Theory of Stochastic Processes
Vol. 21 (37), no. 2, 2016, pp. 84–90

G. V. RIABOV

A REPRESENTATION FOR THE KANTOROVICH–RUBINSTEIN DISTANCE DEFINED BY THE CAMERON–MARTIN NORM OF A GAUSSIAN MEASURE ON A BANACH SPACE

A representation for the Kantorovich–Rubinstein distance between probability measures on a separable Banach space X in the case when this distance is defined by the Cameron–Martin norm of a centered Gaussian measure μ on X is obtained in terms of the extended stochastic integral (or divergence) operator.

1. Introduction

Consider a separable Banach space $(X, \|\cdot\|)$ equipped with a centered Gaussian measure μ on the Borel $\sigma-$field of X. We will assume that supp $\mu = X$. Let $(H, |\cdot|_H)$ be the Cameron–Martin space of μ, i.e. the separable Hilbert space densely and continuously embedded in X and such that

$$\int_X \exp(il(x))\mu(dx) = \exp\Big(-\frac{1}{2}|l|^2_H\Big),\ l \in X^*.$$

Because of continuous embedding of H into X, a functional $l \in X^*$ can be considered as a continuous linear functional on H. In the latter expression $|l|_H$ denotes the norm of l as an element of H^*.

The space $\mathcal{M}(X)$ of Borel probability measures on X is endowed with the Kantorovich-Rubinstein distance [1, §1.2]

$$W_1(\nu_0,\nu_1) = \inf_{\pi\in C(\nu_0,\nu_1)} \int_X\int_X |x_1 - x_0|_H \pi(dx_0, dx_1),$$

where $C(\nu_0, \nu_1)$ is the set of all Borel probability measures on $X \times X$ with marginals ν_0 and ν_1.

The aim of the present paper is to establish the following representation for W_1.

Theorem 1.1. *Consider probability measures $\nu_0, \nu_1 \in \mathcal{M}(X)$ with $\nu_1 - \nu_0 \ll \mu$ and $\frac{d(\nu_1-\nu_0)}{d\mu} \in L^2(X,\mu)$. Then*

$$(1)\qquad W_1(\nu_0,\nu_1) = \inf_{Iu=\frac{d(\nu_1-\nu_0)}{d\mu}} \Big\{\int_X |u(x)|_H \mu(dx)\Big\}.$$

Here I denotes the extended stochastic integral (or the divergence operator, see the next section for precise definitions). We consider the action of I on square integrable $H-$valued vector fields $u : X \to H$ only. Accordingly, the infimum is taken over all $u \in L^2(X,\mu;H)$ that solve the equation

$$(2)\qquad Iu = \frac{d(\nu_1-\nu_0)}{d\mu}.$$

2000 *Mathematics Subject Classification.* Primary 60G15; Secondary 60H07.

Key words and phrases. Gaussian measure, extended stochastic integral, optimal transport.

The author is grateful to Prof. A. A. Dorogovtsev and A. M. Kulik for valuable discussions and comments. The research is partially supported by the Young Scientists Grant of the National Academy of Sciences of Ukraine.

This work was partially motivated by results of [2] where several integral representations for square integrable functions on (X,μ) were derived. Namely, for every function $\alpha \in L^2(X,\mu)$ the equation

$$\alpha = \int_X \alpha d\mu + Iu \tag{3}$$

has infinitely many solutions $u \in L^2(X,\mu;H)$. In the case of the classical Wiener space one particular solution is distinguished. Namely, if $X = C_0([0,1])$ is the space of continuous functions $f : [0,1] \to \mathbb{R}$, $f(0) = 0$ and μ is the Wiener measure, then there is a unique solution u_0 of (3) which is adapted to the natural filtration on $C_0([0,1])$ [3, Ch. V, §3]. When α is the probability density, i.e. $\alpha \geq 0$ and $\int_X \alpha d\mu = 1$, the representation

$$\alpha = 1 + Iu_0$$

is connected to the measure transportation via the Girsanov theorem [3, Ch. VIII, §1]. The mapping $T : C_0([0,1]) \to C_0([0,1])$ defined by

$$T(x)(t) = x(t) - \int_0^t \frac{u_0(s,x)}{1 + I(u_0 1_{\cdot \leq s})(x)} ds$$

sends the measure $\alpha \cdot \mu$ into the Wiener measure μ :

$$(\alpha \cdot \mu) \circ T^{-1} = \mu.$$

Moreover, the mapping T is in a sense optimal [4, 5]. For every mapping $S : X \to X$ such that $S(x) - x \in H$ and $(\alpha \cdot \mu) \circ S^{-1} = \mu$ one has

$$\int_X |T(x) - x|_H^2 \mu(dx) \leq \int_X |S(x) - x|_H^2 \mu(dx).$$

In the general situation there is still a connection between the transportation of measure and the equation (3). An estimate on the Kantorovich–Rubinstein distance in terms of solutions of (3) was obtained in [6]. It was proved that for a sufficiently smooth density α one has

$$W_1(\alpha \cdot \mu, \mu) \leq \int_X |(1+L)^{-1} D\alpha|_H d\mu, \tag{4}$$

where D denotes the stochastic derivative and $(-L)$ is the generator of the Ornstein-Uhlenbeck semigroup. Our result (1) generalizes this inequality. Indeed, the identity [7, Prop. 1.3.1, 1.4.3]

$$DI = 1 + L$$

implies that $(1+L)^{-1} D\alpha$ is a solution to (3). The existence of a solution to the Monge problem associated with W_1 was proved in [8] under assumptions $\nu_0, \nu_1 \ll \mu$, $W_1(\nu_0,\nu_1) < \infty$. Namely, it was proved that a mapping $T : X \to X$ such that

$$\nu_0 \circ T^{-1} = \nu_1, \ W_1(\nu_0,\nu_1) = \int_X |x - T(x)|_H \nu_0(dx)$$

exists. For several other estimates of the quantity $W_1(\nu_0,\nu_1)$ as well as for the extensive treatment of the optimal transport theory we refer to [9].

Another motivation for the undertaken research is the study of geodesics on the space $(\mathcal{M}(X), W_1)$ [11, Ch. 7]. In the case $p > 1$ the differential structure of the space $(\mathcal{M}(X), W_p)$ is studied rather detaily [12, 13, 14, 15]. The assumption $p > 1$ allows to apply powerful technique from convex analysis. In the limit $p \to 1+$ certain results about geodesics in $(\mathcal{M}(X), W_1)$ can be obtained [14]. However, the distance W_1 is not strictly convex. This results in existence of multiple geodesics between different measures while the described approximating approach gives results only for particular W_1-geodesics. In general, the behaviour of geodesics in the space $(\mathcal{M}(X), W_1)$ remains unstudied. Proved identity (1) gives an intrinsic description of the W_1-distance between measures. In our further work it will be applied to the study of W_1-geodesics.

2. Notations and Preliminary Results

For a detailed exposition of the theory of Gaussian measures on Banach spaces we refer to [10].

A function $f : X \to \mathbb{R}$ will be called a smooth cylindrical function, if it has a representation

$$f(x) = \varphi(l_1(x), \ldots, l_d(x)),\ x \in X,$$

where $l_1, \ldots, l_d \in X^*$ and $\varphi \in C^\infty(\mathbb{R}^d)$ is bounded together with all derivatives. The family of all smooth cylindrical functions will be denoted by $\mathcal{FC}^\infty$.

Stochastic derivative D is naturally defined for a fuction $f \in \mathcal{FC}^\infty$ with a representation $f(x) = \varphi(l_1(x), \ldots, l_d(x))$:

$$Df(x) = \sum_{i=1}^d \partial_i \varphi(l_1(x), \ldots, l_d(x)) l_i \in H.$$

Then D is extended to a closed (unbounded) operator

$$D : L^2(X, \mu) \to L^2(X, \mu; H).$$

Functions in the domain of D are called stochastically differentiable. Denote by I the adjoint operator to D,

$$I = D^*.$$

According to such definition we consider the action of I on elements $u \in L^2(X, \mu; H)$ exceptionally.

Following [16] we will call I the extended stochastic integral. In terms of the integration by parts formula one has the characterization (see [10, §5.8] for equivalent definitions of the operator I):

for every stochastically differentiable $f \in L^2(X, \mu)$

$$\int_X (u, Df)_H d\mu = \int_X Iu \cdot f d\mu.$$

Remark 2.1. In [10, 7] the operator $(-I)$ is denoted by δ and is called a divergence operator. The term "extended stochastic integral" is kept for a specific situation when H is an L^2-space. Our terminology is chosen to underline the connection between the operator I and integral representations of random variables (3).

The Ornstein-Uhlenbeck semigroup is denoted by $(T_t)_{t\geq 0}$:

$$T_t h(x) = \int_X h(e^{-t}x + \sqrt{1 - e^{-2t}} y) \mu(dy).$$

For each $p \geq 1$ $(T_t)_{t\geq 0}$ is a strongly continuous semigroup of contractions in $L^p(X, \mu)$ [10]. We will also consider the action of T_t on measures. Given a signed measure ν on X define

$$T_t\nu(A) = \int_X T_t 1_A(x) \nu(dx) = \int_X \int_X 1_A(e^{-t}x + \sqrt{1 - e^{-2t}} y) \mu(dy) \nu(dx).$$

Duality considerations imply that T_t is still a contraction:

$$\|T_t \nu\|_v \leq \|\nu\|_v,$$

where $\| \cdot \|_v$ denotes the total variation norm.

Among integral representations (3) of a random variable α there is a unique representation with a minimal $L^2(X, \mu; H)$ norm [2]. In the next lemma all the needed properties of this representation are gathered.

Lemma 2.1. [2, L. 6,7]. *Define the mapping*

$$v(\alpha) = D\int_0^\infty T_t\alpha dt,\ \ \alpha \in L^2(X,\mu).$$

Then

- $v : L^2(X,\mu) \to L^2(X,\mu;H)$ *is a bounded linear operator of the norm* 1;
- *for every* $\alpha \in L^2(X,\mu)$ $v(\alpha)$ *is a solution to* (3)*:*

$$\alpha = \int_X \alpha d\mu + Iv(\alpha);$$

- *for any solution* u *to* (3) *one has*

$$\int_X |v(\alpha)|_H^2 d\mu \le \int_X |u|_H^2 d\mu.$$

3. Proof of the Theorem 1.1

In this section the proof of the equality (1) is presented.

Proof. For convenience we divide the proof into three steps.

Step 1. The inequality $\le$ *.*

The well-known Kantorovich–Rubinstein theorem [1, Th. 1.14] states that

$$W_1(\nu_0,\nu_1) = \sup\left\{\int_X fd(\nu_1-\nu_0)\right\}, \tag{5}$$

where the supremum is taken over all bounded measurable functions $f : X \to \mathbb{R}$ that satisfy the condition

$$|f(x+h) - f(x)| \le |h|_H,\ \ x \in X, h \in H$$

(we will call such functions 1−Lipschitzian along H [10, §4.5]).

Hence, to prove $\le$ in the representation (1) it is enough to check the inequality

$$\left|\int_X fd\nu_1 - \int_X fd\nu_0\right| \le \int_X |u|_H d\mu, \tag{6}$$

where $f : X \to \mathbb{R}$ is a bounded measurable 1−Lipschitzian function along H and $u \in L^2(X,\mu;H)$ satisfies (2):

$$Iu = \frac{d(\nu_1-\nu_0)}{d\mu}.$$

According to [10, Th. 5.11.2, Cor. 5.4.7] the function f is stochastically differentiable with $|Df|_H \le 1$. This implies the following chain of inequalities.

$$\left|\int_X fd\nu_1 - \int_X fd\nu_0\right| = \left|\int_X \frac{d(\nu_1-\nu_0)}{d\mu}\cdot fd\mu\right| = \left|\int_X Iu\cdot fd\mu\right| =$$
$$= \left|\int_X (Df,u)_H d\mu\right| \le \int_X |Df|_H|u|_H d\mu \le \int_X |u|_H d\mu.$$

The relation (6) together with the inequality $\le$ in (1) are proved.

The relation (5) implies that the distance $W_1(\nu_0,\nu_1)$ depends only on the density $\rho = \frac{d(\nu_1-\nu_0)}{d\mu} \in L^2(X,\mu)$. Thus we will denote the quantity $W_1(\nu_0,\nu_1)$ by $W_1(\rho)$ as well and will consider it as a function on the set $L_0^2(X,\mu) = \{\rho \in L^2(X,\mu) : \int_X \rho d\mu = 0\}$. Accordingly, let us denote the right-hand side of (1) by $\mathcal{N}(\rho)$, i.e.

$$\mathcal{N}(\rho) = \inf_{Iu=\rho}\int_X |u|_H d\mu.$$

Our strategy of proving equality in (1) is to check continuity of W_1 and $\mathcal{N}$ on $L_0^2(X,\mu)$ and to prove the equality for a dense set of functions $\rho \in L_0^2(X,\mu)$.

Step 2. Continuity of functionals W_1 and $\mathcal{N}$.

The supremum in (5) can be reduced to the set of all bounded measurable 1-Lipschitzian functions along H with zero integral

$$\int_X f d\mu = 0.$$

Then the concentration inequality [10, Th. 4.5.7]

$$\mu(|f| > r) \leq 2e^{-\frac{r^2}{2}}, \ r > 0$$

implies the bound

$$\int_X f^2 d\mu \leq 4.$$

Hence for arbitrary $\rho_1, \rho_2 \in L_0^2(X, \mu)$ one has

$$\left| \int_X f\rho_1 d\mu - \int_X f\rho_2 d\mu \right| \leq 2\sqrt{\int_X (\rho_1 - \rho_2)^2 d\mu}.$$

From the representation (5) one has the estimate

$$|W_1(\rho_1) - W_1(\rho_2)| \leq 2\sqrt{\int_X (\rho_1 - \rho_2)^2 d\mu}.$$

For the functional $\mathcal{N}$ similar estimate follows from the existence of the minimal norm representation operator v (see lemma 2.1). Indeed, consider $\rho_1, \rho_2 \in L_0^2(X, \mu)$. For each solution u of $Iu = \rho_1$ one has

$$I(u + v(\rho_2 - \rho_1)) = \rho_2.$$

By the properties of v,

$$\mathcal{N}(\rho_2) \leq \int_X |u + v(\rho_2 - \rho_1)|_H d\mu \leq \int_X |u|_H d\mu + \sqrt{\int_X |v(\rho_2 - \rho_1)|_H^2 d\mu}$$
$$\leq \int_X |u|_H d\mu + \sqrt{\int_X (\rho_2 - \rho_1)^2 d\mu}.$$

Taking infimum in u and repeating the argument we get the inequality

$$|\mathcal{N}(\rho_1) - \mathcal{N}(\rho_2)| \leq \sqrt{\int_X (\rho_1 - \rho_2)^2 d\mu}.$$

It remains to prove the inequality $\geq$ in (1) for a dense family of functions $\rho \in L_0^2(X, \mu)$.

Step 3. The inequality $\geq$.

Let $\{e_n\}$ be the orthonormal basis in H and $\{\hat{e}_n\}$ be the corresponding measurable linear functionals on X. Denote by γ_n the standard Gaussian measure on $\mathbb{R}^n$. Functions of the form

$$\rho(x) = \varkappa(\hat{e}_1(x), \ldots, \hat{e}_n(x)), \ \varkappa \in L_0^2(\mathbb{R}^n, \gamma_n) \tag{7}$$

form a dense set in $L_0^2(X, \mu)$. We will finish the proof by establishing the inequality $\geq$ in (1) for a function ρ of the form (7).

In [17, Proof of Prop. 4.1] the following consequence of the Riesz–Markov–Kakutani representation theorem is derived: there exists an $\mathbb{R}^n-$valued Borel measure π on $\mathbb{R}^n$ such that

(1) for all $f \in \mathcal{FC}^\infty$ one has

$$\int_{\mathbb{R}^n} (Df, d\pi) = \int_{\mathbb{R}^n} f\varkappa d\gamma_n;$$

(2) $W_1(\rho)$ coincides with the total variation of the measure π :

$$W_1(\rho) = \|\pi\|_v = \left(\sum_{i=1}^{n} \|\pi_i\|_v^2\right)^{\frac{1}{2}}.$$

Symmetry of the Ornstein-Uhlenbeck semigroup implies following relations.

$$\int_{\mathbb{R}^n} f(T_t\varkappa)d\gamma_n = \int_{\mathbb{R}^n} (T_t f)\varkappa d\gamma_n = \int_{\mathbb{R}^n} (D(T_t f), d\pi)$$

$$= e^{-t}\int_{\mathbb{R}^n} (T_t Df, d\pi) = \int_{\mathbb{R}^n} \left(Df, e^{-t}\frac{dT_t\pi}{d\gamma_n}\right)d\gamma_n.$$

In other words, the function

$$u(x) = e^{-t}\sum_{j=1}^{n} \left(\frac{dT_t\pi_j}{d\gamma_n}(\hat{e}_1(x), \ldots, \hat{e}_n(x))\right)e_j$$

is a solution to the equation

$$Iu = T_t\rho.$$

In particular,

$$\mathcal{N}(T_t\rho) \le \int_X |u|_H d\mu = e^{-t}\int_{\mathbb{R}^n} \left|\frac{dT_t\pi}{d\gamma_n}\right| d\gamma_n = e^{-t}\|T_t\pi\|_v \le \|\pi\|_v = W_1(\rho).$$

Taking the limit $t \to 0+$ we obtain the inequality $\ge$ in (1). The theorem is proved. □

Acknowledgement. The author is grateful to the referee for suggesting a better scheme of the proof and improving the overall presentation.

References

1. C. Villani, *Topics in Optimal Transportation*, American Mathematical Society, Providence, Rhode Island, 2003.
2. A. A. Dorogovtsev, O. L. Izyumtseva, G. V. Riabov and N. Salhi, *Clark formula for local time for one class of Gaussian processes*, Communications on Stochastic Analysis **10** (2016), no. 2, 195–217.
3. D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, Springer–Verlag, Berlin, 1999.
4. J. Lehec, *Representation formula for the entropy and functional inequalities*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques **49** (2013), no. 3, 885–899.
5. H. Föllmer, *An entropy approach to the time reversal of diffusion processes*, Stochastic Differential Systems Filtering and Control (M. Metivier, E. Pardoux, ed.), Springer, Berlin, pp. 156–163.
6. D. Feyel and A. S. Üstünel, *Monge-Kantorovitch measure transportation and Monge-Ampere equation on Wiener space*, Probability theory and related fields **128** (2004), no. 3, 347–385.
7. D. Nualart, *The Malliavin calculus and related topics*, Springer, Berlin, 2006.
8. F. Cavalletti, *The Monge problem in Wiener space*, Calc. Var. Partial Differential Equations **45** (2012), no. 1–2, 101–124.
9. V. I. Bogachev and A. V. Kolesnikov, *The Monge–Kantorovich problem: achievements, connections, and prospects*, Russian Math. Surveys **67** (2012), no. 5, 785–890.
10. V. I. Bogachev, *Gaussian Measures*, American Mathematical Society, Providence, Rhode Island, 1998.
11. L. Ambrosio, N. Gigli and D. Savaré, *Gradient flows: in metric spaces and in the space of probability measures*, Springer Science & Business Media, Berlin, 2008.
12. M.–K. von Renesse and K.–Th. Sturm, *Transport inequalities, gradient estimates, entropy and Ricci curvature*, Communications on pure and applied mathematics **58** (2005), no. 7, 923–940.
13. Sh. Fang, J. Shao and K.–Th. Sturm, *Wasserstein space over the Wiener space*, Probability theory and related fields **146** (2010), no. 3–4, 535–565.
14. V. Nolot, *Convexities and optimal transport problems on the Wiener space*, Doctoral dissertation, Université de Bourgogne, 2013.
15. Sh. Fang and V. Nolot, *Optimal transport maps on infinite dimensional spaces*, Frontiers of Mathematics in China **10** (2015), no. 4, 715–732.

16. A. V. Skorokhod, *On a generalization of a stochastic integral*, Theory of Probability & Its Applications **20** (1976), no. 2, 219–233.
17. L. Ambrosio, *Lecture notes on optimal transport problems*, Springer, Berlin, 2003.

01004, Ukraine, Kiev–4, 3, Tereschenkivska st.
E-mail address: ryabov.george@gmail.com, ryabov.george@imath.kiev.ua