Rend. Lincei Mat. Appl. 28 (2017), 601–618
DOI 10.4171/RLM/778

Measure and Integration — *Estimates for transportation costs along solutions to Fokker–Planck–Kolmogorov equations with dissipative drifts*, by Oxana A. Manita, communicated on February 10, 2017.

Abstract. — We obtain estimates for the Kantorovich transportation costs along solutions to different Fokker–Planck–Kolmogorov equations for measures with the same diffusion part but with different drifts and different initial conditions. We give applications of such estimates to the study of the well-posedness for nonlinear equations.

Key words: Nonlinear Fokker–Planck equation, dissipative operator; Fokker–Planck–Kolmogorov equation, Kantorovich distance

Mathematics Subject Classification: 35K55, 35Q84, 35Q83

1. Introduction

In the present paper we derive and study estimates for the Kantorovich transportation costs along probability solutions for the linear Fokker–Planck–Kolmogorov (FPK) equations for probability measures μ_t and σ_t on $\mathbb{R}^d$, $t \in [0, T]$, with different drifts and different initial conditions

$$\partial_t \mu_t = \operatorname{trace} D^2(Q(x,t)\mu_t) - \operatorname{div}(B_1(x,t)\mu_t), \quad \mu|_{t=0} = \mu_0$$
$$\partial_t \sigma_t = \operatorname{trace} D^2(Q(x,t)\sigma_t) - \operatorname{div}(B_2(x,t)\sigma_t), \quad \sigma|_{t=0} = \sigma_0.$$

We also develop an alternative method of the study of well-posedness and stability of solutions to the nonlinear FPK equations

$$\partial_t \rho_t = \operatorname{trace} D^2(Q(x,t)\rho_t) - \operatorname{div}(B(\rho,x,t)\rho_t), \quad \rho|_{t=0} = \rho_0, \tag{1.1}$$

based on such estimates for linear equations.

Recently, FPK equations have been actively studied from the functional-analytical, variational and as well from the probabilistic point of view. Interesting connections between diverse approaches have been found (a survey of the current state of studies is provided in [3]). Estimates connecting transportation costs along solutions with cost functional between initial data and even coefficients play a great role not only for the study of such qualitative properties of solutions as uniqueness or stability, but also for numerical simulations. In this context estimates for transportation costs along solutions to equations with different drift terms are particularly interesting.

In Section 2 we derive estimates for the Kantorovich transportation cost functional along solutions of FPK equations with different dissipative drifts. To do this, we partially use ideas from [13]. Since these ideas can not be directly applied in the case of different drifts and to nonlinear equations, new methods and ideas should be used. In the present paper extension to these cases has been done for the Kantorovich functionals with bounded cost functions. Moreover, we admit time-dependent coefficients and a non-constant diffusion matrix Q. We note that the requirement of dissipativity is not really restrictive – in typical physical examples, the drift term is the gradient of a concave function, i.e. is dissipative. Section 3 is concerned with applications of these estimates to the study of the well-posedness of the Cauchy problem for nonlinear FPK equations. Well-posedness for nonlinear equations has been studied by many authors even in a more general setting (see, for example, [7, 11, 9, 10]). However, we present an alternative approach to this problem that is applicable in the case of dissipative drifts. A similar method of treating well-posedness via estimates for the costs along solutions to linear equations was used in [5].

Let us introduce some notation and give basic definitions. By $C_0^\infty(\mathbb{R}^d)$ and $C_0^\infty(\mathbb{R}^d \times (0,T))$ we denote the classes of infinitely smooth compactly supported functions on $\mathbb{R}^d$ and $\mathbb{R}^d \times (0,T)$, respectively. For shortness of notation we shall always drop the subscript $\mathbb{R}^d$ when integrating over the whole space. We shall say that a measure ρ on $\mathbb{R}^d \times [0,T]$ is given by a family of probability measures $(\rho_t)_{t\in[0,T]}$ on $\mathbb{R}^d$ (and write $\rho(dx\,dt) = \rho_t(dx)\,dt$ or simply $\rho = \rho_t\,dt$) if $\rho_t \geq 0$, $\rho_t(\mathbb{R}^d) = 1$, for each Borel set U the function $t \mapsto \rho_t(U)$ is measurable and

$$\int_0^T \int \phi\,d\rho = \int_0^T \int \phi\,d\rho_t\,dt \quad \forall \phi \in C_0^\infty(\mathbb{R}^d \times (0,T)).$$

Given a probability measure ρ_0 on $\mathbb{R}^d$, a symmetric Borel measurable matrix $Q(x,t)$ and a Borel measurable mapping $B(x,t) : \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$, consider the following Cauchy problem for the linear FPK equation:

$$\partial_t \rho_t = \mathrm{trace}(Q(x,t)D^2\rho_t) - \mathrm{div}(B(x,t)\rho_t), \quad \rho|_{t=0} = \rho_0. \tag{1.2}$$

Here D^2 denotes the Hessian matrix with respect to the spacial variables. Denote the elements of the diffusion matrix $Q(x,t)$ by $q^{ij}(x,t)$, $1 \leq i,j \leq d$ and the elements of the vector drift $B(x,t)$ by $b^j(x,t)$, $1 \leq j \leq d$. Set

$$L\phi = q^{ij}(x,t)\partial^2_{x_i x_j}\phi + b^i(x,t)\partial_{x_i}\phi,$$

where the summation over all repeated indices is taken. We shall say that a measure $\rho(dx\,dt) = \rho_t(dx)\,dt$ is a solution to the Cauchy problem (1.2) if the mappings $q^{ij}(x,t)$, $b^i(x,t)$, $1 \leq i,j \leq d$, are Borel and belong to $L^1(\rho, U \times [0,T])$ for each ball $U \subset \mathbb{R}^d$, and for each test function $\varphi \in C_0^\infty(\mathbb{R}^d)$ we have

$$\int \varphi\,d\rho_t = \int \varphi\,d\rho_0 + \int_0^t \int L\varphi\,d\rho_s\,ds \tag{1.3}$$

for all $t \in [0,T]$. Occasionally it is more convenient to use an equivalent definition (see [4]), more precisely, the identity

$$\int \phi(x,t)\,d\rho_t = \int \phi(x,0)\,d\rho_0 + \int_0^t \int [\partial_s \phi + L\phi]\,d\rho_s\,ds, \tag{1.4}$$

for all $t \in [0,T]$ and all test functions $\phi \in C^{2,1}(\mathbb{R}^d \times [0,T)) \cap C(\mathbb{R}^d \times [0,T])$ that are identically zero outside a ball $U \subset \mathbb{R}^d$. If we know a priori that the drift term B is integrable over $\mathbb{R}^d \times [0,T]$ with respect to the measure $d\rho_s\,ds$ and ϕ is not of compact support, but has two continuous bounded derivatives, then (1.4) also holds true for such ϕ (to show this, it suffices to use the standard truncation argument).

2. Estimates for Kantorovich transportation costs along solutions to linear equations with different drifts

In this section, we shall focus on two solutions of the linear FPK equation with different initial conditions and different drifts. Fix $T > 0$. Given probability measures μ_0 and σ_0 on $\mathbb{R}^d$, a symmetric Borel measurable matrix $Q(x,t)$ and Borel measurable mappings $B_\mu, B_\sigma : \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$, consider the two corresponding Cauchy problems

$$\begin{aligned} \partial_t \mu_t &= \operatorname{trace}(Q(x,t)D^2\mu_t) - \operatorname{div}(B_\mu(x,t)\mu_t), \quad \mu|_{t=0} = \mu_0 \\ \partial_t \sigma_t &= \operatorname{trace}(Q(x,t)D^2\sigma_t) - \operatorname{div}(B_\sigma(x,t)\sigma_t), \quad \sigma|_{t=0} = \sigma_0. \end{aligned} \tag{2.1}$$

The indices μ and σ in the drift coefficients are merely used to distinguish the different drifts and don't indicate any dependence on the solution.

Given a monotone nonnegative continuous function h on $\mathbb{R}$ with $h(0) = 0$, we introduce the Kantorovich h-cost functional between the probability measures μ and σ by

$$C_h(\mu,\sigma) := \inf_{\pi \in \Pi(\mu,\sigma)} \int_{\mathbb{R}^d \times \mathbb{R}^d} h(|x-y|)\,d\pi(x,y), \tag{2.2}$$

where $\Pi(\mu,\sigma)$ is the set of couplings between μ and σ. Recall that a probability measures π on $\mathbb{R}^d \times \mathbb{R}^d$ belongs to $\Pi(\mu,\sigma)$ if and only if $\pi(E \times \mathbb{R}^d) = \mu(E)$ and $\pi(\mathbb{R}^d \times E) = \sigma(E)$ for each Borel set $E \subset \mathbb{R}^d$. If h is a concave function with $h(r) > 0$ for $r > 0$, then C_h defines a distance on the space of probability measures and turns it into a complete metric space with the topology that coincides with the usual weak one (see [1, Proposition 7.1.5]). Another important example is given by $h(r) = \min\{|r|^p, 1\}$ for some $p \geq 1$. In this case $C_h^{1/p}$ turns the space of probability measures into a complete metric space. Moreover, convergence with respect to this metric is equivalent to the weak convergence (see [6, Th. 1.1.9]).

We assume that a monotone non-decreasing continuous bounded cost function h with $h(0) = 0$ is fixed. Set $\|h\|_\infty := \sup_{z \in \mathbb{R}^d} h(|z|) < \infty$.

Throughout the paper we assume that the following regularity condition holds:

(A1) The diffusion matrix $Q(x,t)$ has uniformly bounded elements with uniformly bounded first derivatives. Moreover, it is strictly elliptic: there exists $\nu > 0$ such that $\forall (x,t) \in \mathbb{R}^d \times [0,T]$

$$\langle Q(x,t)y, y\rangle \geq \nu |y|^2 \quad \forall y \in \mathbb{R}^d. \tag{2.3}$$

THEOREM 2.1. *Let* (A1) *hold. Let* $(\mu_t)_{t\in[0,T]}$ *and* $(\sigma_t)_{t\in[0,T]}$ *be solutions to* (2.1) *with initial conditions* μ_0 *and* σ_0*, respectively. Suppose that the drift term* B_μ *is* λ*-dissipative in x, i.e.,*

$$\langle B_\mu(x,t) - B_\mu(y,t), x - y\rangle \leq \lambda \|x-y\|^2 \tag{2.4}$$

for all $x, y \in \mathbb{R}^d$ *and all* $t \in [0,T]$*. Let*

$$B_\mu(x,t) - \lambda x, B_\sigma(x,t) - \lambda x \in L^2(\mathbb{R}^d \times [0,T], d(\mu_s + \sigma_s)\, ds) \tag{2.5}$$

Then

$$C_{h_{\lambda t}}(\mu_t, \sigma_t) \leq C_h(\mu_0, \sigma_0) + \|h\|_\infty \cdot \left(\int_0^t \int \nu^{-1} |B_\mu - B_\sigma|^2 \, d\sigma_s \, ds\right)^{1/2} \times \left(1 + \int_0^t \int \nu^{-1} |B_\mu - B_\sigma|^2 \, d\sigma_s \, ds\right)^{1/2}, \tag{2.6}$$

for all $t \in [0,T]$*, where* $h_s(r) := h(re^{-s})$*.*

REMARK 2.1. Bound (2.6) is obviously asymmetric in measure: we impose dissipativity on B_μ, and the integration in the right-hand side is taken over σ. This property might be interesting from the point of view of possible numerical simulations. Indeed, if we want to solve a FPK equation

$$\partial_t \mu_t = \operatorname{trace}(Q(x,t)D^2\mu_t) - \operatorname{div}(B(x,t)\mu_t)$$

with a dissipative drift B, we can approximate the drift by "better" drifts B_n and solve the FPK equations with those drifts. Then (2.6) controls the transportation costs between the desired solution μ_t and the approximative solution μ_t^n in terms of the distance between the drifts integrated with respect to the known solution μ^n.

PROOF. Let $(\mu_t)_{t\in[0,T]}$ and $(\sigma_t)_{t\in[0,T]}$ satisfy the assumptions of the theorem. According to [2], the measures μ_t and σ_t have strictly positive densities with respect to Lebesgue measure on $\mathbb{R}^d$ for each $t \in [0,T]$. We split the proof of (2.6) into several steps.

Step 1. *Reduction to the dissipative case* ($\lambda = 0$). We rescale the problem, keeping the cost function unchanged, in order to reduce the problem to the case of a dissipative drift B_μ. To this end we use the rescaling procedure from [13] with the opposite sign (since our drift term and the drift term in the cited work have the opposite signs). For completeness, we provide this rescaling procedure: for $\lambda \neq 0$ define the change of time

$$s(t) := \int_0^t e^{-2\lambda r}\,dr = \frac{1 - e^{-2\lambda t}}{2\lambda}, \quad t(s) = \frac{-\ln(1-2\lambda s)}{2\lambda}, \quad s \in [0, S_\infty),$$

where $S_\infty = +\infty$ for $\lambda < 0$ and $S_\infty = 1/(2\lambda)$ for $\lambda > 0$. For the measures μ_t and σ_t we introduce their rescaled versions ρ_s^μ and ρ_s^σ: for each Borel set $E \subset \mathbb{R}^d$ set $\rho_s^w(E) := w_{t(s)}(e^{\lambda t(s)}E)$ for $w = \mu, \sigma$. We observe that $C_h(\rho_s^\mu, \rho_s^\sigma) = C_{h_{\lambda t}}(\mu_t, \sigma_t)$. Since B_μ is λ-dissipative, $A_\mu := B_\mu - \lambda I$ is dissipative. Define the rescaled diffusion coefficient by

$$\tilde{Q}(y,s) := Q(t(s), e^{\lambda t(s)}y)$$

and the rescaled drifts by

$$\tilde{B}_w(y,s) := e^{\lambda t(s)}B_w(t(s), e^{\lambda t(s)}y), \quad \tilde{A}_w(y,s) := e^{\lambda t(s)}A_w(t(s), e^{\lambda t(s)}y), \quad w = \mu, \sigma.$$

Note that $\tilde{A}_\mu$ is also a dissipative operator. The measure $\mu = \mu_t\,dt$ is a solution to

$$\partial_t \mu_t = \operatorname{trace}(QD^2\mu_t) - \operatorname{div}(B_\mu \mu_t)$$

if and only if the rescaled version $\rho^\mu = \rho_t^\mu\,dt$ is a solution to

$$\partial_t \rho_t^\mu = \operatorname{trace}(\tilde{Q}D^2\rho_t^\mu) - \operatorname{div}(\tilde{A}_\mu \rho_t^\mu); \tag{2.7}$$

moreover, (2.5) holds true if and only if for all nonnegative $s_1 < s_2 \le S(T) < S_\infty$ one has

$$\int_{s_1}^{s_2} \int |\tilde{A}_\mu(x,s)|^2\,d\rho_s^\mu\,ds < +\infty.$$

The integrability statement follows immediately from the change of variables formula, identity (2.7) can be verified explicitly: it suffices to consider the change of variables $\mathsf{X}(x,t) := (e^{-\lambda t}x, s(t))$ and calculate the derivatives. A similar statement holds for σ and ρ^σ. This means that it is sufficient to prove (2.6) only in the case $\lambda = 0$, i.e., in the case of a dissipative drift term B_μ.

Step 2. *Approximation of the drift term.* We construct a family of smooth (in both variables) bounded Lipschitz (as functions of x) dissipative operators $A_k^\varepsilon(x,t)$ approximating the dissipative drift term $B_\mu(x,t)$.

For each t the operator $B_\mu(\cdot, t)$ can be approximated by Lipschitz in x bounded dissipative operators $A_k(\cdot, t)$ (see [13, Th. 2.4, 2.5]): for each $t \in [0, T]$

$$\lim_{k\to\infty} A_k(x,t) = B_\mu(x,t) \quad \text{for a.e. } (x) \in \mathbb{R}^d \quad \text{and} \quad \sup_{x\in\mathbb{R}^d} |A_k(x,t)| \le k+1. \tag{2.8}$$

Let us fix a nonnegative function $\eta \in C_0^\infty([0,T])$ such that $\|\eta\|_{L^1([0,T])} = 1$ and introduce the family of mollifiers $\eta_\varepsilon(t) := \eta(t/\varepsilon)$. Since for each k the mapping $A_k(x,t)$ is bounded, the mappings $A_k^\varepsilon(x,t) := \eta_\varepsilon(t) * A_k(x,t)$ have bounded derivatives of all orders with respect to t and converge to $A_k(x,t)$ as $\varepsilon \to 0$ for a.e. $(x,t) \in \mathbb{R}^d \times [0,T]$. Notice that A_k^ε also have bounded first order derivatives with respect to the spacial variables. Moreover, A_k^ε are dissipative in x:

$$\langle A_k^\varepsilon(x,t) - A_k^\varepsilon(y,t), x-y\rangle = \eta_\varepsilon(t) * \langle A_k(x,t) - A_k(y,t), x-y\rangle \le 0.$$

Finally, we define operators $\mathscr{L}_k^\varepsilon$ as follows: for $(x,s) \in \mathbb{R}^d \times [0,T]$ we set

$$\mathscr{L}_k^\varepsilon[\phi](x,s) := \operatorname{trace}(Q(x,s)D^2\phi(x,s)) + \langle A_k^\varepsilon(x,s), \nabla_x\phi(x,s)\rangle, \quad \phi(\cdot,s) \in C^2(\mathbb{R}^d).$$

Step 3. *Reduction of the class of test functions.* It is well-known (see, for example, [15, Th. 1.3]) that the problem (2.2) admits a dual formulation: define the class Φ_h as

$$\Phi_h := \{(\phi,\psi) \in L^1(\mu) \times L^1(\nu) : \phi(x) + \psi(y) \le h(|x-y|)\}.$$

Hence

$$C_h(\mu,\sigma) = \sup_{(\phi,\psi)\in\Phi_h} \int \phi\, d\mu + \int \psi\, d\sigma. \tag{2.9}$$

An important observation ([16, Lemma 2.3]) is that in the case of a bounded cost function h the supremum in the dual problem (2.9) can be taken over the following smaller class of functions Φ_h^δ for any $\delta > 0$:

$$\Phi_h^\delta := \Phi_h \cap C_0^\infty(\mathbb{R}^d) \cap \{(\phi,\psi) : \inf\psi > -\delta \text{ and } \sup\psi \le \|h\|_\infty\}. \tag{2.10}$$

The proof is based on the fact that functions φ and ψ can be shifted by different constants and truncated in such a way that the new pair (φ_0, ψ_0) is still admissible and the value in (2.9) does not decrease:

$$\int \varphi_0\, d\mu + \int \psi_0\, d\sigma \ge \int \varphi\, d\mu + \int \psi\, d\sigma \quad \text{and}$$

$$\inf_{\mathbb{R}^d} \psi_0 = 0, \quad \sup_{\mathbb{R}^d} \psi_0 \le \|h\|_\infty, \quad \sup_{\mathbb{R}^d} \varphi_0 \ge 0.$$

If we want to deal with smooth compactly supported functions, then the bounds become a bit worse and lead to the class (2.10).

In the sequel we shall take the supremum in (2.9) over the class Φ_h^b of admissible pairs of $C_0^\infty(\mathbb{R}^d)$-functions such that $\|\psi\|_\infty \le \|h\|_\infty$.

Step 4. *The adjoint problem.* Fix an admissible pair $(\phi, \psi) \in \Phi_h^b$. The smoothness of operators A_k^ε imply (see [14, Th. 3.2.1]) that the following adjoint problems have solutions $g, f \in C_b^{2,1}(\mathbb{R}^d \times [0,t])$:

$$(2.11)\qquad \partial_s g + \mathscr{L}_k^\varepsilon g = 0, \quad g(\cdot, t) = \phi(\cdot) \quad \text{and} \quad \partial_s f + \mathscr{L}_k^\varepsilon f = 0, \quad f(\cdot, t) = \psi(\cdot)$$

First, due to the maximum principle (see [14, Th. 3.1.1]) we have

$$(2.12)\qquad \sup_{\mathbb{R}^d \times [0,t]} |g| \le \sup_{\mathbb{R}^d} |\phi|, \qquad \sup_{\mathbb{R}^d \times [0,t]} |f| \le \sup_{\mathbb{R}^d} |\psi|.$$

Let us derive some bounds for $|\nabla g|$ and $|\nabla f|$. The method of doing this is inspired by the Bernstein estimates. Set for shortness $A_k^\varepsilon := (\alpha^1, \ldots, \alpha^d)$. Set $v(x,t) := |\nabla g|^2 + \kappa g^2 - t$, where κ will be chosen below. Explicit computations give (the summation over all repeated indices is assumed)

$$(2.13)\qquad \begin{aligned} (\partial_s - \mathscr{L}_k^\varepsilon) v \overset{(2.11)}{=} {} & 2\partial_{x_k} g(\partial_{x_k} q^{ij} \partial^2_{x_i x_j} g + \partial_{x_k} \alpha^i \partial_{x_i} g) \\ & - 2 q^{ij} \partial^2_{x_k x_i} g \partial^2_{x_k x_j} g - 2\kappa q^{ij} \partial_{x_i} g \partial_{x_j} g - 1. \end{aligned}$$

Due to dissipativity, DA_k^ε defines a negative quadratic form and

$$\partial_{x_k} g \partial_{x_k} \alpha^i \partial_{x_i} g = (DA_k^\varepsilon \nabla g, \nabla g) \le 0.$$

By this observation, (2.3) and the Cauchy inequality $2ab \le ca^2 + c^{-1}b^2$ with $c = 2v$, the right-hand side of (2.13) is dominated by

$$\begin{aligned} & \omega c^{-1} |\nabla g|^2 + c \sum_{i,j} (\partial^2_{x_i x_j} g) - 2v \sum_{i,j} (\partial^2_{x_i x_j} g) - 2v\kappa |\nabla g|^2 - 1 \\ & \quad = \Omega |\nabla g|^2 - 2v\kappa |\nabla g|^2 - 1, \end{aligned}$$

where $\omega := 2\max\{|\partial_{x_k} q^{ij}|\}$ and $\Omega := \omega \cdot (2v)^{-1}$ depend only on the diffusion matrix and not on the drift. Letting $\kappa := \Omega \cdot (2v)^{-1}$, we get

$$(\partial_s - \mathscr{L}_k^\varepsilon) v \le -1.$$

Therefore, the maximum principle ensures

$$\max_{\mathbb{R}^d \times [0,t]} |v| \le \max_{\mathbb{R}^d} |v(x,0)| \equiv \max_{\mathbb{R}^d} |\nabla \phi|^2 + \kappa \max_{\mathbb{R}^d} |\phi|^2,$$

hence

$$(2.14)\qquad \sup_{\mathbb{R}^d \times [0,t]} |\nabla g(x,s)| \le \left(\max_{\mathbb{R}^d} |\nabla \phi|^2 + \kappa \max_{\mathbb{R}^d} |\phi|^2 \right)^{1/2} =: C_1.$$

Similarly,

$$\sup_{\mathbb{R}^d\times[0,t]} |\nabla f(x,s)| \le \Big(\max_{\mathbb{R}^d} |\nabla\psi|^2 + \kappa \max_{\mathbb{R}^d} |\psi|^2\Big)^{1/2} + t =: C_2. \tag{2.15}$$

Set $l := C_1 + C_2$.

Finally, let us prove the crucial assertion: if the pair (φ, ψ) is admissible and f and g solve (2.11), then $g(x,0) + f(y,0) \le h(|x-y|)$. In the case $Q \equiv I$ this was proved in [13, Th. 3.1]. In the general case the proof almost repeats the case $Q = I$, but we sketch it for completeness. By approximating h from above, we can assume without loss of generality that $h \in C^1(\mathbb{R})$. Let $H(y_1, y_2) := h(|y_1 - y_2|)$ and

$$0 \le \xi(y_1,y_2) = \xi(y_2,y_1) = \begin{cases} \frac{h(|y_1-y_2|)}{|y_1-y_2|} & \text{if } y_1 \neq y_2 \\ 0 & \text{if } y_1 = y_2 \end{cases}.$$

First assume that

$$\partial_s g + \mathscr{L}_k^\varepsilon g > 0, \quad \partial_s f + \mathscr{L}_k^\varepsilon f > 0.$$

Suppose that $\zeta(y_1, y_2, s) := g(y_1, s) + f(y_2, s) - H(y_1, y_2)$ attains a local maximum at (Y_1, Y_2, S) and $S < t$. Then $\partial_s\zeta(Y_1, Y_2, S) = \partial_s g(Y_1, S) + \partial_s f(Y_2, S) \le 0$,

$$\begin{aligned} &\nabla_{y_1}\zeta(Y_1, Y_2, S) = \nabla_{y_2}\zeta(Y_1, Y_2, S) = 0 \\ &\quad \Rightarrow \nabla_{y_1} g(Y_1, S) = -\nabla_{y_2} f(Y_2, S) = \xi(Y_1, Y_2)(Y_1 - Y_2) \end{aligned}$$

and, due to dissipativity,

$$\begin{aligned} &A_k^\varepsilon(Y_1,S)\nabla_{y_1} g(Y_1,S) + A_k^\varepsilon(Y_2,S)\nabla_{y_2} f(Y_2,S) \\ &\quad = \xi(Y_1,Y_2)\langle A_k^\varepsilon(Y_1,S) - A_k^\varepsilon(Y_2,S), Y_1 - Y_2\rangle \le 0. \end{aligned}$$

Since $\zeta(Y_1 + z, Y_2 + z, S)$ as a function of z has a local maximum at $z = 0$ and Q is positive definite, $\operatorname{trace} \overline{Q} D^2\zeta = \operatorname{trace} Q(Y_1, S)D^2 g + \operatorname{trace} Q(Y_2, S)D^2 f \le 0$, where

$$\overline{Q}(y_1, y_2, s) := \begin{pmatrix} Q(y_1) & 0 \\ 0 & Q(y_2) \end{pmatrix}.$$

Summing up, we get

$$(\partial_s g + \mathscr{L}_k^\varepsilon g) + (\partial_s f + \mathscr{L}_k^\varepsilon f) \le 0;$$

this contradiction means that the local maximum can be attained only at $S = t$. Now we proceed to the equality. Setting for some $\varepsilon, \delta > 0$

$$\begin{aligned} g_{\varepsilon,\delta}(y_1, s) &:= g(y_1, s) - \delta(t-s) - \varepsilon e^{-s}|y_1|^2, \\ f_{\varepsilon,\delta}(y_2, s) &:= f(y_2, s) - \delta(t-s) - \varepsilon e^{-s}|y_2|^2 \end{aligned}$$

and computing $\partial_s + \mathscr{L}_k^\varepsilon$, we arrive at the previous case for ε, δ small enough (since all the coefficients of the differential operator are bounded). Passing to the limit as $\varepsilon, \delta \to 0$, we obtain the desired assertion.

Step 5. *Deriving the estimate*-1. Plugging solutions of (2.11) into identity (1.4), we get

$$\int \phi \, d\mu_t - \int g(x,0)\, d\mu_0 = -\int_0^t \int (A_k^\varepsilon(x,s) - B_\mu) \cdot \nabla g(x,s)\, d\mu_s\, ds,$$

$$\int \psi \, d\sigma_t - \int f(x,0)\, d\sigma_0 = -\int_0^t \int (A_k^\varepsilon(x,s) - B_\sigma) \cdot \nabla f(x,s)\, d\sigma_s\, ds.$$

Because of (2.14),

$$\int \phi \, d\mu_t - \int g(x,0)\, d\mu_0 \le l \int_0^t \int |A_k^\varepsilon - B_\mu|\, d\mu_s\, ds.$$

Note that

$$\begin{aligned} &\int \psi \, d\sigma_t - \int f(x,0)\, d\sigma_0 \\ &\quad \le \int_0^t \int |A_k^\varepsilon - B_\mu| \cdot |\nabla f|\, d\sigma_s\, ds + \int_0^t \int |B_\mu - B_\sigma| \cdot |\nabla f|\, d\sigma_s\, ds \\ &\quad \overset{(2.15)}{\le} l \int_0^t \int |A_k^\varepsilon - B_\mu|\, d\sigma_s\, ds + \int_0^t \int |B_\mu - B_\sigma| \cdot |\nabla f|\, d\sigma_s\, ds. \end{aligned}$$

Summing up these inequalities, we get

$$\begin{aligned} \int \phi \, d\mu_t + \int \psi \, d\sigma_t \le \int g(x,0)\, d\mu_0 + \int f(x,0)\, d\sigma_0 + l \cdot R_k^\varepsilon \\ + \int_0^t \int |B_\mu - B_\sigma| \cdot |\nabla f|\, d\sigma_s\, ds, \end{aligned}$$

where

$$R_k^\varepsilon := \int_0^t \int |A_k^\varepsilon - B_\mu|\, d(\mu_s + \sigma_s)\, ds.$$

According to Step 4 we have $g(x,0) + f(y,0) \le h(|x-y|)$. Thus,

$$\int g(x,0)\, d\mu_0 + \int f(x,0)\, d\sigma_0 \le C_h(\mu_0, \sigma_0). \tag{2.16}$$

So we get

$$\int \phi \, d\mu_t + \int \psi \, d\sigma_t \le C_h(\mu_0, \sigma_0) + l \cdot R_k^{\varepsilon} + \int_0^t \int |B_\mu - B_\sigma| \cdot |\nabla f| \, d\sigma_s \, ds. \tag{2.17}$$

Step 6. Integral bound for ∇f. The last term in the right-hand side of (2.17) is dominated by

$$\Big(\int_0^t \int v^{-1} |B_\mu - B_\sigma|^2 \, d\sigma_s \, ds\Big)^{1/2} \cdot \Big(\int_0^t \int |\sqrt{Q}\nabla f|^2 \, d\sigma_s \, ds\Big)^{1/2}.$$

To estimate the second multiplier, i.e., $\|\sqrt{Q}\nabla f\|^2_{L^2(\mathbb{R}^d \times [0,T]; d\sigma_s \, ds)}$, note that f^2 is a function of class $C_b^{2,1}(\mathbb{R}^d \times [0,t)) \cap C(\mathbb{R}^d \times [0,T])$ and it can be plugged into identity (1.4) for the measure σ:

$$\begin{aligned}
&\int \psi^2 \, d\sigma_t - \int f^2(x,0) \, d\sigma_0 \\
&\quad = \int_0^t \int (\partial_s + L_\sigma) f^2 \, d\sigma_s \, ds \\
&\quad = \int_0^t \int 2f(\partial_s f + \operatorname{trace}(QD^2 f) + \langle B_\sigma, \nabla f \rangle) + 2|\nabla f|^2 \, d\sigma_s \, ds \\
&\quad = -\int_0^t \int 2f \langle A_k^{\varepsilon} - B_\sigma, \nabla f \rangle + 2|\nabla f|^2 \, d\sigma_s \, ds.
\end{aligned}$$

Hence

$$\begin{aligned}
2\int_0^t \int |\nabla f|^2 \, d\sigma_s \, ds \le{}& \int \psi^2 \, d\sigma_t - \int f^2(x,0) \, d\sigma_0 \\
&+ 2 \max|f(x,t)| \int_0^t \int v^{-1/2} |A_k^{\varepsilon} - B_\sigma| \cdot |\sqrt{Q}\nabla f| \, d\sigma_s \, ds.
\end{aligned}$$

The maximum principle (2.12) and relation (2.10) imply that $\max|f(x,t)| \le \max|\psi(x)| \le \|h\|_\infty$. Taking into account that $ab \le 2^{-1}\gamma a^2 + (2\gamma)^{-1} b^2$ with $\gamma = \|h\|_\infty$, we arrive at

$$\begin{aligned}
2\int_0^t \int |\nabla f|^2 \, d\sigma_s \, ds \le{}& \|h\|_\infty^2 + \frac{\|h\|_\infty^2}{v} \int_0^t \int |A_k^{\varepsilon} - B_\sigma|^2 \, d\sigma_s \, ds \\
&+ \int_0^t \int |\sqrt{Q}\nabla f|^2 \, d\sigma_s \, ds.
\end{aligned}$$

Cancelling alike terms and recalling (2.17), we get

$$\int \phi \, d\mu_t + \int \psi \, d\sigma_t \le C_h(\mu_0, \sigma_0) + l \cdot R_k^\varepsilon + \|h\|_\infty \nu^{-1/2} \cdot r_k^\varepsilon \cdot \Big(\int_0^t \int |B_\mu - B_\sigma|^2 \, d\sigma_s \, ds \Big)^{1/2}, \tag{2.18}$$

where

$$r_k^\varepsilon := \Big(1 + \nu^{-1} \int_0^t \int |A_k^\varepsilon - B_\sigma|^2 \, d\sigma_s \, ds \Big)^{1/2}.$$

Step 7. Limits as $\varepsilon \to 0$ and $k \to \infty$. Deriving the estimate-2. First of all, we recall that

$$A_k^\varepsilon(x,t) \to A_k(x,t) \quad \text{for a.e. } (x,t) \in \mathbb{R}^d \times [0,T],$$

and the measures $d\sigma_s \, ds$ and $d(\mu_s + \sigma_s) \, ds$ have strictly positive densities on $\mathbb{R}^d \times [0,T]$ with respect to Lebesgue measure. Thus,

$$A_k^\varepsilon(x,t) \to A_k(x,t) \, d\sigma_s \, ds\text{-a.e.} \quad \text{and} \quad d(\mu_s + \sigma_s) \, ds\text{-a.e.}$$

Since for each k the mappings A_k^ε and A_k are bounded, Lebesgue's dominated convergence theorem yields

$$\int_0^t \int |A_k^\varepsilon - B_\sigma|^2 \, d\sigma_s \, ds \to \int_0^t \int |A_k - B_\sigma|^2 \, d\sigma_s \, ds, \quad \varepsilon \to 0,$$

$$\int_0^t \int |A_k^\varepsilon - B_\mu| \, d(\mu_s + \sigma_s) \, ds \to \int_0^t \int |A_k - B_\mu| \, d(\mu_s + \sigma_s) \, ds, \quad \varepsilon \to 0.$$

Next, we recall that

$$\lim_{k \to \infty} A_k(x,t) = B_\mu(x,t) \quad \text{for a.e. } (x,t) \in \mathbb{R}^d \times [0,T].$$

Similarly, taking into account (2.5), one can apply Lebesgue's dominated convergence theorem and get

$$\lim_{k \to \infty} \lim_{\varepsilon \to 0} R_k^\varepsilon = 0, \quad \lim_{k \to \infty} \lim_{\varepsilon \to 0} r_k^\varepsilon = \Big(1 + \int_0^t \int \nu^{-1} \cdot |B_\mu - B_\sigma|^2 \, d\sigma_s \, ds \Big)^{1/2}.$$

Hence one can pass in (2.18) to the limits as $\varepsilon \to 0$, then let $k \to \infty$ and obtain

$$\int \phi \, d\mu_t + \int \psi \, d\sigma_t \le C_h(\mu_0, \sigma_0) + \|h\|_\infty \Big(\int_0^t \int \nu^{-1} |B_\mu - B_\sigma|^2 \, d\sigma_s \, ds \Big)^{1/2} \times \Big(1 + \nu^{-1} \int_0^t \int |B_\mu - B_\sigma|^2 \, d\sigma_s \, ds \Big)^{1/2}. \tag{2.19}$$

Passing to the supremum over $(\phi, \psi) \in \Phi_h^b$ and using Step 3, we obtain

$$C_h(\mu_t, \sigma_t) \leq C_h(\mu_0, \sigma_0) + \|h\|_\infty \Big(\int_0^t \int v^{-1}|B_\mu - B_\sigma|^2 \, d\sigma_s \, ds\Big)^{1/2}$$
$$\times \Big(1 + v^{-1} \int_0^t \int |B_\mu - B_\sigma|^2 \, d\sigma_s \, ds\Big)^{1/2},$$

which is the required estimate (2.6) with $\lambda = 0$. □

3. Applications to nonlinear equations

In this section we focus on applications of the obtained estimates to the study of the well-posedness of the Cauchy problem for nonlinear FPK equations.

Given a continuous positive function α on $[0, T]$, $\tau \in (0, T]$ and a nonnegative continuous function $V(x)$ on $\mathbb{R}^d$ with $V(x) \to +\infty$ as $|x| \to +\infty$, let us define the classes of measures

$$M_{\tau,\alpha}(V) = \Big\{\mu = (\mu_t)_{t\in[0,\tau]} : \int V(x)\, d\mu_t \leq \alpha(t), t \in [0, \tau]\Big\},$$

$$M_\tau(V) = \Big\{\mu = (\mu_t)_{t\in[0,\tau]} : \sup_{t\in[0,\tau]} \int V(x)\, d\mu_t < +\infty\Big\}.$$

Throughout this section we assume that a non-degenerate $d \times d$-matrix $Q(x, t)$ satisfying (A1) is fixed. Suppose that for each measure $\mu = \mu_t \, dt \in M_T(V)$ a Borel mapping

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

is defined. Consider the Cauchy problem for the nonlinear FPK equation

$$\partial_t \mu_t = \mathrm{trace}(Q(x, t)D^2\mu_t) - \mathrm{div}(B(\mu, x, t)\mu_t), \quad \mu_t|_{t=0} = \mu_0. \tag{3.1}$$

Again denote the elements of the diffusion matrix $Q(x, t)$ by $q^{ij}(x, t)$, $1 \leq i, j \leq d$ and the elements of the vector drift $B(\mu, x, t)$ by $b^j(\mu, x, t)$, $1 \leq j \leq d$. Set

$$L_\mu \phi = q^{ij}(x, t)\partial^2_{x_i x_j}\phi + b^i(\mu, x, t)\partial_{x_i}\phi,$$

where the summation over all repeated indices is taken. As above, we call a measure $\mu = \mu_t \, dt$, $t \in [0, T]$ a solution to (3.1) if identity (1.3) holds with L_μ in place of L. Let us introduce the following assumptions about the drift:

(B1) The drift term B is λ-dissipative in x, i.e., for every measure $\mu \in M_T(V)$ one has

$$\langle B(\mu, x, t) - B(\mu, y, t), x - y\rangle_{\mathbb{R}^d} \leq \lambda\|x - y\|^2 \tag{3.2}$$

for all $x, y \in \mathbb{R}^d$ and all $t \in [0, T]$.

(B2) for all measures μ and σ from $M_T(V)$ one has

$$B(\mu, x, t) - \lambda x \in L^2(\mathbb{R}^d \times [0, T], d(\mu_s + \sigma_s)\, ds). \tag{3.3}$$

We start with the question of uniqueness and stability of probability solutions to (3.1). As above, we assume that a continuous non-decreasing monotone bounded cost function h with $h(0) = 0$ is fixed. Given a non-negative non-decreasing function G, set

$$G_*(r) := \int_r^1 \frac{du}{G^2(\sqrt{u})}.$$

Corollary 3.1. *Fix a non-negative continuous function $V(x)$ on $\mathbb{R}^d$ with $V(x) \to +\infty$ as $|x| \to +\infty$ such that $V \in L^1(\mathbb{R}^d; \mu_0) \cap L^1(\mathbb{R}^d; \sigma_0)$. Assume that the coefficients of equation* (3.1) *satisfy* (A1), (B1) *and* (B2) *with this V. Moreover, assume that that for each two measures $\mu = (\mu_t)_{t \in [0,T]}$ and $(\sigma_t)_{t \in [0,T]}$ in $M_T(V)$ one has*

$$|B(\mu, x, t) - B(\sigma, x, t)| \leq \sqrt{V(x)}\, G(C_h(\mu_t, \sigma_t)) \tag{3.4}$$

for some non-negative increasing function G such that $G_*(0) = +\infty$.

Then every two solutions $(\mu_t)_{t \in [0,T]}$ and $(\sigma_t)_{t \in [0,T]}$ of problem (3.1) *in the class $M_T(V)$ with initial data μ_0 and σ_0, respectively, satisfy the inequality*

$$C_{h_{\lambda t}}(\mu_t, \sigma_t) \leq \left(G_*^{-1}(G_*(2(C_h(\mu_0, \sigma_0))^2) - ct)\right)^{1/2}$$

for all $t \in [0, T]$, where G_*^{-1} is the inverse function to G^* and $c > 0$ is a positive constant.

Example 3.1. Assumptions (B1), (B2) and (3.4) are fulfilled, for example, for the drift terms of the form

$$B(\mu, x, t) = H(x) \int k(x, y)\, d\mu_t(y)$$

with $0 \leq H(x) \leq \sqrt{V(x)}$ and a λ-dissipative in the first variable kernel $k(\cdot\,,\cdot)$ such that

$$|k(x, y) - k(z, y)| \leq h(|x - y|).$$

Proof of Corollary 3.1. First of all, if μ is a solution to (3.1) and the assumptions of Corollary 3.1 are fulfilled, then the linear FPK equation

$$\partial_t \rho_t = \operatorname{trace}(Q(x, t) D^2 \rho_t) - \operatorname{div}(B(\mu, x, t)\rho_t), \quad \rho_t|_{t=0} = \mu_0$$

has a solution $\rho = \mu$ that satisfies the assumptions of Theorem 2.1; similarly σ is a solution to a linear equation with the drift term $B(\sigma, \cdot\,, \cdot)$. Hence one can apply (2.6) with $B_\mu(\cdot\,,\cdot) = B(\mu, \cdot\,,\cdot)$ and $B_\sigma(\cdot\,,\cdot) = B(\sigma, \cdot\,,\cdot)$.

Next, arguing as at Step 1 of the proof of Theorem 2.1, we can assume that the drift term B is dissipative. With condition (3.4) in hand, estimate (2.6) takes the form

$$C_h(\mu_t,\sigma_t) \le C_h(\mu_0,\sigma_0) + \|h\|_\infty \sqrt{\nu^{-1}a}\Big(\int_0^t G^2(C_h(\mu_s,\sigma_s))\,ds\Big)^{1/2} \times \Big(1+\nu^{-1}a\int_0^t G^2(C_h(\mu_s,\sigma_s))\,ds\Big)^{1/2}, \tag{3.5}$$

where $a = \sup_{t\in[0,T]} \displaystyle\int V(x)\,d\mu_t < +\infty$ and ν is the ellipticity constant of Q. Note that $C_h(\mu_t,\sigma_t) \le \|h\|_\infty$. Then (3.5) can be reduced to a weaker inequality

$$C_h(\mu_t,\sigma_t) \le C_h(\mu_0,\sigma_0) + K\Big(\int_0^t G^2(C_h(\mu_s,\sigma_s))\,ds\Big)^{1/2} \tag{3.6}$$

with $K = \|h\|_\infty\sqrt{\nu^{-1}\alpha}\cdot(1+\nu^{-1}\cdot TG^2(\|h\|_\infty))^{1/2}$. Squaring (3.6) and using the inequality $(b+c)^2 \le 2b^2+2c^2$, we get

$$C_h(\mu_t,\sigma_t)^2 \le 2C_h(\mu_0,\sigma_0)^2 + 2K^2\int_0^t G^2(C_h(\mu_s,\sigma_s))\,ds.$$

If $\mu_0=\sigma_0$, then uniqueness follows immediately by the explicit integration. In the general case a Gronwall type inequality (see, for example, [8, Th. 27]) implies that

$$C_h(\mu_t,\sigma_t) \le (G_*^{-1}(G_*(2(C_h(\mu_0,\sigma_0))^2) - 2K^2t))^{1/2}$$

This completes the proof. □

A particular case $G(u)=u$ of this latter estimate is especially interesting:

COROLLARY 3.2. *Let μ and σ be two solutions to* (3.1) *as in Theorem* 3.1 *with $G(u)=u$. Then for some $N>0$ one has*

$$C_{h_{\lambda t}}(\mu_t,\sigma_t) \le \sqrt{2}C_h(\mu_0,\sigma_0)e^{Nt}.$$

In particular, if the drift is dissipative ($\lambda=0$) *or* $\lambda<0$, *then*

$$C_h(\mu_t,\sigma_t) \le \sqrt{2}C_h(\mu_0,\sigma_0)e^{Nt}.$$

In some cases estimate (2.6) enables us to establish existence of a solution to the nonlinear equation (3.1). To show this, consider $h(r)=\min\{|r|^p,1\}$ for some $p\ge 1$. Recall that in this case $C_h^{1/p}(\mu_t,\sigma_t)$ is a metric and turns the space of probability measures into a complete metric space. Moreover, convergence with respect to this metric is equivalent to weak convergence (see [6, Th. 1.1.9]).

COROLLARY 3.3. *Suppose there exists a function $V \in C^2(\mathbb{R}^d)$, $V \geq 1$ such that $V(x) \to +\infty$ as $|x| \to +\infty$ and there exists a positive function Λ on $[0, +\infty)$ such that*

$$(L_\mu V)(x, t) \leq \Lambda(\alpha(t))(1 + V(x))$$

for each $\alpha \in C^+[0, T]$, $\tau \in [0, T]$, each $(x, t) \in \mathbb{R}^d \times [0, T]$ and each $\mu \in M_{\tau,\alpha}(V)$. Assume that the coefficients in (3.1) *satisfy* (A1), (B1) *and* (B2) *with this function V. Assume also that $B(\sigma^n) \to B(\sigma)$ in $L^2(\mathbb{R}^d \times [0, T], d\sigma_s\, ds)$ as $n \to \infty$ if measures $\sigma^n(dx\, dt) = \sigma_t^n(dx)\, dt$ converge weakly to a measure $\sigma(dx\, dt) = \sigma_t(dx)\, dt$ on the strip $\mathbb{R}^d \times [0, T]$. Then, for every probability measure μ^* such that $V \in L^1(\mathbb{R}^d; \mu^*)$, there exists a* (*local*) *probability solution $\mu = (\mu_t)_{t \in [0,\tau]}$ to* (3.1) *with initial condition μ^*.*

EXAMPLE 3.2. Let $k(x, y)$ be a bounded function that is λ-dissipative in the first variable for every $y \in \mathbb{R}^d$. Let $Q(x, t)$ be a matrix satisfying (A1). Then the Cauchy problem (3.1) with

$$B(\mu, x, t) = \int k(x, y)\, d\mu_t(y)$$

satisfies all assumptions of Theorem 3.3 with $V(x) = 1 + |x|^2$ and any probability measure ν with finite second moment.

EXAMPLE 3.3. Let $V > 0$ be a convex C^2-function on $\mathbb{R}^d$. Let $g(x)$ be a λ-dissipative function such that $|g| \leq \sqrt{V}$. Let $Q(x, t)$ be a matrix satisfying (A1) and $k(y)$ be a nonnegative continuous bounded function. Then the Cauchy problem (3.1) with

$$B(\mu, x, t) = g(x) \int k(y)\, d\mu_t(y)$$

satisfies all assumptions of Theorem 3.3 with any probability measure ν that integrates V.

EXAMPLE 3.4. Fix $\alpha \in (0, 1)$ and a matrix Q satisfying (A1). Then the Cauchy problem (3.1) with

$$B(\mu, x, t) = -(|x|^{\alpha-1} x) * \mu_t$$

satisfies all assumptions of Theorem 3.3 with $V(x) = 1 + |x|^2$ and any probability measure ν with finite second moment (cf. [11, Proposition 2.1]).

PROOF OF COROLLARY 3.3. As above, without loss of generality we can assume that the drift term is dissipative. Let $\sigma \in M_{\tau,\alpha}(V)$ for some τ, α. Consider the equation

$$\partial_t \mu_t = \operatorname{trace}(Q(x, t) D^2 \mu_t) - \operatorname{div}(B(\sigma, x, t)\mu_t), \quad \mu_0 = \mu^*.$$

Note that the dissipativity of the drift ensures that it is bounded locally in (x, t). Hence under the assumptions of the theorem there exists a unique probability solution $\mu = (\mu_t)_{t \in [0, \tau]}$ in $M_\tau(V)$ (see [12, Theorem 3.6]). Therefore, the mapping $\Theta : M_{\tau,\alpha}(V) \to M_\tau(V)$ given by

$$\mu = \Theta(\sigma) \Leftrightarrow \partial_t \mu_t = \mathrm{trace}(Q(x,t) D^2 \mu_t) - \mathrm{div}(B(\sigma, x, t)\mu_t), \quad \mu_0 = \mu^*$$

is well-defined. It is obvious that the solutions to (3.1) are exactly the fixed points of the mapping Θ.

Let us define the subclass $N_{\tau,\alpha}(V)$ of the class $M_{\tau,\alpha}(V)$ as follows:

$$N_{\tau,\alpha}(V) := \left\{ \mu \in M_{\tau,\alpha}(V) : \left| \int \varphi(x)\, d(\mu_t - \mu_s) \right| \le K(\tau, \alpha, \varphi) \cdot |t - s| \ \forall \varphi \in C_0^\infty(\mathbb{R}^d) \right\},$$

where

$$K(\tau, \alpha, \varphi) := \sup\{|L_\mu \varphi(x,t)|, (x,t) \in \mathbb{R}^d \times [0, \tau], \mu \in M_{\tau,\alpha}(V)\}.$$

Obviously, $N_{\tau,\alpha}$ is a convex set. By [11, Corollary 4] there exist $\bar{\alpha}(t) > 0$ and $\bar{\tau} \in (0, T]$ such that $\Theta(N_{\bar{\tau},\bar{\alpha}}(V)) \subset N_{\bar{\tau},\bar{\alpha}}(V)$. Moreover, the class $N_{\bar{\tau},\bar{\alpha}}(V)$ is a compact set in the topology of weak convergence of measures on the strip $\mathbb{R}^d \times [0, \tau]$ by [11, Corollary 1]. Let us verify the continuity of the mapping Θ on $N_{\bar{\tau},\bar{\alpha}}(V)$. Suppose that a sequence $\sigma^n = (\sigma_t^n) \in N_{\bar{\tau},\bar{\alpha}}(V)$ converges weakly to $\sigma = (\sigma_t) \in N_{\bar{\tau},\bar{\alpha}}(V)$. Set $\mu^n := \Theta(\sigma^n)$, $\mu := \Theta(\sigma)$. Due to (2.17) we have (with $b = B(\sigma)$ and $b_n = B(\sigma_n)$ for shortness)

$$C_h(\mu_t^n, \mu_t) \le \left(\int_0^{\bar{\tau}} \int |b_n - b|^2 \, d\sigma_s \, ds \right)^{1/2} \times \left(1 + \int_0^{\bar{\tau}} \int |b_n - b|^2 \, d\sigma_s \, ds \right)^{1/2}.$$

Our conditions imply that the right-hand side tends to zero as $n \to \infty$. Hence μ_t^n converges to μ_t with respect to the metric $C_h^{1/p}$ and thus converges weakly. Let us show that μ^n converges to μ on the strip $\mathbb{R}^d \times [0, \bar{\tau}]$. Fix a continuous bounded function $\zeta(x,t)$. Then for each $t \in [0, \bar{\tau}]$ we have

$$\int \zeta(x,t) \, d\mu_t^n \to \int \zeta(x,t) \, d\mu_t, \quad n \to \infty.$$

Since the measures μ_t^n are probability measures and ζ is bounded, the integrals on the right-hand side are uniformly bounded and converge pointwise (with respect to $t \in [0, \bar{\tau}]$) to $\int \zeta(x,t)\, d\mu_t$. Therefore, Lebesgue's dominated convergence theorem ensures that

$$\int_0^{\bar{\tau}} \int \zeta(x,t) \, d\mu_t^n \, dt \to \int_0^{\bar{\tau}} \int \zeta(x,t) \, d\mu_t \, dt, \quad n \to \infty.$$

By definition this means that the sequence μ^n converges weakly to μ on the strip $\mathbb{R}^d \times [0, \bar{\tau}]$.

Summarizing, we have a continuous mapping Θ of the convex compact set $N_{\bar{\tau},\bar{\alpha}}(V)$ into itself. The Schauder fixed-point theorem ensures that there exists a fixed point of Θ in $N_{\bar{\tau},\bar{\alpha}}(V)$, i.e., there exists a solution $\mu = (\mu)_{t\in[0,\bar{\tau}]}$ to (3.1) with initial condition μ^*. □

ACKNOWLEDGMENTS. The work was partially supported by RFBR grants 14-01-00237 and 15-31-20082. The author is grateful to M. Zaal for fruitful discussions and to A. I. Nazarov for valuable comments and remarks.

REFERENCES

[1] L. AMBROSIO - N. GIGLI - G. SAVARÉ, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Lectures in Mathematics. Birkhäuser, Zürich, 2008.

[2] V. BOGACHEV - N. KRYLOV - M. RÖCKNER, *On regularity of transition probabilities and invariant measures of singular diffusions under minimal conditions*, Comm. PDEs 26 (2001), 11, 2037–2080.

[3] V. BOGACHEV - N. KRYLOV - M. RÖCKNER - S. SHAPOSHNIKOV, *Fokker–Planck–Kolmogorov equations*, vol. 207 of Mathematical Surveys and Monographs, AMS, 2015.

[4] V. BOGACHEV - M. RÖCKNER - S. SHAPOSHNIKOV, *On uniqueness problems related to the Fokker–Planck–Kolmogorov equation for measures*, Journal of Mathematical Sciences, 179 (2001), 1, 7–47.

[5] V. BOGACHEV - M. RÖCKNER - S. SHAPOSHNIKOV, *Distances between transition probabilities of diffusions and applications to nonlinear Fokker–Planck–Kolmogorov equations*, Journal of Functional Analysis 271 (2016), 1262–1300.

[6] V. BOGACHEV - A. KOLESNIKOV, *The Monge–Kantorovich problem: achievements, connections, and perspectives*, Russian Mathematical Surveys 67 (2012), 5, 785–890.

[7] R. DOBRUSHIN, *Vlasov equations*, Functional Analysis and Its Applications, 13 (1979), 2, 115–123.

[8] S. DRAGOMIR, *Some Gronwall type inequalities and applications*, Victoria University of Technology, Australia (2002).

[9] T. FUNAKI, *A certain class of diffusion processes associated with nonlinear parabolic equations*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 67 (1984), 3, 331–348.

[10] O. MANITA - M. ROMANOV - S. SHAPOSHNIKOV, *On uniqueness of solutions to nonlinear Fokker–Planck–Kolmogorov equations*, Nonlinear Analysis: Theory, Methods and Applications, 128 (2015), 199–226.

[11] O. MANITA - S. SHAPOSHNIKOV, *Nonlinear parabolic equations for measures*, St. Petersburg Math. J. 25 (2014), 43–62.

[12] O. MANITA - S. SHAPOSHNIKOV, *On the Cauchy problem for Fokker–Planck–Kolmogorov equations with potential terms on arbitrary domains*, Journal of Dynamics and Differential Equations, DOI:10.1007/s10884-015-9453-y.

[13] L. NATILE - M. PELETIER - G. SAVARÉ, *Contraction of general transportation costs along solutions to Fokker–Planck equations with monotone drifts*, Journal de Mathematiques Pures et Appliquees, 95 (2011), 1, 18–35.

[14] D. Stroock - S. R. S. Varadhan, *Multidimensional diffusion processes*, No. 233 in Grundlehren der Mathematischen Wissenschaften. Springer, Berlin, 1979. MR:532498. Zbl:0426.60069.

[15] C. Villani, *Topics in Optimal Transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, Rhode Island, 2003.

[16] M. M. Zaal, *On the role of cylindrical functions in Kantorovich duality*, Arxiv Preprint 1504.02600.

Received 9 March 2016,
and in revised form 21 January 2017.

Oxana A. Manita
Faculty of Mechanics and Mathematics
Lomonosov Moscow State University
119991 Moscow, Russia
o.manita@lambda.msu.ru