n} \cdot \nu$. Therefore, we obtain a triangular mapping T that transforms the measure $\mu = \sum_{n=1}^\infty (I_{D_n} \cdot \mu_1) \otimes \mu'$ to the measure $\sum_{n=1}^\infty I_{E_n} \cdot \nu = \nu$. On $D_n \times \mathbb{R}^1 \times \mathbb{R}^1 \times \cdots$ the mapping T equals the composition of Ψ_n and Λ_n. Since Ψ_n has the aforementioned form, we obtain $x - T(x) \in l^2$. $\qquad \square$

Corollary 4.2. *Let γ be a centered Radon Gaussian measure on a locally convex space X, let H be the Cameron–Martin space of the measure γ, let $\{e_n\}$ be an orthogonal basis in H, and let $\widehat{e}_n$ be the corresponding measurable linear functionals (see [3, Ch. 2]). Then for every probability measure ν that is absolutely continuous with respect to the measure γ, there exists a Borel mapping $T\colon X \to X$ of the form $T(x) = x + F(x)$, where $F\colon X \to H$, such that $\nu = \gamma \circ T^{-1}$ and T is triangular with respect to $\{e_n\}$, i.e., $\widehat{e}_n \circ T$ is a Borel function of $\widehat{e}_1, \ldots, \widehat{e}_n$. In other words, $\big(F(x), e_n\big)_H = \varphi_n\big(\widehat{e}_1(x), \ldots, \widehat{e}_n(x)\big)$, where φ_n is a Borel function on $\mathbb{R}^n$. Finally, if for $f := d\nu/d\gamma$ we have $f \log f \in L^1(\gamma)$, then*

$$\int |F(x)|_H^2 \, \gamma(dx) \leq 2 \int f(x) \log f(x) \, \gamma(dx).$$

The established result means that every probability measure that is absolutely continuous with respect to γ is representable in the sense of Definition 2.7.1 in [22]. It remains unclear whether one can transform γ to any absolutely continuous probability measure ν by an increasing triangular mapping or by a monotone mapping of the type considered in [11] if we do not impose additional conditions on f.

The method employed above is applicable to proving some generalizations of the logarithmic Sobolev inequality (cf. [9], [19]).

Proposition 4.3. *Suppose we are given three probability measures $\mu = \exp(-\Phi)\,dx$, $f \cdot \mu$ and $g \cdot \mu$ on $\mathbb{R}^n$, where $f > 0$, $g > 0$, Φ is a function from the class C^2 on $\mathbb{R}^n$, and $f \log f$, $g \log g \in L^1(\mu)$. Assume also that we are given canonical triangular locally Lipschitzian mappings T_f and T_g, where T_g has a locally Lipschitzian inverse mapping T_g^{-1}, such that $\mu \circ T_f^{-1} = f \cdot \mu$ and $\mu \circ T_g^{-1} = g \cdot \mu$. Let $T = (T_1, \ldots, T_n) := T_f \circ T_g^{-1}$. Suppose that the functions*

$$\partial_{x_i} T_i, \ \log \det DT, \ \partial_{x_i}\Phi(x)\big(T_i(x) - x_i\big), \ \partial_{x_i} g(x)\big(T_i(x) - x_i\big)$$

are $g \cdot \mu$-integrable. Then there holds the inequality

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

In particular, in the case $f = 1$ we have $T = T_g^{-1}$ and

$$\int_{\mathbb{R}^n} g \log g \, d\mu \leq - \int_{\mathbb{R}^n} \big\langle \nabla g(x), T(x) - x \big\rangle \mu(dx) \tag{4.2}$$
$$- \int_{\mathbb{R}^n} \big\langle \Lambda[e^{-\Phi}, T(x), x]\big(T(x) - x\big), T(x) - x \big\rangle g(x)\mu(dx).$$

Proof. Let $T := T_f \circ T_g^{-1}$. Since the mappings T_f and T_g are canonical, one has $\det DT > 0$. In addition, it follows by the definition of T that $f \cdot \mu = (g \cdot \mu) \circ T^{-1}$. Therefore, equality (3.6) is fulfilled for T. Adding and subtracting $\big\langle \nabla\Phi(x), T(x) - x \big\rangle$ on the right-hand side of (3.6) and integrating the obtained formula with respect to the measure $g \cdot \mu$ we obtain the relationship

$$\int_{\mathbb{R}^n} f \log f - g \log g \, d\mu = \int_{\mathbb{R}^n} \big\langle \nabla\Phi(x), T(x) - x \big\rangle g(x) \, \mu(dx)$$
$$+ \int_{\mathbb{R}^n} \Big[\Phi\big(T(x)\big) - \Phi(x) - \big\langle \nabla\Phi(x), T(x) - x \big\rangle\Big] g(x) \, \mu(dx) - \int_{\mathbb{R}^n} \log \det DT(x) g(x) \, \mu(dx).$$

Let us transform the first integral on the right in the following way:

$$\int_{\mathbb{R}^n} \big\langle \nabla\Phi(x), T(x) - x \big\rangle g(x) \, \mu(dx) = - \int_{\mathbb{R}^n} \big\langle g(x)\nabla e^{-\Phi(x)}, T(x) - x \big\rangle \, dx$$
$$= - \int_{\mathbb{R}^n} \big\langle \nabla\big(g(x)e^{-\Phi(x)}\big), T(x) - x \big\rangle \, dx + \int_{\mathbb{R}^n} \big\langle e^{-\Phi(x)}\nabla g(x), T(x) - x \big\rangle \, dx.$$

Integrating by parts, which is possible under the imposed conditions according to [3, Theorem 5.1.2], we arrive at the equality

$$-\int_{\mathbb{R}^n}\Big\langle\nabla\big(g(x)e^{-\Phi(x)}\big),T(x)-x\Big\rangle\,dx=\int_{\mathbb{R}^n}(\mathrm{Tr}DT-n)g\,d\mu.$$

By using identity (3.7) with $a=T(x)$ and $b=x$, we obtain

$$\int_{\mathbb{R}^n}\Big[\Phi\big(T(x)\big)-\Phi(x)-\big\langle\nabla\Phi(x),T(x)-x\big\rangle\Big]g(x)\,\mu(dx)$$

$$=\int_{\mathbb{R}^n}\Big[\int_0^1 s\big\langle D^2\Phi\big(sx+(1-s)T(x)\big)\big(x-T(x)\big),x-T(x)\big\rangle\,ds\Big]g(x)\,\mu(dx)$$

$$=\int_{\mathbb{R}^n}\Big\langle\Lambda[e^{-\Phi},T(x),x]\big(T(x)-x\big),T(x)-x\Big\rangle g(x)\,\mu(dx)$$

$$=\int_{\mathbb{R}^n}\Big\langle\Lambda[e^{-\Phi},T_f,T_g]\big(T_f-T_g\big),T_f-T_g\Big\rangle\,d\mu.$$

By using the established equalities, after a change of variables we arrive at the final relationship

$$\int_{\mathbb{R}^n}(f\log f-g\log g)\,d\mu=\int_{\mathbb{R}^n}\Big\langle\nabla g(T_g),T_f-T_g\Big\rangle\frac{1}{g(T_g)}\,d\mu$$

$$+\int_{\mathbb{R}^n}\Big\langle\Lambda[e^{-\Phi},T_f,T_g]\big(T_f-T_g\big),T_f-T_g\Big\rangle\,d\mu+\int_{\mathbb{R}^n}(\mathrm{Tr}DT-n-\log\det DT)g\,d\mu.$$

As shown above,

$$\int_{\mathbb{R}^n}(\mathrm{Tr}DT-n-\log\det DT)g\,d\mu=-\int_{\mathbb{R}^n}\log\det_2\Big[DT_f(DT_g)^{-1}\Big]\,d\mu.$$

Since for the triangular canonical mappings one has $0<\det_2\Big[DT_f(DT_g)^{-1}\Big]\le 1$, the integral on the left-hand side of this equality is nonnegative. The proposition is proven. $\square$

Theorem 4.4. *Let μ be a probability measure on $\mathbb{R}^n$ with a density of the form $e^{-\Phi}$, where Φ is a smooth convex function with a nondegenerate second derivative. Let ψ be a smooth positive function such that $\psi^2\cdot\mu$ is a probability measure, and let T be the canonical triangular mapping transforming $\psi^2\cdot\mu$ to μ. Then one has*

$$\mathrm{Ent}_\mu(\psi^2)\le\int_{\mathbb{R}^n}\Big\langle\Lambda[e^{-\Phi},T(x),x]^{-1}\big(\nabla\psi(x)\big),\nabla\psi(x)\Big\rangle\,\mu(dx). \tag{4.3}$$

Proof. Let $g:=\psi^2$. Applying inequality (4.2), we obtain

$$\int_{\mathbb{R}^n}g\log g\,d\mu\le-\int_{\mathbb{R}^n}\big\langle\nabla g(x),T(x)-x\big\rangle\,\mu(dx)$$

$$-\int_{\mathbb{R}^n}\Big\langle\Lambda[e^{-\Phi},T(x),x]\big(T(x)-x\big),T(x)-x\Big\rangle g(x)\,\mu(dx)$$

$$=-\int_{\mathbb{R}^n}\Big|\Lambda[e^{-\Phi},T(x),x]^{1/2}\big(T(x)-x\big)+\frac{1}{2}\Lambda[e^{-\Phi},T(x),x]^{-1/2}\Big(\frac{\nabla g(x)}{g(x)}\Big)\Big|^2 g(x)\,\mu(dx)$$

$$+\frac{1}{4}\int_{\mathbb{R}^n}\Big\langle\Lambda[e^{-\Phi},T,x]^{-1}\big(\nabla g(x)\big),\frac{\nabla g(x)}{g(x)}\Big\rangle\,\mu(dx)$$

$$\le\frac{1}{4}\int_{\mathbb{R}^n}\Big\langle\Lambda[e^{-\Phi},T(x),x]^{-1}\big(\nabla g(x)\big),\frac{\nabla g(x)}{g(x)}\Big\rangle\,\mu(dx).$$

Returning to ψ we arrive at estimate (4.3). $\square$

Corollary 4.5. *Let $\mu=\exp\Big[-V\big(\sum_{i=1}^n p_i(x_i)\big)\Big]$ be a probability measure on $\mathbb{R}^n$, where V and p_i are smooth even convex functions on the real line, $p_i\ge 0$, and V, p_i and p_i'' are increasing on the*

half-line. Let ψ be a smooth function on $\mathbb{R}^n$ that is invariant with respect to all transformations of the form $(x_1, \cdots, x_n) \mapsto (\pm x_1, \cdots, \pm x_n)$, and let $\int \psi^2 \, d\mu = 1$. Then

$$\mathrm{Ent}_\mu(\psi^2) \leq \int_{\mathbb{R}^n} \left\langle A\nabla\psi, \nabla\psi \right\rangle d\mu,$$

where $A(x)$ is the diagonal matrix with the numbers $\left[\int_0^1 sV'\left(\sum_{i=1}^n p_i(sx_i)\right)p_k''(sx_k) \, ds\right]^{-1}$, $k = 1, \ldots, n$, at the diagonal.

Proof. Let us consider the probability measure $\psi^2 \cdot \mu$ and let $\Phi = V\left(\sum_{i=1}^n p_i(x_i)\right)$. We observe that

$$\frac{\partial^2 \Phi}{\partial x_i \partial x_j} = V''\left(\sum_{k=1}^n p_k(x_k)\right)\frac{\partial p_i}{\partial x_i}\frac{\partial p_j}{\partial x_j} + V'\left(\sum_{k=1}^n p_k(x_k)\right)p_i''(x_i)\delta_{ij}.$$

It is clear that

$$D^2\Phi \geq V'\left(\sum_{i=1}^n p_i(x_i)\right)\mathrm{diag}\{p_1''(x_1), \cdots, p_n''(x_n)\}.$$

The invariance of ψ with respect to the transformations of the indicated form yields that the triangular mapping $T = \left(T_1(x_1), T_2(x_1, x_2), \ldots, T_n(x_1, \ldots, x_n)\right)$ that transforms the measure $\psi^2 \cdot \mu$ to the measure μ has the following property: $T_i(x_1, \cdots, x_{i-1}, -x_i) = -T_i(x_1, \ldots, x_{i-1}, x_i)$. Since the functions p_i and p_i'' increase on the half-line, we obtain

$$p_i\left((1-s)T_i(x) + sx_i\right) \geq p_i(sx_i), \quad p_i''\left((1-s)T_i(x) + sx_i\right) \geq p_i''(sx_i).$$

Thus, $\dfrac{\partial^2 \Phi}{\partial x_i \partial x_j}\left((1-s)T(x) + sx\right) \geq V'\left(\sum_{i=1}^n p_i(sx_i)\right)p_i''(sx_i)\delta_{ij}$. Since $A(x)$ is the diagonal matrix with the numbers $\left[\int_0^1 sV'\left(\sum_{i=1}^n p_i(sx_i)\right)p_k''(sx_k) \, ds\right]^{-1}$ at the diagonal, one has

$$\int_0^1 sD^2\left[V\left(\sum_{i=1}^n p_i(x_i)\right)\right]\left((1-s)T(x) + sx\right) ds \geq A(x)^{-1}.$$

Therefore, $\Lambda\left[\exp\left(-V\left(\sum_{i=1}^n p_i(x_i)\right)\right), T(x), x\right]^{-1} \leq A(x)$, and our assertion follows by the preceding theorem. $\square$

Some other results close to this work are obtained in [5], [6], where the infinite dimensional case is considered, in particular, transformations of uniformly convex and Gaussian measures on infinite dimensional spaces, and in [17], where gradient type mappings are studied.

References

[1] Aleksandrova D.E. *Convergence of triangular transformations of measures.* Theory Probab. and Appl. (to appear)

[2] Bobkov S.G. *Large deviations via transference plans.* Advances in Math. Research. V. 2. P. 151–175. Nova Sci. Publ., New York, 2003.

[3] Bogachev V.I. Gaussian measures. Amer. Math. Soc., Providence, Rhode Island, 1998.

[4] Bogachev V.I. Foundations of measure theory, v. 1,2. Regular and Chaotic Dynamics, Moscow, 2003.

[5] Bogachev V.I., Kolesnikov A.V. *Nonlinear transformations of convex measures and entropy of the Radon–Nikodym densities.* Dokl. Russian Acad. Sci. 2004. V. 397, N 2. P. 155–159 (in Russian); English transl.: Dokl. Math. 2004. V. 70, N 1. P. 524–528.

[6] Bogachev V.I., Kolesnikov A.V. *On nonlinear transformations of convex measures.* BiBoS Preprint 03-12-138, Universität Bielefeld, 14 pp.

[7] Bogachev V.I., Kolesnikov A.V., Medvedev K.V. *On triangular transformations of measures.* Dokl. Russian Acad. Sci. 2004. V. 396, N 6. P. 727–732 (in Russian); English transl.: Dokl. Math. 2004. V. 69, N 3. P. 438–442.

[8] Brenier Y. *Polar factorization and monotone rearrangement of vector valued functions*. Comm. Pure Appl. Math. 1991. V. 44. P. 375–417.

[9] Cordero-Erausquin D. *Some applications of mass transport to Gaussian-type inequalities*. Arch. Rat. Mech. Anal. 2002. V. 161. P. 257–269.

[10] Fernique X. *Extension du théorème de Cameron–Martin aux translations aléatoires*. Ann. Probab. 2003. V. 31, N 3. P. 1296–1304.

[11] Feyel D., Üstünel A.S. *Transport of measures on Wiener space and the Girsanov theorem*. C. R. Acad. Sci. Paris. 2002. T. 334, N 1. P. 1025–1028.

[12] Feyel D., Üstünel A.S. *Monge–Kantorovitch measure transportation and Monge–Ampère equation on Wiener space*. Probab. Theor. Relat. Fields. 2004. V. 128, N 3. P. 347–385.

[13] Hajłasz P. *Change of variables formula under minimal assumptions*. Colloq. Math. 1993. V. 64, N 1. P. 93–101.

[14] Kechris A. Classical descriptive set theory. Springer, Berlin – New York, 1995.

[15] Knote H. *Contributions to the theory of convex bodies*. Michigan Math. J. 1957. V. 4. P. 39–52.

[16] Kolesnikov A.V. *Convexity inequalities and nonlinear transformations of measures*. Dokl. Russian Acad. Sci. 2004. V. 396, N 3. P. 300–304 (in Russian); English transl.: Dokl. Math. 2004. V. 69, N 3. P. 368–372.

[17] Kolesnikov A.V. *Convexity inequalities and optimal transport of infinite dimensional measures*. J. Math. Pures Appl. 2004 (to appear).

[18] McCann R.J. *Existence and uniqueness of monotone measure-preserving maps*. Duke Math. J. 1995. V. 80. P. 309–323.

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

[20] Rachev S.T., Rüschendorf L. Mass transportation problems. Vol. I. Springer-Verlag, New York, 1998.

[21] Talagrand M. *Transportation cost for Gaussian and other product measures*. Geom. Funct. Anal. 1996. V. 6. P. 587–600.

[22] Üstünel A.S., Zakai M. Transformation of measure on Wiener space. Springer, Berlin, 2000.

[23] Villani C. Topics in optimal transportation. Amer. Math. Soc., Rhode Island, Providence, 2003.

Dept. Mechanics and Mathematics, Moscow State University, 119992 Moscow, Russia