ISSN 1064-5624, Doklady Mathematics, 2016, Vol. 93, No. 2, pp. 135–139. © Pleiades Publishing, Ltd., 2016.
Published in Russian in Doklady Akademii Nauk, 2016, Vol. 467, No. 1, pp. 13–17.

Estimates of Distances between Transition Probabilities of Diffusions[1]

V. I. Bogachev[a, b], M. Röckner[c], and S. V. Shaposhnikov[a, b]

Presented by Academician I.A. Ibragimov September 17, 2015

Received October 6, 2015

Abstract—We obtain upper bounds for the total variation and entropy distances between probability solutions to two Fokker–Planck–Kolmogorov equations with different diffusion matrices and drifts

DOI: 10.1134/S1064562416020034

The goal of this paper is to give upper bounds for the total variation and entropy distances between probability solutions $\varrho_1(x, t)$ and $\varrho_2(x, t)$ to two Fokker–Planck–Kolmogorov equations

$$\partial_t \varrho_k(x, t) = \partial_{x_i}\partial_{x_j}(a_k^{ij}(x, t)\varrho_k(x, t)) - \partial_{x_i}(b_k^i(x, t)\varrho_k(x, t)), \quad k = 1, 2, \tag{1}$$

with different diffusion matrices and drifts on $\mathbb{R}^d \times [0, T]$ with fixed $T > 0$. In case of equal initial distributions and identity diffusion matrices, for the entropy of ϱ_2 with respect to ϱ_1 we obtain the estimate

$$\int_{\mathbb{R}^d} \ln\frac{\varrho_2(x, t)}{\varrho_1(x, t)}\varrho_2(x, t)dx \le \frac{1}{2}\int_{\mathbb{R}^d} |b_1(x, t) - b_2(x, t)|^2\varrho_2(x, t)dx,$$

and for the total variation norm we obtain the estimate

$$\|\varrho_1(\cdot, t) - \varrho_2(\cdot, t)\|_{TV}^2 \le \int_0^t\int_{\mathbb{R}^d} |b_1(x, s) - b_2(x, s)|^2\varrho_2(x, s)dxds.$$

In the general case we obtain similar estimates under rather broad assumptions about our coefficients. The principal novelty concerns the case of different diffusion matrices (see Remark 2), but also the simpler case of the same diffusion matrix is new. The main result is applied to nonlinear Fokker–Planck–Kolmogorov equations.

Let us consider a time-dependent second order elliptic operator

$$L_{A, b}u = \sum_{i, j = 1}^d a^{ij}\partial_{x_i}\partial_{x_j}u + \sum_{i = 1}^d b^i\partial_{x_i}u,$$

where $A(x, t) = (a_{ij}(x, t))_{i, j \le d}$ is a positive symmetric matrix (called the diffusion matrix) with Borel measurable entries and $b(x, t) = (b^i(x, t))_{i=1}^d : \mathbb{R}^d \times [0, T] \to \mathbb{R}^d$ is a Borel measurable mapping (called the drift coefficient). Suppose that b is locally bounded and A is locally Lipschitzian in x and locally strictly positive, i.e.,

(H) for every ball $U \subset \mathbb{R}^d$, there exist numbers $\lambda = \lambda(U) \ge 0$, $\alpha = \alpha(U) > 0$ and $m = m(U) > 0$ such that

$$|a^{ij}(x, t) - a^{ij}(y, t)| \le \lambda|x - y|,$$
$$\alpha \cdot I \le A(x, t) \le m \cdot I$$

for all $x, y \in U$ and $t \in [0, T]$.

We study solutions to the Cauchy problem

$$\partial_t\mu = L_{A, b}^*\mu, \quad \mu|_{t=0} = \nu, \tag{2}$$

where ν is a Borel probability measure on $\mathbb{R}^d$. A model example is given by the transition probabilities of a diffusion process. We shall consider measures $\mu(dxdt) = \mu_t(dx)dt$ on $\mathbb{R}^d \times [0, T]$ given by families of probability measures $(\mu_t)_{t \in [0, T]}$ (or with $t \in (0, T)$, which does not matter for our purposes) on $\mathbb{R}^d$, i.e., $t \mapsto \mu_t(B)$ is measurable for every Borel set $B \subset \mathbb{R}^d$ and

[1] The article was translated by the authors.

[a] Department of Mechanics and Mathematics, Moscow State University, Moscow, 119991 Russia
[b] St. Tikhon's Orthodox University, Moscow, Russia
[c] Fakultät für Mathematik, Universität Bielefeld, D-33501 Bielefeld, Germany
e-mail: vibogach@mail.ru, starticle@mail.ru

$$\int_{\mathbb{R}^d \times [0,T]} f(x,t)\mu(dxdt) = \int_0^t \int_{\mathbb{R}^d} f(x,t)\mu_t(dx)dt$$

for every bounded Borel function f on $\mathbb{R}^d \times [0, T]$. Such a measure is called a solution to the Cauchy problem (2) if, for every function φ of class $C_0^\infty(\mathbb{R}^d)$, the equality

$$\int_{\mathbb{R}^d} \varphi d\mu_t = \int_{\mathbb{R}^d} \varphi d\nu + \int_0^t \int_{\mathbb{R}^d} L_{A,b}\varphi d\mu_s ds \quad (3)$$

holds for almost all $t \in (0, T)$.

It is known (see [1]) that the measure μ possesses a continuous positive density ϱ on $\mathbb{R}^d \times (0, T)$ with respect to Lebesgue measure, moreover, for each ball U in $\mathbb{R}^d$, for almost every $t \in (0, T)$ one has $\varrho(\cdot, t) \in W^{p,1}(U)$ for all $p \in [1, +\infty)$ and the function $\|\varrho(\cdot, t)\|^p_{L^p(U)} + \|\nabla\varrho(\cdot, t)\|^p_{L^p(U)}$ is integrable on every compact interval in $(0, T)$. Recall that $W^{p,1}(U)$ consists of all functions that belong to $L^p(U)$ along with their first order Sobolev derivatives. We shall deal with this version of ϱ (in this case $\varrho(\cdot, t)$ is a probability density for almost every t and is integrable for all $t \in (0, T)$. Such a version satisfies the classical equation (1) understood in the weak sense.

Suppose now that $\mu = (\mu_t)_{t \in (0,T)}$ and $\sigma = (\sigma_t)_{t \in (0,T)}$ are two solutions to the Cauchy problem (2) with coefficients A_1, b_1 and A_2, b_2, respectively, and the same initial condition ν (the case of different initial conditions is considered similarly). The corresponding operators will be denoted by L_1 and L_2 for brevity. Suppose throughout that A_1 and A_2 satisfy Condition (H) and b_1 and b_2 are locally bounded Borel measurable.

Let $\mu = \varrho_\mu(x, t)dxdt$ and $\sigma = \varrho_\sigma(x, t)dxdt$. Set

$$v(x,t) = \frac{\varrho_\sigma(x,t)}{\varrho_\mu(x,t)}, \quad \text{i.e. } \sigma = v \cdot \mu.$$

Let us introduce vector mappings

$$h_1 = (h_1^i)_{i=1}^d, \quad h_2 = (h_2^i)_{i=1}^d,$$

$$h_1^i = b_1^i - \sum_{j=1}^d \partial_{x_j} a_1^{ij}, \quad h_2^i = b_2^i - \sum_{j=1}^d \partial_{x_j} a_2^{ij},$$

$$\Phi = \frac{(A_1 - A_2)\nabla\varrho_\sigma}{\varrho_\sigma} - (h_1 - h_2).$$

The distances between μ_t and σ_t will be estimated through the $L^2(\sigma)$-norm of $A_1^{-1/2}\Phi$. In case of equal diffusion matrices we obtain the difference of the drifts: $\Phi = b_1 - b_2$. In case of equal drifts and constant diffusion matrices, only the first term of Φ appears.

The total variation norm $\|\cdot\|_{TV}$ of a measure with a density equals the L^1-norm of the density. Given two probability measures μ_1 and μ_2 on $\mathbb{R}^d$ such that $\mu_2 = w \cdot \mu_1$, the entropy $H(\mu_2|\mu_1)$ is defined by

$$H(\mu_2|\mu_1) = \int w \ln w d\mu_1 = \int \ln w d\mu_2,$$

provided that $w \ln w \in L^1(\mu_1)$. If μ_1 and μ_2 are given by positive densities ϱ_1 and ϱ_2 such that $\varrho_2 \ln\left(\frac{\varrho_2}{\varrho_1}\right) \in L^1(\mathbb{R}^d)$, then $H(\mu_2|\mu_1)$ is the integral of $\varrho_2 \ln\left(\frac{\varrho_2}{\varrho_1}\right)$. Let us formulate our main result.

Theorem 1. *Let* $|A_1^{-1/2}\Phi| \in L^2(\mathbb{R}^d \times [0, T], \sigma)$. *Suppose also that at least one of the following two conditions is fulfilled:*

(a) $(1 + |x|)^{-2}|a_1^{ij}|$, $(1 + |x|)^{-1}|b_1| \in L^1(\mathbb{R}^d \times [0, T], \sigma)$, $(1 + |x|)^{-1}|\Phi| \in L^1(\mathbb{R}^d \times [0, T], \sigma)$;

(b) *there exist a nonnegative function* $V \in C^2(\mathbb{R}^d)$ *and a number* $M \geq 0$ *such that*

$$\lim_{|x|\to\infty} V(x) = +\infty, \quad L_{A_1, b_1}V \leq MV,$$

$$\frac{\langle \Phi, \nabla V\rangle}{1 + V} \in L^1(\mathbb{R}^d \times [0, T], \sigma).$$

Then

$$H(\sigma_t|\mu_t) = \int_{\mathbb{R}^d} v \ln v d\mu_t \leq \frac{1}{2}\int_0^t \int_{\mathbb{R}^d} \left|A_1^{-1/2}\Phi\right|^2 d\sigma_s ds. \quad (4)$$

Corollary 1. *Under the assumptions of the theorem, for every nonnegative measurable function* φ *on* $\mathbb{R}^d \times [0, T]$, *we have*

$$\|\varphi(\mu_t - \sigma_t)\|^2_{TV} \leq (1 + \ln\alpha(t)) \times \int_0^t \int_{\mathbb{R}^d} \left|\frac{A_1^{-1/2}(A_1 - A_2)\nabla\varrho_\sigma}{\varrho_\sigma} - A_1^{-1/2}(h_1 - h_2)\right|^2 d\sigma_s ds,$$

where

$$\alpha(t) := \int_{\mathbb{R}^d} e^{\varphi^2(x,t)}\mu_t(dx).$$

In particular, if $A_1 = A_2 = A$, *then*

$$\|\varphi(\mu_t - \sigma_t)\|^2_{TV} \leq (1 + \ln\alpha(t)) \times \int_0^t \int_{\mathbb{R}^d} \left|A^{-1/2}(b_1 - b_2)\right|^2 d\sigma_s ds$$

and if $A_1 = A_2 = I$, *then*

$$\|\varphi(\mu_t - \sigma_t)\|^2_{TV} \leq (1 + \ln\alpha(t)) \int_0^t \int_{\mathbb{R}^d} |b_1 - b_2|^2 d\sigma_s ds.$$

If $b_1 = b_2 = b$ *and the matrices* A_1, A_2 *do not depend on* x, *then*

$$\|\varphi(\mu_t - \sigma_t)\|_{TV}^2 \le (1 + \ln\alpha(t)) \times \int_0^t \int_{\mathbb{R}^d} \frac{\left|(A_1^{1/2}A_2^{-1/2} - A_1^{-1/2}A_2^{1/2})A_2^{1/2}\nabla\varrho_\sigma\right|^2}{\varrho_\sigma} dxds.$$

The Kantorovich distance $W_p(\mu_1, \mu_2)$ of order $p \in [1, +\infty)$ is defined as the infimum of

$$\left(\int\int |x - y|^p \pi(dxdy)\right)^{1/p}$$

over all probability measures π on $\mathbb{R}^d \times \mathbb{R}^d$ with projections μ_1 and μ_2 on the factors. For $p = 1$ this gives the classical Kantorovich distance, see [2]. In the case $\varphi = 1$, in Corollary 1, we obtain the usual total variation distance, hence the classical Pinsker–Csiszár–Kullback inequality (see, e.g., [3, Theorem 2.12.24])

$$\|\mu - \sigma\|_{TV}^2 \le 2H(\sigma|\mu)$$

can be applied. The estimate in the theorem can be combined with the estimate

$$W_p(\mu_1, \mu_2) \le C[H(\mu_1|\mu_2)^{1/p} + 2^{-1/(2p)}H(\mu_1|\mu_2)^{1/(2p)}],$$

established in [4], where C is a number that depends on the integral of $\exp(\kappa|x|^p)$ against μ_2 for any fixed number κ (so that if we fix κ and consider only measures μ_2 such that the integral of $\exp(\kappa|x|^p)$ against μ_2 does not exceed a fixed number M, then C depends only on κ and M).

Remark 1. The theorem and the corollary involve (through Φ) the logarithmic gradient $\nabla\varrho_\sigma/\varrho_\sigma$ of the measure μ_σ (in the case where A_1 and A_2 are different). If the norms of $A_1 - A_2$ and A_1^{-1} are uniformly bounded, then, up to a constant factor, the right-hand side of (4) is estimated by the $L^2(\sigma)$-norms of $|b_1 - b_2|$, $|\nabla a_1^{ij} - \nabla a_2^{ij}|$, and $\frac{|\nabla\varrho_\sigma|}{\varrho_\sigma}$. Some estimates of the $L^2(\sigma)$-norm of $\frac{\nabla\varrho_\sigma}{\varrho_\sigma}$ were obtained in [5], so there are efficient conditions in terms of the coefficients to verify that the right-hand side of our estimate is finite. In particular, in the previous situation we arrive at the following bound:

$$\|\varphi(\mu_t - \sigma_t)\|_{TV} \le C(t) \sup_{x,t,i,j} [|b_1(x, t) - b_2(x, t)| + |a_1^{ij}(x, t) - a_2^{ij}(x, t)| + |\nabla a_1^{ij}(x, t) - \nabla a_2^{ij}(x, t)|],$$

where $C(t)$ depends also on d, λ, α, $\|b_2\|_{L^2(\sigma)}$, $\|\Lambda\|_{L^1(\sigma)}$, and $\|\ln\varrho_\nu\|_{L^1(\nu)}$.

Remark 2. Let us observe that if $d = 1$, $A = 1$, $\varphi = 1$, and there exist diffusion processes ξ_1 and ξ_2 with drifts b_1 and b_2 and initial distribution ν (which is the case, e.g., for bounded drifts), our estimates agree with the estimates obtained in [6, 7] for the total variation distance between the distributions of ξ_1 and ξ_2 in the space $C[0, T]$. However, the method of [6, 7] is based on the Girsanov theorem and does not extend to the case of different diffusion matrices, because the corresponding distributions in functional spaces can be mutually singular (as in the case $A_1 = I$ and $A_2 = 2I$).

Various estimates for transition probabilities of diffusions involving the total variation distance or Kantorovich-type distances have become popular in the last decade. There are many works on this topic, see, e.g., [8, 9]. The principal novelty of our estimates is that they compare diffusions with different drifts or even with different diffusion matrices, not just with different initial distributions.

Informally, our proof is this: we multiply the equation by $f = v\ln v - v$, integrate by parts, apply the Cauchy inequality and discard certain terms in the obtained inequality. However, a rigorous justification involves some technicalities, which will be presented in a more detailed paper. We just mention that we estimate the entropy and employ the next inequality established in [4]: given two probability measures μ and $\sigma = v \cdot \mu$ on $\mathbb{R}^d$ and a Borel function $\varphi \ge 0$, we have

$$\|\varphi(\mu - \sigma)\|_{TV}^2 \le 2\left(1 + \ln\left(\int_{\mathbb{R}^d} e^{\varphi^2} d\mu\right)\right) \int_{\mathbb{R}^d} v\ln v d\mu.$$

Let us give effective conditions to verify our assumptions (a) or (b).

Corollary 2. $A_1 = A_2 = A$ *be uniformly bounded* (*and satisfy* (H)). *Suppose that for some numbers* $\gamma_1 > 0$ *and* $\gamma_2 > 0$ *we have*

$$\langle b_1(x, t), x\rangle \le \gamma_1 + \gamma_2|x|^2.$$

Then

$$\|\mu_t - \sigma_t\|_{TV}^2 \le 2\int_0^t \int_{\mathbb{R}^d} \left|A^{-1/2}(b_1 - b_2)\right|^2 d\sigma_s ds.$$

Moreover, for any $p \ge 1$ *and* $K > 0$ *the following estimate holds*:

$$\left\|(1 + |x|^p)(\mu_t - \sigma_t)\right\|_{TV}^2 \le 2K^{-1}\left(1 + \ln\left(\int_{\mathbb{R}^d} e^{K(1 + |x|^p)^2}\mu_t(dx)\right)\right) \times \int_0^t \int_{\mathbb{R}^d} \left|A^{-1/2}(b_1 - b_2)\right|^2 d\sigma_s ds.$$

Example 1. In case $A_1 = A_2$ is uniformly bounded, condition (a) is fulfilled if $|b_1(x)| \le C + C|x|$ or if $|b_1(x)| \le C + C|x|^m$ and $|x|^{m-1}$ is integrable with respect to σ. The latter can be verified by using Lyapunov functions (see, e.g., [1, 10]). Also the assumption that $|b_1 - b_2|^2$ is σ-integrable can be verified in these terms. Certainly, the case of bounded b_1 and b_2 is covered by both conditions.

Example 2. Let L_1 be the Ornstein–Uhlenbeck operator $\Delta u(x) - \langle x, \nabla u(x)\rangle$ and let L_2 be its perturbation by a first order term generated by a bounded Borel vector field b_0 on $\mathbb{R}^d$. Then

$$\|\varphi(\mu_t - \sigma_t)\|_{TV}^2 \le (1 + \ln\alpha(t)) \int_0^t \int_{\mathbb{R}^d} |b_0|^2 d\sigma_s ds.$$

In particular, for $\varphi = 1$ we obtain that

$$\|\mu_t - \sigma_t\|_{TV}^2 \le 2 \int_0^t \int_{\mathbb{R}^d} |b_0|^2 d\sigma_s ds.$$

Corollary 3. *Let A_1 and A_2 satisfy* (H). *Suppose that there are numbers $\lambda_1, \lambda_2 > 0$ such that*

$$\lambda_1 \cdot I \le A_1(x, t) \le \lambda_2 \cdot I,$$

$$\lambda_1 \cdot I \le A_2(x, t) \le \lambda_2 \cdot I \ \textit{for all} \ \ (x, t).$$

Assume also that $|x|^m \in L^1(\nu)$, $\nu = \varrho_0 dx$, $\varrho_0 \ln \varrho_0 \in L^1(\mathbb{R}^d)$ and

$$\langle b_1(x, t), x\rangle \le \gamma_1 + \gamma_2 |x|^2,$$

$$|b_2(x, t)| \le \gamma_3 + \gamma_4 |x|^m$$

for some numbers $m, \gamma_i \ge 0$. Then

$$\|\mu_t - \sigma_t\|_{TV}^2 \le C(T) \sup_{x, t} \|A_1 - A_2\|^2$$

$$+ C(T) \int_0^t \int_{\mathbb{R}^d} \left|A_1^{-1/2}(h_1 - h_2)\right|^2 d\sigma_s ds,$$

where $h_1^i - h_2^i = b_1^i - b_2^i - \partial_{x_j}(a_1^{ij} - a_2^{ij})$ and the number $C(T)$ on the right depends on T, m, λ_i, γ_i, $\int |x|^{2m} d\nu$, $\|\varrho_0 \ln \varrho_0\|_{L^1(\mathbb{R}^d)}$.

Suppose now that, for every measure μ on $\mathbb{R}^d \times (0, T)$ given by a family $(\mu_t)_{t \in (0, T)}$ of probability measures on $\mathbb{R}^d$, we are given a locally bounded Borel measurable mapping

$$b(\mu, \cdot, \cdot): \mathbb{R}^d \times [0, T] \to \mathbb{R}^d.$$

Let us consider the Cauchy problem for the nonlinear Fokker–Planck–Kolmogorov equation

$$\partial_t \mu = \Delta\mu - \operatorname{div}(b(\mu, x, t)\mu), \quad \mu|_{t=0} = \nu. \qquad (5)$$

By a solution we mean a measure μ given by a family of probability measures $(\mu_t)_{t \in [0, T]}$ such that the integral identity (3) is fulfilled. The linear case corresponds to a drift independent of measures. To solve this nonlinear equation, we apply the fixed point principle: for each drift $b(\sigma, \cdot, \cdot)$, we can solve the linear equation with this drift and obtain its solution $\mu(\sigma)$, which can differ from σ. But in case of coincidence we obtain a solution to the nonlinear equation. Below we use the notation $L_\mu u = \Delta u + \langle b(\mu), \nabla u\rangle$.

Let $C^+[0, T]$ denote the set of nonnegative continuous functions on $[0, T]$. Suppose that $V \in C^2(\mathbb{R}^d)$ and $V \ge 1$. For $\alpha \in C^+[0, \tau_0]$ and $\tau \in (0, T]$ we set

$$\mathcal{M}_{\tau, \alpha}(V) = \left\{\mu(dxdt) = \mu_t(dx)dt \colon \mu_t \ge 0,\ \mu_t(\mathbb{R}^d) = 1,\right.$$

$$\left.\int_{\mathbb{R}^d} V(x)\mu_t(dx) \le \alpha(t),\ t \in [0, \tau]\right\}.$$

If $V(x) = e^{K|x|^{2p}}$, then the corresponding set $\mathcal{M}_{\tau, \alpha}(V)$ will be denoted by $\mathcal{M}_{\tau, \alpha}^{K, p}$.

Let $\|\mu\|_{p, \tau}$ be the norm defined by

$$\|\mu\|_{p, \tau}^2 := \int_0^\tau \|(1 + |x|^p)\mu_s\|_{TV}^2 ds$$

on the linear space of signed measures for which it is finite. Note that $\mathcal{M}_{\tau, \alpha}^{K, p}$ is a complete metric space with respect to the metric generated by this norm.

Corollary 4. *Let $p \ge 1$, $K > 0$ and suppose that for every function $\alpha \in C^+[0, T]$ there exist numbers $\gamma_1(\alpha) > 0$ and $\gamma_2(\alpha) > 2pK$ such that for every $\tau \in (0, T]$ and $\mu \in \mathcal{M}_{\tau, \alpha}^{K, p}$ one has*

$$\langle b(\mu, x, t), x\rangle \le \gamma_1(\alpha) - \gamma_2(\alpha)|x|^{2p}$$

$$\forall (x, t) \in \mathbb{R}^d \times [0, \tau].$$

Suppose also that

$$|b(\mu, x, t) - b(\sigma, x, t)| \le C e^{K|x|^{2p}/2} \|(1 + |x|^p)(\mu_t - \sigma_t)\|_{TV}.$$

Then, for every probability measure ν on $\mathbb{R}^d$ such that $e^{K|x|^{2p}} \in L^1(\nu)$, there exist $\tau \in (0, T]$ and $\alpha \in C^+[0, T]$ such that a solution to the Cauchy problem (5) *in the class of measures $\mathcal{M}_{\tau, \alpha}^{K, p}$ exists and is unique.*

Note that under different assumptions solutions to nonlinear Vlasov equations were constructed in [11] by employing the contraction mapping theorem for the Kantorovich norm. The existence of not necessarily unique solutions has been proved in [12] by using the Schauder fixed point theorem.

Example 3. Let

$$b(\mu, x, t) = \beta(x, t) + \int_{\mathbb{R}^d} K(x, y)\mu_t(dy),$$

where $\beta\colon \mathbb{R}^d \times [0, T] \to \mathbb{R}^d$ and $K\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ are Borel measurable locally bounded mappings such that there exist numbers $C > 0$, $2p > q > 0$, $\gamma_1 > 0$, $\gamma_2 > 2pK$ for which

$$|K(x, y)| \le C(1 + |x|^q)(1 + |y|^p),$$

$$\langle \beta(x, t), x\rangle \le \gamma_1 - \gamma_2 |x|^{2p}.$$

Then all conditions of the above corollary are fulfilled.

The estimate from Theorem 1 can also be used for proving the differentiability of solutions to the Cauchy problem for linear Fokker–Planck–Kolmogorov equations with respect to a parameter. For a different approach, see [13–15].

Corollary 5. *Suppose that for every* $\alpha \in [0, 1]$ *there exists a mapping*

$$b(\alpha, \cdot, \cdot)\colon \mathbb{R}^d \times [0, T] \to \mathbb{R}^d$$

such that b is continuously differentiable in α *and for every ball U there exists a number* $C(U)$ *such that*

$$\|b(\alpha, \cdot, \cdot)\|_{L^\infty(U \times [0, T])} + \|\partial_\alpha b(\alpha, \cdot, \cdot)\|_{L^\infty(U \times [0, T])} \le C(U).$$

Suppose that for every $\alpha \in [0, 1]$ *there exist numbers* $\gamma_1(\alpha)$ *and* $\gamma_2(\alpha)$ *such that*

$$|b(\alpha, x, t)| \le \gamma_1(\alpha) + \gamma_2(\alpha)|x| \ln(1 + |x|).$$

Let μ^α *be a probability solution to the Cauchy problem* (2) *with* $b(\alpha, x, t)$ *and* $A = I$. *Suppose that for every* $\alpha_0 \in [0, 1]$

$$\lim_{r \to 0} \int_0^T \int_{\mathbb{R}^d} \left| \frac{b(\alpha_0 + r, x, t) - b(\alpha_0, x, t)}{r} - \partial_\alpha b(\alpha_0, x, t) \right|^2 \times d\mu_t^{\alpha_0} dt = 0.$$

Then the density $\varrho(\alpha, x, t)$ *of the measure* μ^α *is differentiable in* α.

ACKNOWLEDGMENTS

Our work was supported by the RFBR project nos. 14-01-00237, 14-01-90406, 15-31-20082, the DFG through the SFB 701 at Bielefeld University, and the Simons Foundation.

REFERENCES

1. V. I. Bogachev, N. V. Krylov, and M. Röckner, Russ. Math. Surv. **64** (6), 973–1078 (2009).
2. V. I. Bogachev and A. V. Kolesnikov, Russ. Math. Surv. **67** (5), 785–890 (2012).
3. V. I. Bogachev, *Measure Theory* (Springer, Berlin, 2007), Vol. 1.
4. F. Bolley and C. Villani, Ann. Fac. Sci. Toulouse Math. **14** (3), 331–352 (2005).
5. V. I. Bogachev, M. Röckner, and S. V. Shaposhnikov, Theory Probab. Appl. **50** (4), 561–581 (2006).
6. Yu. M. Kabanov, R. Sh. Liptser, and A. N. Shiryaev, Probab. Theory Relat. Fields **71** (1), 19–35 (1986).
7. F. Liese, Statistics **17** (1), 63–78 (1986).
8. J. A. Carrillo, R. J. McCann, and C. Villani, Rev. Math. Iberoamer **19**, 971–1018 (2003).
9. L. Natile, M. A. Peletier, and G. Savaré, J. Math. Pures Appl. **95** (1), 18–35 (2011).
10. V. I. Bogachev, M. Röckner, and S. V. Shaposhnikov, J. Math. Sci. (New York) **179** (1), 7–47 (2011).
11. R. L. Dobrushin, Funct. Anal. Appl. **13** (2), 115–123 (1979).
12. O. A. Manita and S. V. Shaposhnikov, St. Petersburg Math. **25** (1), 43–62 (2014).
13. E. Pardoux and A. Yu. Veretennikov, Ann. Probab. **31** (3), 1166–1192 (2003).
14. A. Yu. Veretennikov, J. Math. Sci. (New York) **179** (1), 48–79 (2011).
15. V. I. Bogachev, A. Yu. Veretennikov, and S. V. Shaposhnikov, Dokl. Math. **91** (1), 76–79 (2015).