Journal of Functional Analysis 271 (2016) 1262–1300

Contents lists available at ScienceDirect

Journal of Functional Analysis

www.elsevier.com/locate/jfa

Distances between transition probabilities of diffusions and applications to nonlinear Fokker–Planck–Kolmogorov equations

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

[a] *National Research University Higher School of Economics, Moscow, Russia*
[b] *Fakultät für Mathematik, Universität Bielefeld, D-33501 Bielefeld, Germany*

A R T I C L E   I N F O

Article history:
Received 23 March 2015
Accepted 25 May 2016
Available online 30 May 2016
Communicated by Cédric Villani

MSC:
35K10
35K55
60J60

Keywords:
Fokker–Planck–Kolmogorov equation
Total variation distance
Kantorovich distance
Mean field games

A B S T R A C T

We estimate the total variation and Kantorovich distances between transition probabilities of two diffusions with different diffusion matrices and drifts via a natural quadratic distance between the drifts and diffusion matrices. Applications to nonlinear Fokker–Planck–Kolmogorov equations, optimal control and mean field games are given.

* Corresponding author.
E-mail addresses: vibogach@mail.ru (V.I. Bogachev), roeckner@math.uni-bielefeld.de (M. Röckner), starticle@mail.ru (S.V. Shaposhnikov).

http://dx.doi.org/10.1016/j.jfa.2016.05.016

1. Introduction

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

$$\partial_t \varrho_k(x,t) = \partial_{x_i}\partial_{x_j}(a_k^{ij}(x,t)\varrho_k(x,t)) - \partial_{x_i}(b_k^i(x,t)\varrho_k(x,t)), \quad k=1,2, \tag{1.1}$$

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

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

and for the total variation norm we obtain the estimate

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

In the general case we obtain quite comparable estimates under rather broad assumptions about our coefficients: the diffusion matrices are locally uniformly elliptic and locally Lipschitzian in space, the drifts are locally bounded, and either some mild integrability conditions are imposed or a certain Lyapunov function exists (an advantage of the latter condition is that it is expressed entirely in terms of the coefficients). In examples we give a number of effectively verified conditions. The principal novelty concerns the case of different diffusion matrices (see comments in Remark 1.5), but also the simpler case of the same diffusion matrix seems to be new. The main result is applied to nonlinear Fokker–Planck–Kolmogorov equations, optimal control and mean field games. Let us explain precisely our framework.

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 $b(x,t) = (b^i(x,t))_{i=1}^d \colon\ \mathbb{R}^d\times[0,T]\to\mathbb{R}^d$ is a Borel measurable mapping (called the drift coefficient). Suppose that b is locally bounded, i.e., for every ball $U\subset\mathbb{R}^d$, there is a number $B=B(U)\ge 0$ such that

$$|b(x,t)| \le B(U) \quad \forall\, x\in U,\ t\in[0,T],$$

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 $\lambda = \lambda(U) \geq 0$, $\alpha = \alpha(U) > 0$ and $m = m(U) > 0$ such that

$$|a^{ij}(x,t) - a^{ij}(y,t)| \leq \lambda |x-y|, \quad \alpha \cdot \mathrm{I} \leq A(x,t) \leq m \cdot \mathrm{I} \quad \forall\, x,y \in U,\ t \in [0,T].$$

We study solutions to the Cauchy problem

$$\partial_t \mu = L^*_{A,b}\mu, \quad \mu|_{t=0} = \nu, \tag{1.2}$$

where ν is a Borel probability measure on $\mathbb{R}^d$. A model example is given by the transition probabilities of a diffusion process, although we do not assume the existence of the associated diffusion.

We shall consider measures $\mu(dxdt) = \mu_t(dx)\,dt$ on $\mathbb{R}^d \times [0,T]$ given by a family 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

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

for every bounded Borel function f on $\mathbb{R}^d \times [0,T]$. Such a measure is called a solution to the Cauchy problem (1.2) if, for every function $\varphi \in C_0^\infty(\mathbb{R}^d)$, the equality

$$\int_{\mathbb{R}^d} \varphi\, d\mu_t = \int_{\mathbb{R}^d} \varphi\, d\nu + \int_0^t \int_{\mathbb{R}^d} L_{A,b}\varphi\, d\mu_s\, ds \tag{1.3}$$

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

It is known (see [11] or [12]) 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 (1.1) 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 (1.2) with coefficients A_μ, b_μ and A_σ, b_σ, respectively, and the same initial condition ν (the case of different initial conditions is addressed in Remarks 1.3 and 2.3 just to simplify the obtained estimates). We emphasize that in Sections 1–3 the subindices μ and σ do not mean that the coefficients of our equations depend on the solutions μ and σ, they only label the equations; however, in Section 4 we consider nonlinear equations whose coefficients depend on unknown solutions and in that case we write this dependence as $b(\mu)$ to avoid confusion with the previous sections.

Suppose throughout that

A_μ and A_σ satisfy Condition (H) *and b_μ and b_σ 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

$$h_\mu = (h^i_\mu)^d_{i=1}, \quad h_\sigma = (h^i_\sigma)^d_{i=1}, \quad h^i_\mu = b^i_\mu - \sum_{j=1}^d \partial_{x_j} a^{ij}_\mu, \quad h^i_\sigma = b^i_\sigma - \sum_{j=1}^d \partial_{x_j} a^{ij}_\sigma,$$

$$\Phi = \frac{(A_\mu - A_\sigma)\nabla\varrho_\sigma}{\varrho_\sigma} - (h_\mu - h_\sigma).$$

The latter mapping is crucial: the distances between μ_t and σ_t will be estimated through the $L^2(\sigma)$-norm of $A_\mu^{-1/2}\Phi$. Observe that in case of equal diffusion matrices we obtain just the difference of the drifts: $\Phi = b_\sigma - b_\mu$. In case of equal drifts and constant diffusion matrices, only the first term of this mapping appears.

Let $\|\cdot\|_{TV}$ denote the total variation norm on bounded measures. Recall that, given two probability measures μ_1 and μ_2 on $\mathbb{R}^d$ such that $\mu_1 = w\cdot\mu_2$, the entropy $H(\mu_1|\mu_2)$ is defined by the formula

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

provided that $w\log w \in L^1(\mu_2)$. If μ_1 and μ_2 are given by positive densities ϱ_1 and ϱ_2 such that $\varrho_1\log(\varrho_1/\varrho_2) \in L^1(\mathbb{R}^d)$, then $H(\mu_1|\mu_2)$ is the integral of $\varrho_1\log(\varrho_1/\varrho_2)$.

Let us formulate our main result.

Theorem 1.1. *Let $|A_\mu^{-1/2}\Phi| \in L^2(\mathbb{R}^d\times[0,T],\sigma)$. Suppose also that at least one of the following two conditions is fulfilled*:

(a) $(1+|x|)^{-2}|a^{ij}_\mu|$, $(1+|x|)^{-1}|b_\mu|$, $(1+|x|)^{-1}|\Phi| \in L^1(\mathbb{R}^d\times[0,T],\sigma)$;
(b) *there exist a nonnegative function $V\in C^2(\mathbb{R}^d)$ and a number $M\geq 0$ such that*

$$\lim_{|x|\to\infty} V(x) = +\infty, \quad L_{A_\mu,b_\mu}V \leq MV, \quad \frac{\langle\Phi,\nabla V\rangle}{1+V} \in L^1(\mathbb{R}^d\times[0,T],\sigma).$$

Then

$$H(\sigma_t|\mu_t) = \int_{\mathbb{R}^d} v\log v\,d\mu_t \leq \frac{1}{2}\int_0^t\int_{\mathbb{R}^d} \left|A_\mu^{-1/2}\Phi\right|^2 d\sigma_s\,ds. \tag{1.4}$$

Corollary 1.2. *Under the assumptions of the theorem, for every nonnegative Borel measurable function φ on $\mathbb{R}^d \times [0, T]$, we have*

$$\|\varphi(\mu_t - \sigma_t)\|_{TV}^2 \le \big(1 + \log \alpha(t)\big) \int_0^t \int_{\mathbb{R}^d} \left| \frac{A_\mu^{-1/2}(A_\mu - A_\sigma)\nabla \varrho_\sigma}{\varrho_\sigma} - A_\mu^{-1/2}(h_\mu - h_\sigma) \right|^2 d\sigma_s\, ds, \tag{1.5}$$

where

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

In particular, if $A_\mu = A_\sigma = A$, then

$$\|\varphi(\mu_t - \sigma_t)\|_{TV}^2 \le \big(1 + \log \alpha(t)\big) \int_0^t \int_{\mathbb{R}^d} \left| A^{-1/2}(b_\mu - b_\sigma) \right|^2 d\sigma_s\, ds,$$

and if $A_\mu = A_\sigma = \mathrm{I}$, then

$$\|\varphi(\mu_t - \sigma_t)\|_{TV}^2 \le \big(1 + \log \alpha(t)\big) \int_0^t \int_{\mathbb{R}^d} |b_\mu - b_\sigma|^2\, d\sigma_s\, ds.$$

If $b_\mu = b_\sigma = b$ and A_μ, A_σ do not depend on x, then

$$\|\varphi(\mu_t - \sigma_t)\|_{TV}^2 \le \big(1 + \log \alpha(t)\big) \int_0^t \int_{\mathbb{R}^d} \frac{\left|(A_\mu^{1/2} A_\sigma^{-1/2} - A_\mu^{-1/2} A_\sigma^{1/2}) A_\sigma^{1/2} \nabla \varrho_\sigma\right|^2}{\varrho_\sigma}\, dx\, ds.$$

Finally, in case $\varphi = 1$ these bounds hold with 1 *in place of* $1 + \log \alpha(t)$.

Remark 1.3. In the case of different initial conditions ν_μ and ν_σ the same reasoning applies (which will be noted in the proof below) and gives a bit longer estimates. In place of (1.4) we obtain

$$\int_{\mathbb{R}^d} v \log v\, d\mu_t \le \frac{1}{2} \int_0^t \int_{\mathbb{R}^d} \left| A_\mu^{-1/2} \Phi \right|^2 d\sigma_s\, ds + H(\nu_\sigma | \nu_\mu). \tag{1.6}$$

The extra term with $H(\nu_\sigma|\nu_\mu)$ will be added also to the integral in the right-hand side of (1.5). It is worth noting that this case does not cover the stationary case considered in [9] (for the unit diffusion matrix; the assumptions about drifts are recalled below), since the right-hand side of (1.6) involves the relative entropy $H(\nu_\sigma|\nu_\mu)$, which is the object of estimating in the stationary case.

It should be also noted that the global integrability conditions in Theorem 1.1 are imposed only on one of the two measures (certainly, the coefficients of both equations are involved in these conditions).

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

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

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 (sometimes mistakenly called the Wasserstein distance); see [10] or [49].

It is known that a bound on the entropy yields many other bounds, see, e.g., [30,49,50], where many additional references can be found. In the case $\varphi=1$ in Corollary 1.2 we obtain the usual total variation distance, hence the classical Pinsker–Csiszár–Kullback inequality (see, e.g., [6, 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

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

established in [24], 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 1.4. 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_μ and A_σ are different). If the norms of $A_\mu-A_\sigma$ and A_μ^{-1} are uniformly bounded, then, up to a constant factor, the right-hand side of (1.4) is estimated by the $L^2(\sigma)$-norms of $|b_\mu-b_\sigma|$, $|\nabla a_\mu^{ij}-\nabla a_\sigma^{ij}|$, and $|\nabla\varrho_\sigma|/\varrho_\sigma$. Let us recall some estimates of the $L^2(\sigma)$-norm of $\nabla\varrho_\sigma/\varrho_\sigma$ obtained in [14] (see also [13, Chapter 7] and [15]). Suppose that λ and α in (H1) can be chosen independent of U and that $|b_\sigma|\in L^2(\sigma)$. Assume also that the function $\Lambda(x):=\log\max(|x|,1)$ belongs to $L^2(\sigma)$ (which is true if, for example, $\langle b_\sigma(x,t),x\rangle\le C_1|x|^2\Lambda(x)+C_2$ with some constants C_1 and C_2 and $\Lambda\in L^2(\nu)$). If the initial distribution ν has finite entropy, i.e., possesses a density ϱ_ν such that $\log\varrho_\nu\in L^1(\nu)$, then for every $\tau<T$ we have

$$\int_0^\tau\int_{\mathbb{R}^d}\frac{|\nabla\varrho_\sigma(x,t)|^2}{\varrho_\sigma(x,t)}\,dx\,dt\le K<\infty,$$

where K is a number that depends on the $L^2(\sigma)$-norm of $|b_\sigma|$, the entropy of ν, and the bounds on the integrals of Λ against σ_t (see estimate (2.12) in [14] or (7.4.13) in [13, Chapter 7]). More precisely,

$$\int_0^\tau \int_{\mathbb{R}^d} \frac{|\nabla \varrho_\sigma|^2}{\varrho_\sigma}\, dx\, dt \le \alpha^{-2}\Bigl(\|b_\sigma\|_{L^2(\sigma)} + \sqrt{3}\lambda d^{5/2}\Bigr)^2$$
$$+ 2\log 2\alpha^{-1} + 2\alpha^{-1}\int_{\mathbb{R}^d} \varrho_\nu(x)\log \varrho_\nu(x)\, dx + 2\alpha^{-1}(d+1)\int_{\mathbb{R}^d} \varrho_\sigma(x,\tau)\Lambda(x)\, dx. \quad (1.7)$$

If the integrals of $\Lambda(x)$ against σ_t over $\mathbb{R}^d$ remain bounded as $t \to T$ (which holds, for example, if $\langle b_\sigma(x,t), x\rangle \le C_1|x|^2 + C_2$ with some constants C_1 and C_2 and $\Lambda \in L^1(\nu)$), then (1.7) is true with $\tau = T$. Therefore, there are efficient conditions in terms of the coefficients to verify that the right-hand side of our estimate is finite. Thus, in the previous situation we arrive at the following bound:

$$\|\varphi(\mu_t - \sigma_t)\|_{TV} \quad (1.8)$$
$$\le C(t) \sup_{x,t,i,j} \Bigl[|b_\mu(x,t) - b_\sigma(x,t)| + |a_\mu^{ij}(x,t) - a_\sigma^{ij}(x,t)| + |\nabla a_\mu^{ij}(x,t) - \nabla a_\sigma^{ij}(x,t)|\Bigr],$$

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

A modification of (1.7) is given by (2.13) in [14]: for almost all $\tau \in [0,T]$ one has

$$\int_0^\tau \int_{\mathbb{R}^d} \Bigl|\frac{A_\sigma^{1/2}\nabla \varrho_\sigma}{\varrho_\sigma}\Bigr|^2 d\sigma$$
$$\le \int_0^\tau \int_{\mathbb{R}^d} |A^{-1/2}h_\sigma|^2\, d\sigma + 2\int_{\mathbb{R}^d} [\varrho_\nu(x)\log \varrho_\nu(x) - \varrho_\sigma(x,\tau)\log \varrho_\sigma(x,\tau)]\, dx.$$

Remark 1.5. 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 [32,34–36] for the total variation distance between the distributions of ξ_1 and ξ_2 in the space $C[0,T]$. Assuming for simplicity that the drifts b_1 and b_2 are bounded and do not depend on t, we obtain by the Girsanov theorem that for any fixed t the distributions in $C[0,t]$ of the diffusions governed by the stochastic equations

$$d\xi_1(t) = \sqrt{2}dw_t - b_1(\xi_1(t))dt, \quad d\xi_2(t) = \sqrt{2}dw_t - b_2(\xi_2(t))dt$$

with $\xi_1(0)$ and $\xi_2(0)$ having distribution η are equivalent to the Wiener measure P and the corresponding Radon–Nikodym densities are given by

$$\varrho_i(w) = \exp\Bigl(\int_0^t b_i(w(s))dw_s - \frac{1}{2}\int_0^t |b_i(w(s)|^2 ds\Bigr).$$

This enables one to estimate the $L^1(P)$-norm of the function $\varrho_1(w) - \varrho_2(w)$, which yields an estimate on the total variation norm of the measure $\mu_t^1 - \mu_t^2$, where μ_t^i is the distribution of $\xi_i(t)$. However, in spite of this explicit expression, the derivation of the desired estimate is not trivial, and the Hellinger distance and the associated Hellinger processes are employed in the cited papers. Apparently, this method extends to multidimensional diffusions, but in this way it is impossible to deal with the case where the diffusion processes have different diffusions matrices, because in such a case their distributions in the functional space are typically mutually singular, as happens, e.g., if $A_1 = \mathrm{I}$ and $A_2 = 2\mathrm{I}$ (in the one-dimensional case with different analytic A_1 and A_2, they are always mutually singular, see [7, Section 4.4]).

Remark 1.6. It should be also noted that if we are interested only in W_p-estimates, then a simpler approach is possible. Let $\mu_t\,dt$ and $\sigma_t\,dt$ be solutions on $[0,T]\times\mathbb{R}^d$ to the continuity equations

$$\partial_t\mu_t + \operatorname{div}(b\mu_t) = 0, \quad \partial_t\sigma_t + \operatorname{div}(h\sigma_t) = 0$$

with initial distribution ν. Suppose that b and h are Lipschitz mappings on $\mathbb{R}^d$ with constant λ. Then $\mu_t = \nu\circ x_t^{-1}$ and $\sigma_t = \nu\circ y_t^{-1}$, where

$$\dot{x}_t(z) = b(x_t(z)), \quad x_0(z) = z, \quad \dot{y}_t(z) = h(y_t(z)), \quad y_0(z) = z.$$

We observe that

$$\begin{aligned}\frac{d}{dt}\frac{|x_t - y_t|^p}{p} &\le |b(x_t) - b(y_t)|\,|x_t - y_t|^{p-1} + |b(y_t) - h(y_t)|\,|x_t - y_t|^{p-1}\\ &\le \Bigl(\lambda + \frac{p-1}{p}\Bigr)|x_t - y_t|^p + \frac{1}{p}|b(y_t) - h(y_t)|^p.\end{aligned}$$

Therefore,

$$|x_t - y_t|^p \le \int_0^t e^{c_p(t-s)}|b(y_s) - h(y_s)|^p\,ds, \quad c_p = p\lambda + p - 1.$$

From the definition of the metric W_p we find that

$$W_p^p(\mu_t,\sigma_t) \le \int |x_t - y_t|^p\,d\nu \le \int_0^t e^{c_p(t-s)}\int |b(y_s) - h(y_s)|^p\,d\nu\,ds.$$

Since $\sigma_t = \nu \circ y_t^{-1}$, we arrive at the inequality

$$W_p(\mu_t, \sigma_t) \le C(\lambda, T)\Bigl(\int_0^t \int |b(y) - h(y)|^p \, d\sigma_s \, ds\Bigr)^{1/p},$$

where $C(\lambda, T) = e^{c_p T/p}$.

In a similar manner one can obtain upper bounds for solutions $\mu_t \, dt$ and $\sigma_t \, dt$ to the Fokker–Planck–Kolmogorov equation

$$\partial_t \mu_t = \Delta \mu_t - \operatorname{div}(b(x)\mu_t), \quad \partial_t \sigma_t = \Delta \sigma_t - \operatorname{div}(h(x)\sigma_t)$$

with initial distribution ν. Suppose again that b and h are Lipschitzian with constant λ. Consider the solutions x_t and y_t to the stochastic equations

$$dx_t = \sqrt{2} dw_t + b(x_t) \, dt, \quad dy_t = \sqrt{2} dw_t + h(y_t) \, dt$$

with the same initial distribution ν. We observe that $d(x_t - y_t) = (b(x_t) - h(y_t)) \, dt$. Repeating the previous reasoning we obtain

$$W_p^p(\mu_t, \sigma_t) \le \mathbb{E}|x_t - y_t|^p \le C(\lambda, T) \int_0^t \int |b(y) - h(y)|^p \, d\sigma_t.$$

On the other hand, if we are interested only in the total variation norm, then another approach is possible (see Remarks 2.2 and 2.3).

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., [1,4,21–23,25,26,28,42,43,45]. The principal novelty of our estimates is that they compare diffusions with different drifts or even different diffusion matrices, not with different initial distributions. Analogous results for stationary Fokker–Planck–Kolmogorov equations have been obtained in [9], where in the case of the unit diffusion matrix and the same assumptions about drifts (which a bit simplify due to the equalities $A_\mu = A_\nu = \mathrm{I}$ and $\Phi = b_\sigma - b_\mu$), it is proved that

$$\int_{\mathbb{R}^d} \frac{|\nabla v|^2}{v} \, d\mu \le \int_{\mathbb{R}^d} |b_\mu - b_\nu|^2 \, d\nu.$$

If, in addition, both solutions have second moments and, moreover, μ satisfies the log-Sobolev inequality with constant C, then

$$W_2(\mu, \nu)^2 \le \frac{C^2}{4} \int_{\mathbb{R}^d} |b_\mu - b_\nu|^2 \, d\nu, \quad \|\mu - \nu\|^2 \le \frac{C}{2} \int_{\mathbb{R}^d} |b_\mu - b_\nu|^2 \, d\nu.$$

In terms of drifts, the log-Sobolev inequality with constant $C = 2/\kappa$ is ensured by the estimate $\langle b(x) - b(y), x - y\rangle \le -\kappa|x - y|^2$. As we see, no Sobolev inequality is needed in the present parabolic case.

The proof of the main theorem is given in Section 2; Section 3 contains some additional corollaries and examples. In Section 4 some applications to nonlinear Fokker–Planck–Kolmogorov equations are considered. Also applications to differentiability of solutions with respect to a parameter, optimal control and mean field games are given.

2. Proof of the main result

Informally, our proof is this: we multiply the equation by $f = v \log v - v$, integrate by parts, apply the Cauchy inequality and discard certain terms in the obtained inequality. However, a rigorous justification involves some technicalities.

The proof of Theorem 1.1 is based on the two lemmas below; the corollary then follows from the next estimate established in [24]: given two probability measures μ and $\sigma = v\cdot\mu$ on $\mathbb{R}^d$ and a Borel function $\varphi \ge 0$, we have

$$\|\varphi(\mu - \sigma)\|_{TV}^2 \le 2\biggl(1 + \log\Bigl(\int_{\mathbb{R}^d} e^{\varphi^2}\, d\mu\Bigr)\biggr) \int_{\mathbb{R}^d} v \log v\, d\mu. \tag{2.1}$$

It should be noted that a bit less compact, but stronger estimate is proved in [24]:

$$\|\varphi(\mu - \sigma)\|_{TV} \le \Bigl(\frac{3}{2} + \log \int_{\mathbb{R}^d} e^{2\varphi}\, d\mu\Bigr)\Bigl(H(\sigma, \mu)^{1/2} + \frac{1}{2} H(\sigma, \mu)\Bigr). \tag{2.2}$$

This estimate can be used in place of (2.1) in the proof of Corollary 1.2, which will result in longer expressions in the corresponding estimates, but the function α involved in those estimates will involve 2φ in place of φ^2. Note that for $\varphi = 1$ the bound from [24] gives the extra factor 2 on the right as compared to the Pinsker–Csiszár–Kullback inequality.

The next lemma needed in the proof of the main theorem can be of independent interest (recall that we admit arbitrary initial probability distributions).

Lemma 2.1. *For every ball U we have*

$$\lim_{\tau\to 0} \int_U |\varrho_\mu(x, \tau) - \varrho_\sigma(x, \tau)|\, dx = 0.$$

Proof. Let U' be an open ball containing the closure of U and let $\varphi \in C_0^\infty(U')$ be such that $\varphi \ge 0$, $\varphi(x) = 1$ on U, and $|\nabla\varphi(x)|\varphi(x)^{-1/2} \le C(\varphi)$. We can assume that the functions a_μ^{ij} are extended outside of $U' \times [0, T]$ to all of $\mathbb{R}^{d+1}$ in such a way that the extended functions a_μ^{ij} are bounded, the matrix A_μ is uniformly nondegenerate and Lipschitz in x with a constant independent of t. Below we deal with such A_μ.

Let $\psi_k \in C_0^\infty(\mathbb{R}^d)$, $|\psi_k| \le 1$ and $\psi_k(x) \to \operatorname{sgn}(\varrho_\mu(x,\tau) - \varrho_\sigma(x,\tau))$ as $k \to \infty$ for almost all $x \in U'$. Let us consider the Cauchy problem

$$\partial_t f + a_\mu^{ij}\partial_{x_i}\partial_{x_j} f = 0, \quad f|_{t=\tau} = \psi_k.$$

By [33, Theorem 2.8] there is a solution f_k on $(-1,\tau]$ such that f_k is continuous on $\mathbb{R}^d \times (-1,\tau]$, for each $t \in (-1,\tau)$ we have $f_k(\,\cdot\,,t) \in C^2(\mathbb{R}^d)$ and the Sobolev derivative $\partial_t f_k$ has a modification that is locally bounded and continuous in x uniformly in t. By the maximum principle $|f_k| \le 1$. We observe that in the sense of distributions

$$\partial_t \varrho_\sigma = \partial_{x_i}\partial_{x_j}(a_\mu^{ij}\varrho_\sigma) - \partial_{x_i}((b_\mu^i + \Phi^i)\varrho_\sigma).$$

Hence for the difference $\eta = \varrho_\mu - \varrho_\sigma$ we have (in the sense of distributions)

$$\partial_t \eta = \partial_{x_i}\partial_{x_j}(a_\mu^{ij}\eta) - \partial_{x_i}(b_\mu^i \eta) + \partial_{x_i}(\Phi^i \varrho_\sigma)$$

and for each continuous bounded function g with compact support we have

$$\lim_{t\to 0}\int g(x)\eta(x,t)\,dx = \lim_{t\to 0}\int g(x)\varrho_\mu(x,t)\,dx - \lim_{t\to 0}\int g(x)\varrho_\sigma(x,t)\,dx = 0.$$

Note that $\|\eta(\,\cdot\,,t)\|_{L^1(\mathbb{R}^d)} \le 2$. We need a bound on the expression

$$\int_0^\tau \int |A_\mu^{1/2}\nabla f_k|^2 \varphi(\varrho_\mu + \varrho_\sigma)\,dx\,ds.$$

By virtue of the equation for ϱ_μ and the stated properties of f_k, writing the equation for f_k^2, we find that

$$\int \varphi(x)\psi_k(x)^2\varrho_\mu(x,\tau)\,dx = 2\int_0^\tau \int |A_\mu^{1/2}\nabla f_k|^2 \varphi\varrho_\mu\,dx\,ds$$
$$+ \int_0^\tau \int \Big[f_k^2 L_{A_\mu,b_\mu}\varphi + 4f_k\langle A_\mu\nabla f_k, \nabla\varphi\rangle + 2f_k\langle b_\mu, \nabla f_k\rangle\varphi\Big]\varrho_\mu\,dx\,ds. \quad (2.3)$$

By the Cauchy inequality for every $\gamma > 0$ we have

$$4|\langle A_\mu\nabla f_k, \nabla\varphi\rangle| + 2|\langle b_\mu, \nabla f_k\rangle|\varphi$$
$$\le 3\gamma|A_\mu^{1/2}\nabla f_k|^2\varphi + \gamma^{-1}\Big(2|A_\mu^{1/2}\nabla\varphi|^2\varphi^{-1} + |A_\mu^{-1/2}b_\mu|^2\varphi\Big).$$

Since $|\varphi| \le 1$ and $|\psi_k| \le 1$, one has

$$\int \varphi(x)\psi_k(x)^2 \varrho_\mu(x,\tau)\,dx \le 1.$$

Applying these estimates in (2.3) we obtain

$$(2-3\gamma)\int_0^\tau \int |A_\mu^{1/2}\nabla f_k|^2 \varphi\varrho_\mu\,dx\,ds$$
$$\le 1 + \int_0^\tau \int \Big[|L_{A_\mu,b_\mu}\varphi| + \gamma^{-1}(2|A_\mu^{1/2}\nabla\varphi|^2\varphi^{-1} + |A_\mu^{-1/2}b_\mu|^2\varphi)\Big]\varrho_\mu\,dx\,ds. \quad (2.4)$$

Setting $\gamma = 1/3$ in (2.4) we obtain the estimate

$$\int_0^\tau \int |\sqrt{A_\mu}\nabla f|^2 \varphi\varrho_\mu\,dx\,ds \le 1 + C_1(\tau),$$

where

$$C_1(\tau) = \int_0^\tau \int \Big[|L_{A_\mu,b_\mu}\varphi| + 6|A_\mu^{1/2}\nabla\varphi|^2\varphi^{-1} + 3|A_\mu^{-1/2}b_\mu|^2\varphi\Big]\varrho_\mu\,dx\,ds$$

is independent of k. We observe that $\lim\limits_{\tau\to 0} C_1(\tau) = 0$. Similarly we obtain the estimate

$$\int_0^\tau \int |A_\mu^{1/2}\nabla f_k|^2 \varphi\varrho_\sigma\,dx\,ds \le 1 + C_2(\tau),$$

where

$$C_2(\tau) = \int_0^\tau \int \Big[|L_{A_\mu,b_\mu}\varphi| + 8|A_\mu^{1/2}\nabla\varphi|^2\varphi^{-1}$$
$$+ 4|A_\mu^{-1/2}b_\mu|^2\varphi + 4|A_\mu^{-1/2}\nabla\Phi|^2\varphi + |\Phi|\,|\nabla\varphi|\Big]\varrho_\sigma\,dx\,ds$$

and $\lim\limits_{\tau\to 0} C_2(\tau) = 0$. Hence we arrive at the inequality

$$\int_0^\tau \int |A_\mu^{1/2}\nabla f_k|^2 \varphi(\varrho_\mu + \varrho_\sigma)\,dx\,ds \le 2 + C_1(\tau) + C_2(\tau).$$

By virtue of the equation for η and the stated properties of f_k we have

$$\int \varphi(x)\psi_k(x)\eta(x,\tau)\,dx$$
$$= \int_0^\tau \int \Big[f_k L_{A_\mu,b_\mu}\varphi + 2\langle A_\mu \nabla f_k, \nabla\varphi\rangle + \langle b_\mu, \nabla f_k\rangle\varphi\Big]\eta\,dx\,ds$$
$$- \int_0^\tau \int \Big[\langle \Phi, \nabla f_k\rangle \varphi \varrho_\sigma + \langle \Phi, \nabla\varphi\rangle f_k \varrho_\sigma\Big]\,dx\,ds. \qquad (2.5)$$

Note that

$$2|\langle A_\mu \nabla f_k, \nabla\varphi\rangle| + |\langle b_\mu, \nabla f_k\rangle\varphi| \le \Big(2|A_\mu^{1/2}\nabla\varphi|\varphi^{-1/2} + |A_\mu^{-1/2}b_\mu|\varphi^{1/2}\Big)|A_\mu^{1/2}\nabla f_k|\varphi^{1/2}.$$

By the Cauchy inequality we have

$$\int_0^\tau \int \Big[2|\langle A_\mu \nabla f_k, \nabla\varphi\rangle| + |\langle b_\mu, \nabla f_k\rangle\varphi|\Big]\eta\,dx\,ds$$
$$\le \Big(\int_0^\tau \int \Big(2|A_\mu^{1/2}\nabla\varphi|\varphi^{-1/2} + |A_\mu^{-1/2}b_\mu|\varphi^{1/2}\Big)^2 |\eta|\,dx\,ds\Big)^{1/2}$$
$$\times \Big(\int_0^\tau \int |A_\mu^{1/2}\nabla f_k|^2 \varphi|\eta|\,dx\,ds\Big)^{1/2}.$$

Similarly we obtain the estimate

$$\int_0^\tau \int |\langle \Phi, \nabla f_k\rangle|\varphi\varrho_\sigma\,dx\,ds$$
$$\le \Big(\int_0^\tau \int |A_\mu^{-1/2}\Phi|^2 \varphi\varrho_\sigma\Big)^{1/2} \Big(\int_0^\tau \int |A_\mu^{1/2}\nabla f_k|^2 \varphi\varrho_\sigma\,dx\,ds\Big)^{1/2}.$$

Substituting these estimates in (2.5) we find that

$$\int \varphi(x)\psi_k(x)\eta(x,\tau)\,dx \le \int_0^\tau \int |L_{A_\mu,b_\mu}\varphi|\,|\eta|\,dx\,ds$$
$$+ (2 + C_1(\tau) + C_2(\tau))^{1/2}\Big(\int_0^\tau \int \Big[2|A_\mu^{-1/2}\nabla\varphi|\varphi^{-1/2} + |A_\mu^{-1/2}b_\mu|\varphi^{1/2}\Big]^2 |\eta|\,dx\,ds\Big)^{1/2}$$
$$+ (1 + C_2(\tau))^{1/2}\Big(\int_0^\tau \int |A_\mu^{-1/2}\Phi|^2 \varphi\varrho_\sigma\,dx\,ds\Big)^{1/2} + \int_0^\tau \int |\Phi|\,|\nabla\varphi|\varrho_\sigma\,dx\,ds.$$

The last inequality can be written as

$$\int \varphi(x)\psi_k(x)\eta(x,\tau)\,dx \le C_3(\tau), \quad \lim_{\tau\to 0} C_3(\tau) = 0,$$

where $C_3(\tau)$ does not depend on f. Letting $k\to\infty$ we arrive at the inequality

$$\int \varphi(x)|\eta(x,\tau)|\,dx \le C_3(\tau),$$

which completes the proof of the lemma. □

Remark 2.2. Developing the method of proving this lemma, one can also obtain (under somewhat different assumptions about the coefficients) certain estimates for the total variation distance between solutions to two Fokker–Planck–Kolmogorov equations analogously to Corollary 1.2. We explain this approach, which is a version of Holmgren's method, in the following simplified situation. Suppose that the coefficients are sufficiently regular and bounded. Let ϱ_μ and ϱ_σ be solutions to two Fokker–Planck–Kolmogorov equations as above. Their difference $\eta = \varrho_\mu - \varrho_\sigma$ satisfies the equation

$$\partial_t\eta = \partial_{x_i}\partial_{x_j}(a^{ij}_\mu \eta) - \partial_{x_i}(b^i_\mu\eta) + \partial_{x_i}(\Phi^i\varrho_\sigma).$$

As in the lemma, we take the solution to the Cauchy problem

$$\partial_t f + a^{ij}_\mu\partial_{x_i}\partial_{x_j} f + b^i_\mu\partial_{x_i} f = 0, \quad f|_{t=\tau} = \psi,$$

where ψ is a smooth function with compact support and $|\psi|\le 1$. We have

$$\int \psi(x)\eta(x,\tau)\,dx = \int_0^\tau \langle\Phi,\nabla f\rangle\varrho_\sigma\,dx\,ds.$$

For estimating the gradient of f we can write a similar equality for f^2:

$$\int \psi(x)^2\varrho_\sigma(x,\tau)\,dx = \int_0^\tau\int \Big[2|A_\mu^{1/2}\nabla f|^2 + 2f\langle\Phi,\nabla f\rangle\Big]\varrho_\sigma\,dx\,ds.$$

Since $|f|\le 1$ and

$$2|\langle\Phi,\nabla f\rangle| \le |A_\mu^{-1/2}\Phi|^2 + |A_\mu^{1/2}\nabla f|^2,$$

we arrive at the estimate

$$\int_0^\tau\int |A_\mu\nabla f|^2\varrho_\sigma\,dx \le 1 + \int_0^\tau\int |A_\mu^{-1/2}\Phi|^2\varrho_\sigma\,dx.$$

Using this estimate and the Cauchy inequality, we obtain

$$\int \psi(x)\eta(x,\tau)\,dx \le \Bigl(\int_0^\tau \int |A_\mu^{-1/2}\Phi|^2 \varrho_\sigma\,dx\Bigr)^{1/2} \Bigl(1+\int_0^\tau \int |A_\mu^{-1/2}\Phi|^2 \varrho_\sigma\,dx\Bigr)^{1/2}.$$

Since ψ was arbitrary, this gives a bound for $\|\eta(\,\cdot\,,\tau)\|_{L^1(\mathbb{R}^d)}$.

It should be noted that such estimates are actually obtained in [39] (see Theorem 3.1 and Remark 5.5) in the study of uniqueness of solutions to nonlinear Fokker–Planck–Kolmogorov equations. Unlike our main result, here we obtain a bound only on the total variation norm, in addition, such estimates require global bounds on the coefficients. As shown in [38], combining the Holmgren method and estimates of the square of the gradient of the solution to the adjoint equation of the Fokker–Planck–Kolmogorov equation, one can obtain estimates for the Kantorovich distance between different solutions for more general cost functions.

Remark 2.3. Applying the reasoning from the previous remark we can obtain estimates in the case where μ and σ satisfy the same equation, but with different initial conditions ν_μ and ν_σ. Let a^{ij}, b^i be bounded on $U\times[0,T]$ for every ball U. Assume that the matrix A satisfies condition (H) and

$$(1+|x|)^{-2}|a^{ij}(x,t)| + (1+|x|)^{-1}|b^i(x,t)| \in L^1(\mathbb{R}^d\times[0,T],\mu+\sigma). \tag{2.6}$$

Then

$$\|\mu_t-\sigma_t\|_{TV} \le \|\nu_\mu-\nu_\sigma\|_{TV}.$$

Actually this estimate was established in [18] and [13, Theorem 9.3.6] in the proof of uniqueness (see also [46]).

Moreover, let now μ and σ satisfy different equations with different initial conditions, where the coefficients satisfy condition (2.6) and condition (i) in Theorem 1.1. Combining the last estimate with estimates in Corollary 1.2, we obtain

$$\|\mu_t-\sigma_t\|_{TV} \le \|\nu_\mu-\nu_\sigma\|_{TV} + \int_0^t \int_{\mathbb{R}^d} \Bigl|A_\mu^{-1/2}\Phi\Bigr|^2\,d\sigma_s\,ds.$$

Unlike (1.6), this is a bound only on the total variation norm.

We need also the following auxiliary result.

Lemma 2.4.

(i) *Let $f\in C^2(0,+\infty)$. Then, for any function $\psi\in C_0^\infty(\mathbb{R}^d)$ and any compact interval $[\tau,t]\subset(0,T)$, we have*

$$\int_{\mathbb{R}^d} f(v(x,t))\psi(x)\varrho_\mu(x,t)\,dx + \int_\tau^t \int_{\mathbb{R}^d} |A_\mu^{1/2}\nabla v|^2 f''(v)\psi\varrho_\mu\,dx$$

$$= \int_{\mathbb{R}^d} f(v(x,\tau))\psi(x)\,\varrho_\mu(x,\tau)dx + \int_\tau^t \int_{\mathbb{R}^d} f(v)L_{A_\mu,b_\mu}\psi\varrho_\mu\,dx\,ds$$

$$+ \int_\tau^t \int_{\mathbb{R}^d} \big[\langle\Phi,\nabla v\rangle f''(v)\psi + \langle\Phi,\nabla\psi\rangle f'(v)\big] v\varrho_\mu\,dx\,ds. \qquad (2.7)$$

(ii) *If* $f'' \geq 0$, $\sup_t\big[|f'(t)| + |tf''(t)|\big] < \infty$ *and* $\psi \geq 0$, *then*

$$\int_{\mathbb{R}^d} f(v(x,t))\psi(x)\varrho_\mu(x,t)\,dx + \frac{1}{2}\int_0^t \int_{\mathbb{R}^d} |A_\mu^{1/2}\nabla v|^2 f''(v)\psi\varrho_\mu\,dx$$

$$\leq f(1)\int_{\mathbb{R}^d} \psi\,d\nu + \int_0^t \int_{\mathbb{R}^d} f(v)L_{A_\mu,b_\mu}\psi\varrho_\mu\,dx\,ds$$

$$+ \frac{1}{2}\int_0^t \int_{\mathbb{R}^d} |A_\mu^{-1/2}\Phi|^2 f''(v)\psi v^2\varrho_\mu\,dx\,ds + \int_0^t \int_{\mathbb{R}^d} \langle\Phi,\nabla\psi\rangle f'(v)v\varrho_\mu\,dx\,ds. \qquad (2.8)$$

In particular,

$$\int_{\mathbb{R}^d} f(v(x,t))\psi(x)\varrho_\mu(x,t)\,dx \leq f(1)\int_{\mathbb{R}^d} \psi\,d\nu + \int_0^t \int_{\mathbb{R}^d} f(v)L_{A_\mu,b_\mu}\psi\varrho_\mu\,dx\,ds$$

$$+ \frac{1}{2}\int_0^t \int_{\mathbb{R}^d} |A_\mu^{-1/2}\Phi|^2 f''(v)\psi v^2\varrho_\mu\,dx\,ds + \int_0^t \int_{\mathbb{R}^d} \langle\Phi,\nabla\psi\rangle f'(v)v\varrho_\mu\,dx\,ds. \qquad (2.9)$$

In case of different initial conditions ν_μ *and* ν_σ *such that* $\nu_\sigma \ll \nu_\mu$ *the first terms in the right-hand sides of the last two relations will be replaced by* $\int_{\mathbb{R}^d} \psi f\Big(\frac{d\nu_\sigma}{d\nu_\mu}\Big)d\nu_\mu$.

Proof. We start with (2.7). Suppose first that a_μ^{ij}, a_σ^{ij}, b_μ^i, b_σ^i are infinitely differentiable in x on $\mathbb{R}^d \times (0,T)$, A_μ and A_σ are nondegenerate (but no uniform Lipschitzness and uniform ellipticity up to $t = 0$ is assumed at this stage). The assumption that we deal with smooth coefficients is employed in the calculations below and ensures the existence of the regarded integrals. Recall that both densities ϱ_μ and ϱ_ν are continuous and positive on the sets of the form $\{x\colon\ |x| \leq R\} \times (0,T)$, so that v is also continuous on these sets. Hence v is bounded on $\operatorname{supp}\psi \times [\tau,t]$. Note that

$$\partial_t \varrho_\mu = L^*_{A_\mu, b_\mu} \varrho_\mu \quad \text{and} \quad \partial_t \varrho_\sigma = L^*_{A_\mu, b_\mu} \varrho_\sigma - \operatorname{div}(\Phi \varrho_\sigma).$$

Our assumptions about the coefficients enable us to write $L^*_{A_\mu, b_\mu} \xi$ directly just as a usual operator $\partial_{x_i} \partial_{x_j} (a^{ij}_\mu \xi) - \partial_{x_i} (b^i_\mu \xi)$ with summation over repeated indices. With respect to t both densities are absolutely continuous.

For any $\xi, \eta \in C^\infty(\mathbb{R}^d \times (0,T))$ and $\varphi \in C^\infty(\mathbb{R})$ we have by direct calculations

$$L^*_{A_\mu, b_\mu}(\varphi(\xi)) = \varphi'(\xi) L^*_{A_\mu, b_\mu} \xi + \varphi''(\xi) \langle A_\mu \nabla \xi, \nabla \xi \rangle + (\xi \varphi'(\xi) - \varphi(\xi)) \operatorname{div} h_\mu,$$
$$L^*_{A_\mu, b_\mu}(\xi \cdot \eta) = \eta L^*_{A_\mu, b_\mu} \xi + \xi L^*_{A_\mu, b_\mu} \eta + 2 \langle A_\mu \nabla \xi, \nabla \eta \rangle + \xi \eta \operatorname{div} h_\mu.$$

Multiplying the equation $\partial_t \varrho_\mu = L^*_{A_\mu, b_\mu} \varrho_\mu$ by v and subtracting the obtained relation from the equation $\partial_t \varrho_\sigma = L^*_{A_\mu, b_\mu} \varrho_\sigma - \operatorname{div}(\Phi \varrho_\sigma)$, we arrive at the equation

$$\varrho_\mu \partial_t v = \varrho_\mu L^*_{A_\mu, b_\mu} v + 2 \langle A_\mu \nabla \varrho_\mu, \nabla v \rangle + \varrho_\mu v \operatorname{div} h_\mu - \operatorname{div}(\Phi \varrho_\sigma).$$

Multiplying this by $f'(v)$ and taking into account the equalities

$$\partial_t f(v) = f'(v) \partial_t v, \quad \nabla f(v) = f'(v) \nabla v,$$

we obtain that

$$\varrho_\mu \partial_t (f(v)) = \varrho_\mu f'(v) L^*_{A_\mu, b_\mu} v + 2 \langle A_\mu \nabla \varrho_\mu, \nabla f(v) \rangle + \varrho_\mu v f'(v) \operatorname{div} h_\mu - f'(v) \operatorname{div}(\Phi \varrho_\sigma).$$

Since

$$f'(v) L^*_{A_\mu, b_\mu} v = L^*_{A_\mu, b_\mu} f(v) - f''(v) \langle A_\mu \nabla v, \nabla v \rangle - (v f'(v) - f(v)) \operatorname{div} h_\mu,$$

we have

$$\varrho_\mu \partial_t (f(v)) = \varrho_\mu L^*_{A_\mu, b_\mu} f(v) + 2 \langle A_\mu \nabla \varrho_\mu, \nabla f(v) \rangle$$
$$+ \varrho_\mu f(v) \operatorname{div} h_\mu - \varrho_\mu f''(v) \langle A_\mu \nabla v, \nabla v \rangle - f'(v) \operatorname{div}(\Phi \varrho_\sigma).$$

Summing up the last equation and $f(v) \partial_t \varrho_\mu = f(v) L^*_{A_\mu, b_\mu} \varrho$, we find that

$$\partial_t (\varrho_\mu f(v)) = L^*_{A_\mu, b_\mu} (\varrho_\mu f(v)) - \varrho_\mu f''(v) \langle A_\mu \nabla v, \nabla v \rangle - f'(v) \operatorname{div}(\Phi \varrho_\sigma).$$

Multiplying this equation by the function ψ and integrating, we arrive at the equality

$$\int_\tau^t \int_{\mathbb{R}^d} \partial_t (f(v) \varrho_\mu) \psi \, dx \, dt + \int_\tau^t \int_{\mathbb{R}^d} \varrho_\mu \langle A_\mu \nabla v, \nabla v \rangle f''(v) \psi \, dx \, d\tau$$
$$= \int_\tau^t \int_{\mathbb{R}^d} \psi L^*_{A_\mu, b_\mu} (\varrho_\mu f(v)) \, dx \, ds - \int_\tau^t \int_{\mathbb{R}^d} \psi f'(v) \operatorname{div}(\Phi \varrho_\sigma) \, dx \, ds.$$

Applying the Newton–Leibniz formula, we obtain

$$\int_\tau^t \int_{\mathbb{R}^d} \partial_t (f(v)\varrho_\mu)\psi \, dx \, dt = \int_{\mathbb{R}^d} f(v(x,t))\varrho_\mu(x,t)\psi(x)\, dx - \int_{\mathbb{R}^d} f(v(x,\tau))\psi(x)\varrho_\mu(x,\tau)\, dx.$$

Since $L^*_{A_\mu,b_\mu}$ is the adjoint to L_{A_μ,b_μ}, one has

$$\int_\tau^t \int_{\mathbb{R}^d} \psi L^*_{A_\mu,b_\mu}(\varrho_\mu f(v))\, dx\, ds = \int_\tau^t \int_{\mathbb{R}^d} \varrho_\mu f(v) L_{A_\mu,b_\mu}\psi \, dx \, ds.$$

In addition, one has

$$-\int_\tau^t \int_{\mathbb{R}^d} \psi f'(v) \mathrm{div}(\Phi\varrho_\sigma)\, dx\, ds = \int_\tau^t \int_{\mathbb{R}^d} \big[\langle \Phi, \nabla v\rangle f''(v)\psi\varrho_\sigma + f'(v)\langle\Phi,\nabla\psi\rangle\varrho_\sigma\big]\, dx\, ds.$$

Therefore, we obtain the following equality:

$$\int_{\mathbb{R}^d} f(v(x,t))\varrho_\mu(x,t)\psi(x)\, dx + \int_\tau^t \int_{\mathbb{R}^d} \varrho_\mu \langle A_\mu \nabla v, \nabla v\rangle f''(v)\psi\, dx\, d\tau$$
$$= \int_{\mathbb{R}^d} f(v(x,\tau))\psi(x)\varrho_\mu(x,\tau)\, dx + \int_\tau^t \int_{\mathbb{R}^d} \varrho_\mu f(v) L_{A_\mu,b_\mu}\psi\, dx\, ds$$
$$+ \int_\tau^t \int_{\mathbb{R}^d} \big[\langle \Phi, \nabla v\rangle f''(v)\psi\varrho_\sigma + f'(v)\langle\Phi,\nabla\psi\rangle\varrho_\sigma\big]\, dx\, ds,$$

which is the desired identity. As noted above, in case of different initial conditions the integral against ν must be replaced by the indicated expression.

In the general case, where the coefficients are not smooth, it suffices to show that our solutions can be obtained as limits of solutions (with convergent derivatives) to the Cauchy problems with coefficients and data of the respective classes. To this end, let us take an even probability density $\omega \in C_0^\infty(\mathbb{R}^d)$ with support in the unit ball such that $|\nabla\omega|^2/\omega \in L^1(\mathbb{R}^d)$, set $\omega_\varepsilon(x) = \varepsilon^{-d}\omega(x/\varepsilon)$, $\varepsilon \in (0,1)$, and for any locally integrable function g on $\mathbb{R}^d$ set $(g)_\varepsilon := g * \omega_\varepsilon$. It is readily verified that the measures with densities $(\varrho_\mu)_\varepsilon$ and $(\varrho_\sigma)_\varepsilon$, where the convolution is taken with respect to x, are solutions to the Cauchy problems corresponding to the operators with smooth coefficients

$$A_\mu^\varepsilon = \frac{(A_\mu \varrho_\mu)_\varepsilon}{(\varrho_\mu)_\varepsilon}, \quad b_\mu^\varepsilon = \frac{(b_\mu \varrho_\mu)_\varepsilon}{(\varrho_\mu)_\varepsilon}, \quad A_\sigma^\varepsilon = \frac{(A_\sigma \varrho_\sigma)_\varepsilon}{(\varrho_\sigma)_\varepsilon}, \quad b_\sigma^\varepsilon = \frac{(b_\sigma \varrho_\sigma)_\varepsilon}{(\varrho_\sigma)_\varepsilon},$$

respectively, and initial condition $\nu * \omega_\varepsilon$. Indeed, let $\varphi \in C_0^\infty(\mathbb{R}^d)$. For every fixed y, by (1.3) we have

$$\int_{\mathbb{R}^d} \varphi(y+x)\, \varrho_\mu(x,t)\, dx = \int_{\mathbb{R}^d} \varphi(y+x)\, \nu(dx)$$
$$+ \int_0^t \int_{\mathbb{R}^d} (a_\mu^{ij}(x,s)\partial_{x_i}\partial_{x_j}\varphi(y+x) + b_\mu^i(x,s)\partial_{x_i}\varphi(y+x))\, \varrho_\mu(x,s)\, dx\, ds.$$

Changing variables $z = x + y$ and integrating in y with respect to $\omega_\varepsilon(y)dy$, we arrive at (1.3) for $(\varrho_\mu)_\varepsilon$ and the new operator with A_μ^ε and b_μ^ε, because $A_\mu^\varepsilon(\varrho_\mu)_\varepsilon = (A_\mu \varrho_\mu)_\varepsilon$ and $b_\mu^\varepsilon(\varrho_\mu)_\varepsilon = (b_\mu \varrho_\mu)_\varepsilon$. The same is true for σ. Moreover, since the original coefficients are locally bounded, these new coefficients are bounded uniformly in ε on every set $U \times [0,T]$, where U is a ball in $\mathbb{R}^d$.

Since identity (2.7) holds for these approximations, it remains valid for the original solutions, because, as $\varepsilon \to 0$, we have $(\varrho_\mu)_\varepsilon \to \varrho_\mu$ uniformly on the set $U \times [\tau, t]$ and these functions are separated from zero on this set, in addition, $\partial_{x_i}(\varrho_\mu)_\varepsilon \to \partial_{x_i}\varrho_\mu$ in $L^2(U \times [\tau, t])$ and the same is true for σ. Thus, (2.7) is proved completely.

Let us show (2.8). Recall that $\varrho_\sigma = v\varrho_\mu$. Applying in (2.7) the inequality

$$|\langle v\Phi, \nabla v\rangle| \le \frac{1}{2}|A_\mu^{-1/2}\Phi|^2 v^2 + \frac{1}{2}|A_\mu^{1/2}\nabla v|^2$$

and taking into account that $f'' \ge 0$ and $\psi \ge 0$, we obtain an inequality similar to (2.8), but with the integrals over $[\tau, t]$ in place of $[0, t]$ and with the integral of $f(v(x,\tau))\psi(x)\varrho_\mu(x,\tau)$ in place of the integral of $f(1)\psi$ against ν. Obviously, the integrals over $[\tau, t]$ converge to the respective integrals over $[0, t]$ as $\tau \to 0$. So we have to show that

$$\lim_{\tau \to 0} \int_{\mathbb{R}^d} f(v(x,\tau))\psi(x)\varrho_\mu(x,\tau)\, dx = f(1) \int_{\mathbb{R}^d} \psi(x)\, \nu(dx). \tag{2.10}$$

It follows from (1.3) that

$$\lim_{\tau \to 0} \int_{\mathbb{R}^d} \psi(x)\varrho_\mu(x,\tau)\, dx = \int_{\mathbb{R}^d} \psi(x)\, \nu(dx),$$

and for proving (2.10) it is enough to show that

$$\int_{\mathbb{R}^d} |f(v(x,\tau)) - f(1)|\psi\varrho_\mu(x,\tau)\, dx \to 0.$$

We have $|f(v) - f(1)| \le C|v-1|$, since f' is bounded. Hence

$$\int_{\mathbb{R}^d} |f(v(x,\tau)) - f(1)|\psi\varrho_\mu(x,\tau)\,dx \le C\int_{\mathbb{R}^d} |\varrho_\sigma(x,\tau) - \varrho_\mu(x,\tau)|\psi\,dx$$

and by Lemma 2.1 the right-hand side tends to zero. □

Proof of Theorem 1.1. We would like to apply the second assertion of the last lemma to the functions

$$f(v) = v\log v - v$$

and ψ_N such that $\psi_N \to 1$ and $L_{A_\mu,b_\mu}\psi_N \to 0$. Then, letting $N\to\infty$ and using (2.9) and the equality $f''(v) = 1/v$, we arrive at the desired estimate. However, this function f does not satisfy the conditions of assertion (ii) in Lemma 2.4 and some additional justification is needed.

Let $m, k > 1$ and

$$f_{m,k}(t) = \begin{cases} -t\log k & \text{if } t \le k^{-1}, \\ t\log t - t + k^{-1} & \text{if } k^{-1} < t < m, \\ t\log m - m + k^{-1} & \text{if } t \ge m. \end{cases}$$

We observe that $f'_{m,k}(t) = \log\bigl((k^{-1}\vee t)\wedge m\bigr)$ and $f''_{m,k}(t) = t^{-1}I_{\{k^{-1}<t<m\}}$, where $I_{\{k^{-1}<t<m\}}$ is the indicator function of the set $\{k^{-1} < t < m\}$. In addition,

$$f_{m,k}(1) = -1 + k^{-1} \quad \text{and} \quad |f_{m,k}(t)| \le C(m,k)t$$

for fixed m, k.

Let us now consider separately cases (a) and (b). Suppose first that (a) is fulfilled. Set $\psi_N(x) = \psi(x/N)$, where $\psi \in C_0^\infty(\mathbb{R}^d)$, $\psi \ge 0$, $\psi(x) = 1$ whenever $|x| < 1$ and $\psi(x) = 0$ whenever $|x| \ge 2$. Let us substitute in (2.9) the functions $f_{m,k}$ and ψ_N for f and ψ, respectively, and let first $N\to\infty$ and then $m,k\to\infty$. We observe that $|f_{m,k}(v)| \le C(m,k)v$ and

$$\Bigl|\int_0^t\int_{\mathbb{R}^d} \varrho_\mu f_{m,k}(v) L_{A_\mu,b_\mu}\psi_N\,dx\,ds\Bigr| \le C(m,k)\int_0^t\int_{\mathbb{R}^d} |L_{A_\mu,b_\mu}\psi_N|\,d\sigma_s\,ds.$$

Since

$$|L_{A_\mu,b_\mu}\psi_N| \le N^{-2}|a_\mu^{ij}|\,|\partial^2_{x_ix_j}\psi| + N^{-1}|b_\mu|\,|\nabla\psi| \le C(\psi)\sum_{i,j\le d}\frac{|a_\mu^{ij}|}{(1+|x|)^2} + C(\psi)\frac{|b_\mu|}{1+|x|},$$

we have

$$\lim_{N\to\infty}\int_0^t\int_{\mathbb{R}^d}\varrho_\mu f_{m,k}(v)L_{A_\mu,b_\mu}\psi_N\,dx\,ds=0.$$

In addition, by the estimate

$$|\nabla\psi_N(x)|=N^{-1}|\nabla\psi(x/N)|\le 3(1+|x|)^{-1}\max|\nabla\psi|,$$

which holds because $\psi(x/N)=0$ whenever $|x|\ge 2N$, we have

$$\Bigl|\langle\Phi,\nabla\psi_N\rangle\log\bigl((k^{-1}\vee v)\wedge m\bigr)\Bigr|\le 3(\log k+\log m)\max|\nabla\psi|(1+|x|)^{-1}|\Phi|,$$

hence by the integrability of $(1+|x|)^{-1}|\Phi|$ with respect to σ and the Lebesgue dominated convergence theorem we have

$$\lim_{N\to\infty}\int_0^t\int_{\mathbb{R}^d}\langle\Phi,\nabla\psi_N\rangle\log\bigl((k^{-1}\vee v)\wedge m\bigr)\,d\sigma_s\,ds=0.$$

Finally, note that

$$\int_{\mathbb{R}^d}f_{m,k}(v)\varrho_\mu\,dx\ge -\log k\int_{v<k^{-1}}v\varrho_\mu\,dx+\int_{k^{-1}<v<m}v\log v\varrho_\mu\,dx-\int_{k^{-1}<v<m}v\varrho_\mu\,dx.$$

It is clear that the first term on the right tends to zero as $k\to\infty$, since it is dominated by $k^{-1}\log k$ in absolute value. The second term converges to the integral of $I_{\{v<m\}}v\log v\varrho$, because the function $v\log v$ is bounded on the set $\{v<m\}$. The last term tends to -1 if $k,m\to\infty$. Recall that $f_{m,k}(1)=-1+k^{-1}$. Thus, passing to the limit in (2.9) in N (with fixed m and k) and then in m and k, we obtain the estimate

$$\int_{\mathbb{R}^d}v(x,t)\log v(x,t)\varrho_\mu(x,t)\,dx\le\frac{1}{2}\int_0^t\int_{\mathbb{R}^d}|A_\mu^{-1/2}\Phi|^2\,d\sigma_s\,ds.$$

Let us now consider case (b). Set $\psi_N(x)=\zeta(V(x)/N)$, where $\zeta\in C^\infty(\mathbb{R}^1)$, $\zeta'\le 0$, $\zeta''\ge 0$, $\zeta(0)=1$ and $\zeta(t)=0$ if $t>1$. We observe that

$$L_{A_\mu,b_\mu}\psi_N=N^{-1}\zeta'(V/N)L_{A_\mu,b_\mu}V+N^{-2}\zeta''(V/N)|A_\mu^{-1/2}\nabla V|^2.$$

Let us substitute in (2.9) the functions $f_{m,k}$ and ψ_N for f and ψ, respectively, and let first $N\to\infty$ and then $m,k\to\infty$.

Note that since ϱ_μ and $v\varrho_\mu=\varrho_\sigma$ are solutions to the respective Cauchy problems, for any numbers α and β one has

$$\int_0^t\int_{\mathbb{R}^d}\varrho_\mu(f_{m,k}(v)-\alpha v-\beta)L_{A_\mu,b_\mu}\psi_N\,dx\,ds$$
$$=\int_0^t\int_{\mathbb{R}^d}\varrho_\mu f_{m,k}(v)L_{A_\mu,b_\mu}\psi_N\,dx\,ds-\alpha\int_0^t\int_{\mathbb{R}^d}\langle\Phi,\nabla\psi_N\rangle\varrho_\sigma\,dx\,ds$$
$$-\beta\int_{\mathbb{R}^d}\psi_N(x)\varrho_\mu(x,t)\,dx+\beta\int_{\mathbb{R}^d}\psi_N\,d\nu$$
$$-\alpha\int_{\mathbb{R}^d}\psi_N(x)v(x,t)\varrho_\mu(x,t)\,dx+\alpha\int_{\mathbb{R}^d}\psi_N\,d\nu. \qquad (2.11)$$

Applying (2.11) with $\alpha=\log m$ and $\beta=k^{-1}$ we have

$$\overline{\lim_{N\to\infty}}\int_0^t\int_{\mathbb{R}^d}\varrho_\mu f_{m,k}(v)L_{A_\mu,b_\mu}\psi_N\,dx\,ds$$
$$\le\overline{\lim_{N\to\infty}}\int_0^t\int_{\mathbb{R}^d}\varrho_\mu\bigl(f_{m,k}(v)-v\log m-k^{-1}\bigr)L_{A_\mu,b_\mu}\psi_N\,dx\,ds$$
$$+\log m\,\overline{\lim_{N\to\infty}}\int_0^t\int_{\mathbb{R}^d}\langle\Phi,\nabla\psi_N\rangle\,d\sigma_s\,ds.$$

It is readily seen that

$$f_{m,k}(v)-v\log m-k^{-1}\le 0,\quad |f_{m,k}(v)-v\log m-k^{-1}|\le C(m,k)(1+v)$$

and

$$\varrho_\mu(f_{m,k}(v)-v\log m-k^{-1})L_{A_\mu,b_\mu}\psi_N\le C(\zeta,m,k)MN^{-1}(\varrho_\mu+v\varrho_\mu)V.$$

For every $\delta\in(0,1)$ we have

$$N^{-1}\int_0^t\int_{V<N}V\,d(\mu+\sigma)\le\delta\int_0^t\int_{V<\delta N}1\,d(\mu_s+\sigma_s)\,ds+\int_0^t\int_{\delta N<V<N}1\,d(\mu_s+\sigma_s)\,ds.$$

Therefore,

$$\overline{\lim_{N\to\infty}}N^{-1}\int_0^t\int_{V<N}V\,d(\mu_s+\sigma_s)\,ds\le 2t\delta.$$

Since δ was arbitrary, we obtain that

$$\lim_{N\to\infty} N^{-1}\int_0^t \int_{V<N} V\, d(\mu_s+\sigma_s)\, ds = 0$$

and

$$\overline{\lim_{N\to\infty}} \int_0^t \int_{\mathbb{R}^d} \varrho_\mu (f_{m,k}(v) - v\log m - k^{-1}) L_{A_\mu,b_\mu}\psi_N\, dx\, dt \le 0.$$

Finally, note that $\langle \Phi, \nabla\psi_N\rangle = N^{-1}\zeta'(V/N)\langle \Phi, \nabla V\rangle$ and for every $\delta \in (0,1)$

$$N^{-1}\int_0^t \int_{V<N} \big|\langle \Phi, \nabla V\rangle\big|\, d\sigma_s\, ds \le (N^{-1}+\delta)\int_0^t \int_{V<\delta N} \big|\langle \Phi, \nabla V\rangle\big|(1+V)^{-1}\, d\sigma_s\, ds$$
$$+ (N^{-1}+1)\int_0^t \int_{\delta N<V<N} \big|\langle \Phi, \nabla V\rangle\big|(1+V)^{-1}\, d\sigma_s\, ds.$$

As above, we conclude that

$$\overline{\lim_{N\to\infty}} \int_0^t \int_{\mathbb{R}^d} \langle \Phi, \nabla\psi_N\rangle\, d\sigma_s\, ds \le 0.$$

Thus, passing in (2.9) to the limit in N for fixed m and k and then in m and k, we obtain the desired bound. □

It is worth noting that it has been actually proved above that the entropy at time t is estimated by the entropy at time $\tau < t$ with the added integral with Φ.

3. Corollaries and examples

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

Corollary 3.1. *Let $A_\mu = A_\sigma = A$ be uniformly bounded (and satisfy* (H)*). Suppose that for some numbers $\gamma_1 > 0$ and $\gamma_2 > 0$ we have*

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

Then

$$\|\mu_t - \sigma_t\|_{TV}^2 \le \int_0^t \int_{\mathbb{R}^d} |A^{-1/2}(b_\mu - b_\sigma)|^2 \, d\sigma_s \, ds.$$

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

$$\|(1 + |x|^p)(\mu_t - \sigma_t)\|_{TV}^2 \le 2K^{-1}\Bigl(1 + \log\Bigl(\int_{\mathbb{R}^d} e^{K(1+|x|^p)^2} \mu_t(dx)\Bigr)\Bigr) \int_0^t \int_{\mathbb{R}^d} |A^{-1/2}(b_\mu - b_\sigma)|^2 \, d\sigma_s \, ds.$$

Proof. Condition (b) in Theorem 1.1 is fulfilled with $V(x) = |x|^2$, so we apply Corollary 1.2 with $\varphi = 1$ to obtain the first estimate. The second one is similar, we take $V(x) = |x|^2$ and $\varphi(x) = \sqrt{K}(1 + |x|^p)$. □

Example 3.2. In case $A_\mu = A_\sigma$ is uniformly bounded, condition (a) is fulfilled if $|b_\mu(x)| \le C + C|x|$ or if $|b_\mu(x)| \le C + C|x|^m$ and $|x|^{m-1}$ is integrable with respect to σ. The latter can be verified by using Lyapunov functions (see, e.g., [8,12,18]). Also the assumption that $|b_\mu - b_\sigma|^2$ is σ-integrable can be verified in these terms. Certainly, the case of bounded b_μ and b_σ is covered by both conditions.

Example 3.3. Let L_μ be the Ornstein–Uhlenbeck operator $\Delta u(x) - \langle x, \nabla u(x)\rangle$ and let L_σ be its perturbation by a first order term generated by a bounded Borel vector field b_0 on $\mathbb{R}^d$. Then

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

In case $\varphi = 1$ we obtain that

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

This estimate extends to the infinite-dimensional case, which will be considered in a separate paper along with some generalizations to locally unbounded drifts.

Remark 3.4. If A_μ is uniformly bounded and for some $p \ge 1$, $K > 0$, $\gamma_1 > 0$ and $\gamma_2 > 2pK$ we have

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

then for some $C > 0$ and all $t \in [0, T]$ one has by Gronwall's inequality (see, e.g., [47])

$$\int_{\mathbb{R}^d} e^{K|x|^{2p}}\,\mu_t(dx) \le e^{Ct} + e^{Ct}\int_{\mathbb{R}^d} e^{K|x|^{2p}}\,\nu(dx).$$

Corollary 3.5. *Let A_μ and A_σ satisfy* (H). *Suppose that there are numbers $\lambda_1, \lambda_2 > 0$ such that*

$$\lambda_1 \cdot \mathrm{I} \le A_\mu(x,t) \le \lambda_2 \cdot \mathrm{I}, \quad \lambda_1 \cdot \mathrm{I} \le A_\sigma(x,t) \le \lambda_2 \cdot \mathrm{I} \quad \text{for all } (x,t).$$

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

$$\langle b_\mu(x,t), x\rangle \le \gamma_1 + \gamma_2|x|^2, \quad |b_\sigma(x,t)| \le \gamma_3 + \gamma_4|x|^m$$

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

$$\|\mu_t - \sigma_t\|_{TV}^2 \le C(T)\sup_{x,t}\|A_\mu - A_\sigma\|^2 + C(T)\int_0^t\int_{\mathbb{R}^d} |A_\mu^{-1/2}(h_\mu - h_\sigma)|^2\,d\sigma_s\,ds,$$

where $h_\mu^i - h_\sigma^i = b_\mu^i - b_\sigma^i - \partial_{x_j}(a_\mu^{ij} - a_\sigma^{ij})$ and the number $C(T)$ on the right depends on T, m, λ_i, γ_i, $\int |x|^{2m}\,d\nu$, and $\|\varrho_0 \ln \varrho_0\|_{L^1(\mathbb{R}^d)}$.

Proof. This follows easily from Corollary 1.2 combined with estimate (1.7) and Gronwall's inequality. □

Remark 3.6. We observe that passing from (2.8) to (2.9) we have merely discarded the nonnegative term with $|A_\mu^{1/2}\nabla v|^2$. Keeping this term, we obtain the integrability of the function $|A_\mu^{1/2}\nabla v|^2/v$ with respect to μ (actually, already known from [14]), which means membership of v in the corresponding weighted Sobolev class.

4. Applications

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

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

Then we can consider the Cauchy problem for the nonlinear Fokker–Planck–Kolmogorov equation

$$\partial_t \mu = \Delta\mu - \operatorname{div}(b(\mu,x,t)\mu), \quad \mu|_{t=0} = \nu. \tag{4.1}$$

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 (1.3) is fulfilled. The linear case considered above corresponds to a drift independent of measures. As warned above, the notation b_μ in the previous sections indicated only that μ was a solution for a given drift, but now the drift may depend on the unknown solution, which makes our equation nonlinear. For example, if in the one-dimensional case $b(\mu, x, t) = \mu$, then we obtain the equation $\partial_t \mu = \mu'' - (\mu^2)'$ with a quadratic nonlinearity. One can also think that a solution is a measure satisfying a linear equation, but its drift has been preassigned to this measure. This opens a way of solving the nonlinear equation by means of a fixed point principle: for each drift b_σ, we solve the linear equation with this drift and obtain its solution $\mu(\sigma)$, which, of course, in general differs from σ. But in case of coincidence we obtain a solution to the nonlinear equation. Certainly, nonlinear equations with nonconstant second order coefficients $a^{ij}(\mu, x, t)$ can be considered similarly. We deal with $A = \mathrm{I}$ just for simplicity.

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 $V \geq 1$. For $\alpha \in C^+[0,T]$ and $\tau \in (0,T]$ we set

$$\mathcal{M}_{\tau,\alpha}(V) = \Big\{\mu(dxdt) = \mu_t(dx)\,dt\colon\ \mu_t \geq 0, \mu_t(\mathbb{R}^d) = 1, \\ \int_{\mathbb{R}^d} V(x)\,\mu_t(dx) \leq \alpha(t), t \in [0,\tau]\Big\}.$$

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

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

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

on the linear space of signed measures for which it is finite (to see that it is a norm, one can use the triangle inequality for the total variation norm and the Cauchy inequality or apply the fact that here we deal with the L^2-norm of a mapping with values in a normed space). Note that $\mathcal{M}^{K,p}_{\tau,\alpha}$ is a complete metric space with respect to the metric generated by this norm, which follows by the completeness of the space $L^2([0,\tau], X)$ of L^2-mappings with values in a Banach space X.

For shortening notation, we shall write occasionally $b(\mu)$ in place of $b(\mu, x, t)$.

Corollary 4.1. *Let $p \geq 1$, $K > 0$ 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 $\mu \in \mathcal{M}^{K,p}_{\tau,\alpha}$ one has*

$$\langle b(\mu, x, t), x\rangle \leq \gamma_1(\alpha) - \gamma_2(\alpha)|x|^{2p} \quad \forall\ (x,t) \in \mathbb{R}^d \times [0,\tau].$$

Suppose also that

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

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 (4.1) in the class of measures* $\mathcal{M}^{K,p}_{\tau,\alpha}$ *exists and is unique.*

Proof. Let us define a mapping $F\colon \mathcal{M}^{K,p}_{\tau,\alpha} \to \mathcal{M}^{K,p}_{\tau,\alpha}$ by

$$\mu = F(\sigma) \Longleftrightarrow \partial_t\mu = \Delta\mu - \operatorname{div}(b(\sigma)\mu),\ \ \mu|_{t=0} = \nu.$$

We have to find τ and α such that F will take values in $\mathcal{M}^{K,p}_{\tau,\alpha}$. Let

$$\alpha(t) = e\Big(1 + \int_{\mathbb{R}^d} e^{K|x|^{2p}}\, \nu(dx)\Big).$$

According to Remark 3.4 we have

$$\int_{\mathbb{R}^d} e^{K|x|^{2p}}\, \mu_t(dx) \le e^{qt} + e^{qt}\int_{\mathbb{R}^d} e^{K|x|^{2p}}\, \nu(dx),$$

where $q > 0$ depends only on $\gamma_1(\alpha)$, $\gamma_2(\alpha)$, p and K. Let $\tau < 1/q$. Then

$$\int_{\mathbb{R}^d} e^{K|x|^{2p}}\, \mu_t(dx) \le \alpha(t) \quad \forall t \in [0,\tau]$$

and F takes values in $\mathcal{M}^{K,p}_{\tau,\alpha}$.

Applying Corollary 1.2 and Remark 3.4 we obtain that

$$\|(1+|x|^p)(\mu^1_t - \mu^2_t)\|^2_{TV} \le \widetilde{C}\int_0^t\int_{\mathbb{R}^d} |b(\sigma^1) - b(\sigma^2)|^2\, d\sigma \le \widehat{C}\|\sigma^1 - \sigma^2\|^2_{p,\tau},$$

where $\widehat{C}$ does not depend on τ, but only on T. Integrating in t over $[0,\tau]$, we find that

$$\|F(\sigma^1) - F(\sigma^2)\|^2_{p,\tau} \le \tau\widehat{C}\|\sigma^1 - \sigma^2\|^2_{p,\tau}.$$

For $\tau < 1/\widehat{C}$ the mapping F is contracting, therefore, in $\mathcal{M}^{K,p}_{\tau,\alpha}$ there exists a unique solution. □

Note that under different assumptions a solution to a nonlinear Vlasov equation was constructed in [27] by employing the contraction mapping theorem for the Kantorovich norm. The existence of not necessarily unique solutions has been proved in [16] and [40] by using the Schauder fixed point theorem.

Example 4.2. 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

$$|K(x, y)| \le C(1 + |x|^q)(1 + |y|^p), \quad \langle \beta(x, t), x\rangle \le \gamma_1 - \gamma_2 |x|^{2p}.$$

Then all conditions of the above corollary are fulfilled.

Corollary 4.3. *Suppose that there exists a Lyapunov function* $V \in C^2(\mathbb{R}^d)$ *such that* $\lim_{|x|\to\infty} V(x) = +\infty$ *and for every* $\tau \in (0, T]$ *and* $\alpha \in C^+[0, T]$ *the following conditions are fulfilled*:

- (i) *for every* $\mu \in \mathcal{M}_{\tau,\alpha}(V)$ *there exists a number* $M(\alpha)$ *such that* $L_\mu V \le M(\alpha)V$;
- (ii) *for every* μ *and* σ *in* $\mathcal{M}_{\tau,\alpha}(V)$ *one has* $b(\mu) \in L^2(\sigma)$ *and, in addition, for every ball* $U \subset \mathbb{R}^d$ *there exists a number* $C(U, \alpha, \tau) > 0$ *such that*

$$\sup_{\mu\in\mathcal{M}_{\tau,\alpha}(V)} \|b(\mu)\|_{L^\infty(U\times[0,\tau])} \le C(U, \tau, \alpha);$$

- (iii) *if a sequence of measures* $\{\mu^n\}$ *and a measure* μ *in* $\mathcal{M}_{\tau,\alpha}(V)$ *are such that*

$$\lim_{n\to\infty} \int_0^\tau \|\mu_t^n - \mu_t\|_{TV}^2\, dt = 0,$$

then the mappings $b(\mu_n)$ *converge to* $b(\mu)$ *in* $L^2(\mu, \mathbb{R}^d)$.

Then there exists a number $\tau \in (0, T]$ *such that on the interval* $[0, \tau]$ *the Cauchy problem* (4.1) *has a solution.*

Proof. Let us consider the mapping F from the proof of the previous corollary and choose as above τ and α such that $F\colon \mathcal{M}_{\tau,\alpha}(V) \to \mathcal{M}_{\tau,\alpha}(V)$. By assumption, the drift $b(\sigma)$ is bounded on each set of the form $U\times[0, T]$ (where U is a ball) uniformly in $\sigma \in \mathcal{M}_{\tau,\alpha}(V)$. Therefore (see [11]), for every ball U and every compact interval $[\tau_1, \tau_2] \subset (0, \tau)$, there exists a number $\widetilde{C}(U, \alpha, \tau, \tau_1, \tau_2)$ bounding the Hölder norm (in both variables) of the density of the solution $\mu = F(\sigma)$ on $U \times [\tau_1, \tau_2]$.

Let us consider the subset $\mathcal{L}_{\tau,\alpha}(V)$ of the set $\mathcal{M}_{\tau,\alpha}(V)$ consisting of measures given by Hölder continuous densities ϱ such that on every set of the form $U\times[\tau_1, \tau_2]$ the Hölder norm is bounded by the number $\widetilde{C}(U, \alpha, \tau, \tau_1, \tau_2)$. We observe that the set $\mathcal{L}_{\tau,\alpha}(V)$ is convex and compact in the normed space of finite measures $\mu(dxdt) = \mu_t(dx)\,dt$ on $\mathbb{R}^d \times [0,T]$ with

$$\|\mu\| = \sqrt{\int_0^\tau \|\mu_t\|_{TV}^2\,dt} < \infty.$$

Indeed, each sequence in this set contains a subsequence uniformly convergent on the sets of the form $U \times [\tau_1, \tau_2]$, where U is a ball in $\mathbb{R}^d$ and $[\tau_1, \tau_2] \subset (0,T)$, because this sequence has uniformly bounded Hölder norms on such sets. Hence this subsequence converges in the indicated norm. It remains to note that F maps $\mathcal{L}_{\tau,\alpha}(V)$ into itself and by Corollary 1.2 is continuous. Therefore, by the Schauder fixed point theorem F has a fixed point, i.e., a solution to the Cauchy problem. □

Let $\mathcal{M}_T(V)$ denote the set of nonnegative measures $\mu(dxdt) = \mu_t(dx)\,dt$ such that

$$\sup_{t\in[0,T]} \int_{\mathbb{R}^d} V\,d\mu_t < \infty.$$

Corollary 4.4. *Let $V > 1$ and $W = \sqrt{\log V}$. Suppose that for every measure μ in $\mathcal{M}_T(V)$ there exist a positive function $\Psi_\mu \in C^2(\mathbb{R}^d)$ and a number $\beta(\mu)$ such that $\lim_{|x|\to\infty} \Psi_\mu(x) = +\infty$, $|\nabla\Psi_\mu|\Psi_\mu^{-1}V^{-1}$ is bounded and $L_\mu\Psi_\mu \le \beta(\mu)\Psi_\mu$. Suppose also that there exists an increasing continuous function G on $[0,+\infty)$ such that $G(0) = 0$ and*

$$|b(\mu,x,t) - b(\sigma,x,t)| \le \sqrt{V(x)}G(\|W(\mu_t - \sigma_t)\|_{TV}) \quad \forall\,\mu,\sigma,x,t.$$

If we have

$$\int_{0+} \frac{du}{G^2(\sqrt{u})} = +\infty,$$

then the Cauchy problem (4.1) has at most one solution in the class $\mathcal{M}_T(V)$.

Proof. According to Theorem 1.1, for any two solutions μ and σ one has

$$\|W(\mu_t - \sigma_t)\|_{TV}^2 \le C\int_0^t G^2(\|W(\mu_s - \sigma_s)\|_{TV})\,ds.$$

Hence by Gronwall's inequality $\|W(\mu_t - \sigma_t)\|_{TV} = 0$. □

Under different assumptions, sufficient conditions for uniqueness of solutions to nonlinear Fokker–Planck–Kolmogorov equations have been obtained in [39]. The estimate from Theorem 1.1 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 to this, see [44,48] and the recent papers [19,20].

Corollary 4.5. *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*

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

Let μ^α *be a probability solution to the Cauchy problem* (1.2) *with* $b(\alpha, x, t)$ *and* $A = \mathrm{I}$. *Suppose that for every* $\alpha_0 \in [0,1]$ *we have*

$$\limsup_{r\to 0} \int_0^T \int_{\mathbb{R}^d} \Bigl|\frac{b(\alpha_0 + r, x, t) - b(\alpha_0, x, t)}{r}\Bigr|^2 \, d\mu_t^{\alpha_0}\, dt < \infty.$$

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

Proof. Let us fix α_0. For notational simplicity we write $b(\alpha)$ in place of $b(\alpha, x, t)$. Set

$$\delta_r\varrho(x,t) = \frac{\varrho(\alpha_0 + r, x, t) - \varrho(\alpha_0, x, t)}{r}.$$

Similarly we define $\delta_r b$. By Theorem 1.1 we have

$$\|\delta_r\varrho(\,\cdot\,, t)\|^2_{L^1(\mathbb{R}^d)} \le \int_0^t \int_{\mathbb{R}^d} |\delta_r b|^2 \, d\mu_s^{\alpha_0}\, ds.$$

Therefore, the quantity $\sup_{t\in[0,T]} \|\delta_r\varrho(\,\cdot\,, t)\|_{L^1(\mathbb{R}^d)}$ is uniformly bounded as $r \to 0$. We observe that $\delta_r\varrho$ satisfies the equation

$$\partial_t\delta_r\varrho = \Delta\delta_r\varrho - \operatorname{div}\bigl(b(\alpha_0 + r)\delta_r\varrho\bigr) - \operatorname{div}\bigl(\delta_r b\varrho_{\alpha_0}\bigr).$$

Since the coefficients $b(\alpha_0 + r)$ and $\delta_r b\varrho_{\alpha_0}$ are bounded on every set $U \times [0,T]$ uniformly in $r > 0$, according to the estimates from [11], the functions $\delta_r\varrho$ have uniformly bounded

Hölder norms. Hence there exists a sequence $r_k \to 0$ for which $\delta_{r_k}\varrho$ converges to some function w. This function satisfies the Cauchy problem

$$\partial_t w = \Delta w - \operatorname{div}(bw) - \operatorname{div}(\varrho\partial_{\alpha_0} b), \quad w|_{t=0} = 0.$$

Moreover, $w \in L^\infty([0,T], L^1(\mathbb{R}^d))$. It remains to show that a solution with such properties is unique. The difference of two solutions satisfies the homogeneous equation

$$\partial_t w = \Delta w - \operatorname{div}(bw).$$

According to [17], a solution to this equation in the considered class is unique. □

Remark 4.6. Suppose that under the conditions of Corollary 4.5 the function $V(x) = \ln(4 + |x|)$ is in $L^1(\nu)$. Then $\Delta V + \langle b, \nabla V\rangle \le C + CV$, hence by [13, Theorem 7.1.1]

$$\sup_t \int_{\mathbb{R}^d} \ln(4 + |x|)\, \mu_t^\alpha(dx) < \infty.$$

Let $\varphi(x) = \sqrt{\ln\ln(4 + |x|)}$. Then by Corollary 1.2 there is $C > 0$ such that

$$\|\varphi(\,\cdot\,)\delta_r\varrho(\,\cdot\,, t)\|^2_{L^1(\mathbb{R}^d)} \le C \int_0^t \int_{\mathbb{R}^d} |\delta_r b|^2\, d\mu_s^{\alpha_0}\, ds \quad \forall\, t \in (0,T).$$

Since $\lim\limits_{|x|\to\infty} \varphi(x) = +\infty$, we obtain

$$\int_{\mathbb{R}^d} |\delta_r\varrho(x,t) - \partial_\alpha\varrho(x,t,\alpha_0)|\, dx \to 0 \quad \text{as } r \to 0.$$

Our next application is concerned with optimal control (see [37]). Following [2,3,41], for a given bounded probability density σ on $\mathbb{R}^d$ and $\tau \in (0,1)$, we consider the problem of minimization of the function

$$J(u) = \frac{1}{2}\int_{\mathbb{R}^d} |\varrho(x,\tau,u) - \sigma(x)|^2\, dx + \frac{u^2}{2}, \quad u \in \mathbb{R}, \tag{4.2}$$

in the class of probability densities $x \mapsto \varrho(x,t,u)$ on $\mathbb{R}^d$ such that $\varrho(x,t,u)$ solves the Cauchy problem on $[0,\tau]$ for the Fokker–Planck–Kolmogorov equation

$$\partial_t\varrho = \Delta\varrho - \operatorname{div}(b(\cdot,\cdot,u)\varrho), \quad \varrho|_{t=0} = \varrho_0.$$

We deal with $\mathbb{R}^d$, not with a bounded domain as in [2,3]. The unit diffusion matrix is considered for simplicity, a general case will be considered elsewhere. We assume that the initial condition ϱ_0 is a bounded probability density with $\varrho_0 \ln(4+|x|) \in L^1(\mathbb{R}^d)$ and the drift b depending on the parameter u satisfies the inequality

$$|b(x,t,u)| + |\partial_u b(x,t,u)| + |\partial_u^2 b(x,t,u)| \le M \quad \forall\,(x,t) \in \mathbb{R}^d \times [0,1],\ u \in \mathbb{R}.$$

In particular,

$$|b(x,t,u) - b(x,t,v)| + |\partial_u b(x,t,u) - \partial_u b(x,t,v)| \le 2M|u-v|.$$

For brevity we write $b(u)$ in place of $b(x,t,u)$ and $\varrho(u)$ in place of $\varrho(x,t,u)$. Under our assumptions, for each u there is a unique solution $\varrho(x,t,u)$ on $\mathbb{R}^d \times [0,1]$. Moreover, by [13, Corollary 7.3.8], there is a number $C_0 > 0$ such that $\varrho(x,t,u) \le C_0$ for all $(x,t) \in \mathbb{R}^d \times [0,1]$ and $u \in \mathbb{R}$. Hence $J(u)$ is defined for all u.

Corollary 4.7. *There is $\tau > 0$ such that J from (4.2) has a unique point of minimum.*

Proof. Let $\varrho(x,t,u) \le C_0$ and $\sigma(x) \le C_\sigma$. We observe that

$$\left| \int |\varrho(x,\tau,u) - \sigma(x)|^2\,dx - \int |\varrho(x,\tau,v) - \sigma(x)|^2\,dx \right| \le 2(C_0 + C_\sigma) \int |\varrho(x,\tau,u) - \varrho(x,\tau,v)|\,dx.$$

By Corollary 1.2 we have

$$\left(\int |\varrho(x,\tau,u) - \varrho(x,\tau,v)|\,dx \right)^2 \le \int_0^\tau \int_{\mathbb{R}^d} |b(x,t,u) - b(x,t,v)|^2 \varrho(x,t,u)\,dx \le \tau M^2 |u-v|^2,$$

therefore,

$$\left| \int |\varrho(x,\tau,u) - \sigma(x)|^2\,dx - \int |\varrho(x,\tau,v) - \sigma(x)|^2\,dx \right| \le 2\sqrt{\tau}(C_0 + C_\sigma) M |u-v|.$$

Thus, the function J is continuous and tends to $+\infty$ as $|u| \to +\infty$, which implies the existence of a point of minimum. We now show that for $\tau > 0$ sufficiently small such a point is unique. By Corollary 4.5 the function ϱ is differentiable in u and

$$\|\partial_u \varrho(\,\cdot\,,t,u)\|_{L^1(\mathbb{R}^d)} \le M\sqrt{t} \quad \forall u \in \mathbb{R}, t \in [0,\tau].$$

In addition, for each t, the function $\varrho(x,t,u)$ is differentiable in u in $L^1(\mathbb{R}^d)$. Hence the function J is differentiable, and at the point of minimum

$$J'(u) = \int_{\mathbb{R}^d} (\varrho(x,\tau,u) - \sigma(x))\partial_u \varrho(x,\tau,u)\, dx + u = 0.$$

Let us consider the mapping $G\colon \mathbb{R} \to \mathbb{R}$ given by

$$G(u) = -\int_{\mathbb{R}^d} (\varrho(x,\tau,u) - \sigma(x))\partial_u \varrho(x,\tau,u)\, dx.$$

The points of minimum of J are fixed points of G. We show that for $\tau > 0$ sufficiently small the mapping G is contracting, which yields the uniqueness of a point of minimum. We need some auxiliary estimates. We first show that there is a number N_1, independent of τ, such that

$$\int_{\mathbb{R}^2} |\varrho(x,\tau,u) - \varrho(x,\tau,v)|^2\, dx \le N_1^2 \tau |u - v|^2. \tag{4.3}$$

Set $\eta(x,t) = \varrho(x,t,u) - \varrho(x,t,v)$. Note that

$$\partial_t \eta = \Delta\eta - \operatorname{div}(b(u)\eta) - \operatorname{div}((b(u) - b(v))\varrho(v)), \quad \eta|_{t=0} = 0. \tag{4.4}$$

Let $\psi \in C_0^\infty(\mathbb{R}^d)$ and $\|\psi\|_{L^2(\mathbb{R}^d)} \le 1$. Let f be the solution to the Cauchy problem $\partial_s f + \Delta f = 0$, $f(x,t) = \psi(x)$. Multiplying this equation by f and integrating by parts, we obtain

$$\int_0^t \int_{\mathbb{R}^d} |\nabla f|^2\, dx\, ds \le \int_{\mathbb{R}^d} |\psi|^2\, dx \le 1.$$

Using (4.4), we arrive at the estimate

$$\int_{\mathbb{R}^d} \eta(x,t)\psi(x)\, dx = \int_0^t \int_{\mathbb{R}^d} \Big[\langle b(u), \nabla f\rangle \eta + \langle b(u) - b(v), \nabla f\rangle \varrho(v)\Big]\, dx\, ds.$$

By the Cauchy inequality we estimate the right-hand side by

$$\|\nabla f\|_{L^2(\mathbb{R}^d\times[0,t])} \big\| |\eta b(u)| + |b(u) - b(v)|\varrho(v) \big\|_{L^2(\mathbb{R}^d\times[0,t])}.$$

Since

$$\|\nabla f\|_{L^2(\mathbb{R}^d\times[0,t])} \le 1, \quad |b(u)| \le M, \quad |b(u) - b(v)| \le M|u - v|, \quad |\varrho(v)|^2 \le C_0 \varrho(v)$$

and $\varrho(x,t,v)$ is a probability density in x, there is a number C_1 such that

$$\int_{\mathbb{R}^d} \eta(x,t)\psi(x)\,dx \le C_1\|\eta\|_{L^2(\mathbb{R}^d\times[0,t])} + C_1\sqrt{t}|u-v|.$$

Since $\psi \in C_0^\infty(\mathbb{R}^d)$ with $\|\psi\|_{L^2(\mathbb{R}^d)} \le 1$ was arbitrary, we obtain

$$\|\eta(\,\cdot\,,t)\|_{L^2(\mathbb{R}^d)} \le C_1\|\eta\|_{L^2(\mathbb{R}^d\times[0,t])} + C_1\sqrt{t}|u-v|.$$

Taking the square of this inequality and applying Gronwall's inequality, we obtain (4.3). Note that (4.3) yields the bound

$$\int_{\mathbb{R}^d} |\partial_u\varrho(x,\tau,u)|^2\,dx \le N_1\tau.$$

We now show that there is a number N_2 independent of τ such that

$$\int_{\mathbb{R}^d} |\partial_u\varrho(x,\tau,u) - \partial_u\varrho(x,\tau,v)|\,dx \le N_2\tau^{1/2}|u-v|. \tag{4.5}$$

Set $w(x,t) = \partial_u\varrho(x,t,u) - \partial_u\varrho(x,t,v)$. Then

$$\begin{aligned}\partial_t w &= \Delta w - \operatorname{div}(b(u)w)\\ &\quad - \operatorname{div}((b(u)-b(v))\partial_u\varrho(v)) - \operatorname{div}(\partial_u b(u)\varrho(u) - \partial_u b(v)\varrho(v)), \quad w|_{t=0} = 0.\end{aligned} \tag{4.6}$$

Let $\psi \in C_0^\infty(\mathbb{R}^d)$ and $|\psi| \le 1$. Let h solve the Cauchy problem $\partial_s h + \Delta h = 0$, $h(x,t) = \psi(x)$. Then one has $|\nabla h(x,s)| \le C_2(t-s)^{-1/2}$. Using (4.6), we obtain

$$\begin{aligned}\int_{\mathbb{R}^d} w(x,t)\psi(x)\,dx = \int_0^t\int_{\mathbb{R}^d} \Big[&\langle b(u),\nabla h\rangle w + \langle b(u)-b(v),\nabla h\rangle\partial_u\varrho\\ &+ \langle\partial_u b(u) - \partial_u b(v),\nabla h\rangle\varrho(u) + (\varrho(u)-\varrho(v))\langle\partial_u b(u),\nabla h\rangle\Big]\,dx\,ds.\end{aligned}$$

Using the above estimate on $|\nabla h|$ and the inequalities

$$|b(u)| + |\partial_u b(u)| \le M, \quad |b(u)-b(v)| + |\partial_u b(u) - \partial_u b(v)| \le M|u-v|,$$
$$\|\varrho(\,\cdot\,,t,u) - \varrho(\,\cdot\,,t,v)\|_{L^1(\mathbb{R}^d)} \le \sqrt{t}M|u-v|, \quad \|\partial_u\varrho(\,\cdot\,,t,u)\|_{L^1(\mathbb{R}^d)} \le M\sqrt{t},$$

we estimate the right-hand side of the previous equality by

$$C_3t^{1/2}|u-v| + C_3\int_0^t (t-s)^{-1/2}\int_{\mathbb{R}^d} |w(x,s)|\,dx\,ds.$$

Since $\psi \in C_0^\infty(\mathbb{R}^d)$ with $|\psi| \le 1$ was arbitrary, we obtain

$$\int_{\mathbb{R}^d} |w(x,t)|\,dx \le C_3 t^{1/2}|u-v| + C_3 \int_0^t (t-s)^{-1/2} \int_{\mathbb{R}^d} |w(x,s)|\,dx\,ds.$$

Estimate (4.3) follows from Gronwall's inequality.

We now observe that $|G(u) - G(v)|$ can be estimated by the sum

$$\|\varrho(\,\cdot\,,\tau,u) - \varrho(\,\cdot\,,\tau,v)\|_{L^2(\mathbb{R}^d)} \|\partial_u \varrho(\,\cdot\,,\tau,u)\|_{L^2(\mathbb{R}^d)} \\ + (C_0 + C_\sigma)\|\partial_u \varrho(\,\cdot\,,\tau,u) - \partial_u \varrho(\,\cdot\,,\tau,v)\|_{L^1(\mathbb{R}^d)},$$

which on the account of the estimates obtained above is bounded by $N_3\tau^{1/2}|u-v|$. If $N_3\tau^{1/2} < 1$, then G is a contraction. □

We finally mention yet another possible application, which concerns the so-called mean field games (see [5,29,31]). A typical model for mean field games is the system

$$\begin{cases} \partial_t u + \Delta u - H(x,\nabla u) = F(x,\mu_t), \\ \partial_t \mu_t - \Delta \mu_t - \operatorname{div}(b(x,\nabla u)\mu_t) = 0, \quad (x,t) \in \mathbb{R}^d \times (0,T), \end{cases} \tag{4.7}$$

with initial-terminal conditions $u(x,T) = G(x,\mu_T)$ and $\mu_t|_{t=0} = \nu$, where ν is a Borel probability measure, H, F, G are given functions and b is a vector field, usually $b(x,p) = \partial H(x,p)/\partial p$, but we do not assume this relation. The second equation is (1.1) with $A = I$.

We emphasize that we consider this system on the whole space without boundary conditions.

Let $\mathcal{P}(\mathbb{R}^d)$ be the space of all Borel probability measures on $\mathbb{R}^d$.

Suppose that F and G are functions on $\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$, H is a function on $\mathbb{R}^d \times \mathbb{R}^d$ and $b\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is a vector field such that

(C1) F and G are continuous, $G(x,\mu)$ is continuously differentiable in x and there exist numbers $L_0 > 0$, $L_1 > 0$ such that

$$|F(x,\mu)| + |G(x,\mu)| + |\nabla_x G(x,\mu)| \le L_0,$$
$$|F(x,\mu) - F(x,\sigma)| + |G(x,\mu) - G(x,\sigma)| + |\nabla_x G(x,\mu) - \nabla_x G(x,\sigma)| \\ \le L_1 \|\mu - \sigma\|_{TV}$$

for all $x \in \mathbb{R}^d$, $\mu,\sigma \in \mathcal{P}(\mathbb{R}^d)$;

(C2) H and b are continuous and for every $R > 0$ there exist numbers $M_0(R) > 0$ and $M_1(R) > 0$ such that, whenever $x \in \mathbb{R}^d$ and $|p| \le R$,

$$|H(x,p)| \le M_0(R), \quad |H(x,p) - H(x,q)| + |b(x,p) - b(x,q)| \le M_1(R)|p-q|.$$

A solution to (4.7) is a pair consisting of a mapping $u \in C([0,T], C_b^1(\mathbb{R}^d))$, which we consider also as a function $u(x,t)$ on $\mathbb{R}^d \times [0,T]$ and write ∇u for $\nabla_x u$, and a measure $\mu_t(dx)\,dt$ on $\mathbb{R}^d \times [0,T]$ such that $\mu_t \in \mathcal{P}(\mathbb{R}^d)$ and $\mu_t\,dt$ is a solution to the Cauchy problem

$$\partial_t \mu_t - \Delta\mu_t - \operatorname{div}(b(x,\nabla u)\mu_t) = 0, \quad \mu_t|_{t=0} = \nu,$$

and u satisfies the identity

$$u(x,t) = \int_{\mathbb{R}^d} Z(x-y, T-t) G(y,\mu_T)\,dy + \int_t^T \int_{\mathbb{R}^d} Z(x-y,\tau-t)\big(H(y,\nabla u(y,\tau)) + F(y,\mu_\tau)\big)\,dy\,d\tau,$$

where

$$Z(x,t) = (4\pi t)^{-d/2}\exp(-|x|^2/4t).$$

This interpretation of solutions is consistent with the usual one for smooth solutions.

Under our assumptions, $t \mapsto \mu_t$ has a version with values in $\mathcal{P}(\mathbb{R}^d)$ that is a continuous mapping with respect to $\|\cdot\|_{TV}$ on $(0,T)$. Indeed, due to our assumptions, $b(x,\nabla u(x,t))$ is uniformly bounded and continuous for any $u \in C([0,T], C_b^1(\mathbb{R}^d))$. It follows that the solution $\mu_t(dx) = \varrho(x,t)\,dx$ has a continuous density ϱ on $\mathbb{R}^d \times (0,T)$ and for every ball U one has $\|\varrho(\,\cdot\,,t) - \varrho(\,\cdot\,,s)\|_{L^1(U)} \to 0$ as $t \to s$. Since $\varrho(x,t)$ is a probability density for every t, we have $\|\varrho(\,\cdot\,,t) - \varrho(\,\cdot\,,s)\|_{L^1(\mathbb{R}^d)} \to 0$.

Corollary 4.8. *There is $T > 0$ such that (4.7) has a unique solution on $[0,T]$.*

Proof. We apply the contracting mapping theorem to the mapping

$$\mathcal{F}\colon C([0,T], C_b^1(\mathbb{R}^d)) \to C([0,T], C_b^1(\mathbb{R}^d)),$$

where $C([0,T], C_b^1(\mathbb{R}^d))$ is equipped with its natural norm

$$\|v\| = \sup_{t\in[0,T]} \sup_x \big[|v(x,t)| + |\nabla_x v(x,t)|\big],$$

defined as follows: for each $v \in C([0,T], C_b^1(\mathbb{R}^d))$ we find a solution μ_t (which is unique under our assumptions) to the Cauchy problem

$$\partial_t \mu_t - \Delta\mu_t - \operatorname{div}(b(x,\nabla v)\mu_t) = 0, \quad \mu_t|_{t=0} = \nu,$$

and set

$$\mathcal{F}(v) = \int_{\mathbb{R}^d} Z(x-y, T-t) G(y, \mu_T)\, dy - \int_t^T \int_{\mathbb{R}^d} Z(x-y, \tau - t)\big(H(y, \nabla v(y,\tau)) + F(y, \mu_\tau)\big)\, dy\, d\tau. \quad (4.8)$$

Let $R = 2L_0 + 1$ and $v \in C([0,T], C_b^1(\mathbb{R}^d))$ with $\|v\| \le R$. Then (4.8) yields the estimate

$$\|\mathcal{F}(v)\| \le 2L_0 + (4C\sqrt{T} + T)(M_0(R) + L_0), \quad C = \int_{\mathbb{R}^d} |x| Z(x,1)\, dx.$$

We take $T > 0$ such that $\|\mathcal{F}(v)\| \le R$ and $\mathcal{F}$ maps the set $\{v\colon \|v\| \le R\}$ into itself. Let μ_t^1 and μ_t^2 be two solutions corresponding to v_1 and v_2. According to Corollary 1.2, we have

$$\|\mu_t^1 - \mu_t^2\|_{TV}^2 \le \int_0^t \int_{\mathbb{R}^d} |b(x, \nabla v_1(x,s)) - b(x, \nabla v_2(x,s))|^2\, \mu_s^2(dx)\, ds.$$

It follows that

$$\|\mu_t^1 - \mu_t^2\|_{TV} \le \sqrt{T} M_1(R) \|v_1 - v_2\|.$$

From (4.8) for $u = \mathcal{F}(v)$ we deduce the estimate

$$\|\mathcal{F}(v_1) - \mathcal{F}(v_2)\| \le (\sqrt{T} L_1 M_1(R) + (4C\sqrt{T} + T)(M_1(R) + \sqrt{T} L_1 M_1(R)))\|v_1 - v_2\|.$$

Thus, for sufficiently small $T > 0$ the mapping $\mathcal{F}$ is a contraction and there exists a unique fixed point of $\mathcal{F}$ in $\{v\colon \|v\| \le R\}$. □

Acknowledgments

Our work was supported by the DFG through SFB 701 at Bielefeld University, the RFBR projects 14-01-00237 and 15-31-20082 at Moscow State University. We are most grateful to an anonymous referee for a very thorough reading and a number of very important corrections.

References

[1] L. Ambrosio, G. Savaré, L. Zambotti, Existence and stability for Fokker–Planck equations with log-concave reference measure, Probab. Theory Related Fields 145 (3–4) (2009) 517–564.

[2] M. Annunziato, A. Borzi, Optimal control of probability density functions of stochastic processes, Math. Model. Anal. 15 (4) (2010) 393–407.
[3] M. Annunziato, A. Borzi, A Fokker–Planck control framework for multidimensional stochastic processes, J. Comput. Appl. Math. 237 (2013) 487–507.
[4] A. Arnold, P. Markowich, G. Toscani, A. Unterreiter, On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker–Planck type equations, Comm. Partial Differential Equations 26 (1–2) (2001) 43–100.
[5] A. Bensoussan, J. Frehse, Ph. Yam, Mean Field Games and Mean Field Type Control Theory, Springer, New York, 2013.
[6] V.I. Bogachev, Measure Theory, vols. 1, 2, Springer, Berlin–New York, 2007.
[7] V.I. Bogachev, Differentiable Measures and the Malliavin Calculus, Amer. Math. Soc., Providence, Rhode Island, 2010.
[8] V.I. Bogachev, G. Da Prato, M. Röckner, On parabolic equations for measures, Comm. Partial Differential Equations 33 (1–3) (2008) 397–418.
[9] V.I. Bogachev, A.I. Kirillov, S.V. Shaposhnikov, The Kantorovich and variation distances between invariant measures of diffusions and nonlinear stationary Fokker–Planck–Kolmogorov equations, Math. Notes 96 (6) (2014) 17–25.
[10] V.I. Bogachev, A.V. Kolesnikov, The Monge–Kantorovich problem: achievements, connections, and perspectives, Uspekhi Mat. Nauk 67 (5) (2012) 3–110 (in Russian); English transl.: Russian Math. Surveys 67 (5) (2012) 785–890.
[11] V.I. Bogachev, N.V. Krylov, M. Röckner, On regularity of transition probabilities and invariant measures of singular diffusions under minimal conditions, Comm. Partial Differential Equations 26 (11–12) (2001) 2037–2080.
[12] V.I. Bogachev, N.V. Krylov, M. Röckner, Elliptic and parabolic equations for measures, Russian Math. Surveys 64 (6) (2009) 973–1078.
[13] V.I. Bogachev, N.V. Krylov, M. Röckner, S.V. Shaposhnikov, Fokker–Planck–Kolmogorov Equations, Amer. Math. Soc., Rhode Island, Providence, 2015.
[14] V.I. Bogachev, M. Röckner, S.V. Shaposhnikov, Global regularity and bounds for solutions of parabolic equations for probability measures, Teor. Veroyatn. Primen. 50 (4) (2005) 652–674 (in Russian); English transl.: Theory Probab. Appl. 50 (4) (2006) 561–581.
[15] V.I. Bogachev, M. Röckner, S.V. Shaposhnikov, Estimates of densities of stationary distributions and transition probabilities of diffusion processes, Teor. Veroyatn. Primen. 52 (2) (2007) 240–270 (in Russian); English transl.: Theory Probab. Appl. 52 (2) (2008) 209–236.
[16] V.I. Bogachev, M. Röckner, S.V. Shaposhnikov, Nonlinear evolution and transport equations for measures, Dokl. Ross. Akad. Nauk 429 (1) (2009) 7–11 (in Russian); English transl.: Dokl. Math. 80 (3) (2009) 785–789.
[17] V.I. Bogachev, M. Röckner, S.V. Shaposhnikov, On uniqueness problems related to elliptic equations for measures, J. Math. Sci. (N. Y.) 176 (6) (2011) 759–773.
[18] V.I. Bogachev, M. Röckner, S.V. Shaposhnikov, On uniqueness problems related to the Fokker–Planck–Kolmogorov equation for measures, J. Math. Sci. (N. Y.) 179 (1) (2011) 7–47.
[19] V.I. Bogachev, S.V. Shaposhnikov, A.Yu. Veretennikov, Differentiability of solutions of stationary Fokker–Planck–Kolmogorov equations with respect to a parameter, Discrete Contin. Dyn. Syst. A 36 (7) (2016) 3519–3543.
[20] V.I. Bogachev, A.Yu. Veretennikov, S.V. Shaposhnikov, Differentiability of invariant measures of diffusions with respect to a parameter, Dokl. Akad. Nauk 460 (5) (2015) 507–511 (in Russian); English transl.: Dokl. Math. 91 (1) (2015) 76–79.
[21] F. Bolley, I. Gentil, Phi-entropy inequalities for diffusion semigroups, J. Math. Pures Appl. 93 (5) (2010) 449–473.
[22] F. Bolley, I. Gentil, A. Guillin, Dimensional contraction via Markov transportation distance, J. Lond. Math. Soc. (2) 90 (1) (2014) 309–332.
[23] F. Bolley, A. Guillin, C. Villani, Quantitative concentration inequalities for empirical measures on non-compact spaces, Probab. Theory Related Fields 137 (3–4) (2007) 541–593.
[24] F. Bolley, C. Villani, Weighted Csiszar–Kullback–Pinsker inequalities and applications to transportation inequalities, Ann. Fac. Sci. Toulouse Math. (6) 14 (3) (2005) 331–352.
[25] J.A. Carrillo, R.J. McCann, C. Villani, Kinetic equilibration rates for granular media and related equations: entropy, dissipation and mass transportation estimates, Rev. Mat. Iberoam. 19 (2003) 971–1018.
[26] J.A. Carrillo, G. Toscani, Exponential convergence toward equilibrium for homogeneous Fokker–Planck-type equations, Math. Methods Appl. Sci. 21 (13) (1998) 1269–1286.

[27] R.L. Dobrushin, Vlasov equations, Funktsional. Anal. i Prilozhen. 13 (2) (1979) 48–58 (in Russian); English transl.: Funct. Anal. Appl. 13 (2) (1979) 115–123.
[28] J. Dolbeault, B. Nazaret, G. Savaré, From Poincaré to logarithmic Sobolev inequalities: a gradient flow approach, SIAM J. Math. Anal. 44 (5) (2012) 3186–3216.
[29] D.A. Gomes, J. Saude, Mean field games models — a brief survey, Dyn. Games Appl. 4 (2014) 110–154.
[30] N. Gozlan, Integral criteria for transportation-cost inequalities, Electron. Commun. Probab. 11 (2006) 64–77.
[31] O. Guéant, J.-M. Lasry, P.-L. Lions, Mean field games and applications, in: Lecture Notes in Math., vol. 2003, 2011, pp. 205–266.
[32] Yu.M. Kabanov, R.Sh. Liptser, A.N. Shiryaev, On the variation distance for probability measures defined on a filtered space, Probab. Theory Related Fields 71 (1) (1986) 19–35.
[33] N.V. Krylov, E. Priola, Elliptic and parabolic second-order PDEs with growing coefficients, Comm. Partial Differential Equations 35 (1) (2010) 1–22.
[34] F. Liese, Hellinger integrals of diffusion processes, Statistics 17 (1) (1986) 63–78.
[35] F. Liese, W. Schmidt, A note on the convergence of integral functionals of diffusion processes. An application to strong convergence, Math. Nachr. 161 (1993) 283–289.
[36] F. Liese, W. Schmidt, On the strong convergence, contiguity and entire separation of diffusion processes, Stoch. Stoch. Rep. 50 (3–4) (1994) 185–203.
[37] J.L. Lions, Optimal Control of Systems Governed by Partial Differential Equations, Springer, Berlin, 1971.
[38] O.A. Manita, Estimates for the Kantorovich distances between solutions to the nonlinear Fokker–Planck–Kolmogorov equation with monotone drift, arXiv:1507.04014, 2015.
[39] O.A. Manita, M.S. Romanov, S.V. Shaposhnikov, On uniqueness of solutions to nonlinear Fokker–Planck–Kolmogorov equations, Nonlinear Anal. 128 (2015) 199–226.
[40] O.A. Manita, S.V. Shaposhnikov, Nonlinear parabolic equations for measures, Algebra i Analiz 25 (1) (2013) 64–93 (in Russian); English transl.: St. Petersburg Math. J. 25 (1) (2014) 43–62.
[41] M. Mohammadi, A. Borzi, A Hermite spectral method for a Fokker–Planck optimal control problem in an unbounded domain, Int. J. Uncertain. Quantif. 5 (3) (2015) 233–254.
[42] L. Natile, M.A. Peletier, G. Savaré, Contraction of general transportation costs along solutions to Fokker–Planck equations with monotone drifts, J. Math. Pures Appl. (9) 95 (1) (2011) 18–35.
[43] F. Otto, M. Westdickenberg, Eulerian calculus for the contraction in the Wasserstein distance, SIAM J. Math. Anal. 37 (2005) 1227–1255.
[44] E. Pardoux, A.Yu. Veretennikov, On the Poisson equation and diffusion approximation. II, Ann. Probab. 31 (3) (2003) 1166–1192.
[45] M.-K. von Renesse, K.-T. Sturm, Transport inequalities, gradient estimates, entropy and Ricci curvature, Comm. Pure Appl. Math. 68 (2005) 923–940.
[46] S.V. Shaposhnikov, On the uniqueness of the probabilistic solution of the Cauchy problem for the Fokker–Planck–Kolmogorov equation, Teor. Veroyatn. Primen. 56 (1) (2011) 77–99 (in Russian); English transl.: Theory Probab. Appl. 56 (1) (2012) 96–115.
[47] S.V. Shaposhnikov, The Fokker–Planck–Kolmogorov equations with a potential and a non-uniformly elliptic diffusion matrix, Tr. Mosk. Mat. Obs. 74 (1) (2013) 17–34 (in Russian); English transl.: Trans. Moscow Math. Soc. (2013) 15–29.
[48] A.Yu. Veretennikov, On Sobolev solutions of Poisson equations in $\mathbb{R}^d$ with a parameter, J. Math. Sci. (N. Y.) 179 (1) (2011) 48–79.
[49] C. Villani, Optimal Transport. Old and New, Springer-Verlag, Berlin, 2009.
[50] F.-Y. Wang, From super Poincaré to weighted log-Sobolev and entropy-cost inequalities, J. Math. Pures Appl. 90 (3) (2008) 270–285.