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

V.I. Bogachev[a1], M. Röckner[b], S.V. Shaposhnikov[a]

[a]*Department of Mechanics and Mathematics, Moscow State University, 119991 Moscow, Russia and National Research University Higher School of Economics, Moscow, Russia*
[b]*Fakultät für Mathematik, Universität Bielefeld, D-33501 Bielefeld, Germany*

Abstract

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 are given.

Keywords: Fokker–Planck–Kolmogorov equation, Transition probability, Diffusion, Total variation distance, Kantorovich distance
MSC: 35K10, 35K55, 60J60

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. 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 and A is locally Lipschitzian in x and locally strictly positive, i.e.,

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

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

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

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

We study solutions to the Cauchy problem

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

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 [7] or [8]) 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 Remark 1.3 just to simplify the obtained estimates). The corresponding operators will be denoted by L_μ and L_σ for brevity.

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: $\Phi=b_\mu-b_\sigma$. 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) $(1+|x|)^{-2}|a_\mu^{ij}|,\ (1+|x|)^{-1}|b_\mu| \in L^1(\mathbb{R}^d\times[0,T],\mu)$,
$(1+|x|)^{-1}|\Phi| \in L^1(\mathbb{R}^d\times[0,T],\sigma)$.

(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

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

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

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

where

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

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

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

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

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

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

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

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

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

The extra term with $H(\nu_\sigma|\nu_\mu)$ will be added also to the integral in the right-hand side of (1.5).

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

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

over all probability measures π on $\mathbb{R}^d\times\mathbb{R}^d$ with projections μ_1 and μ_2 on the factors. For $p=1$ this gives the classical Kantorovich distance (sometimes mistakenly called the Wasserstein distance); see [6] or [37].

It is known that a bound on the entropy yields many other bounds, see, e.g., [24], [37], and [38], where many additional references can be found. In the case $\varphi=1$ we obtain the usual total variation distance, hence the classical Pinsker–Csiszár–Kullback inequality (see, e.g., [3, Theorem 2.12.24])

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

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

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

established in [19], 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 [10] (see also [9, Chapter 7]). 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 [10] or (7.4.13) in [9, Chapter 7]). More precisely,

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

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:

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

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

A modification of (1.7) is given by (2.13) in [10]: for almost all $\tau \in [0, T]$ one has

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

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 [25], [26], [27], and [28] 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

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

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 [4, 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 Lipschitzian 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

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

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

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

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

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], [2], [16], [17], [18], [20], [21], [23], [30], and [33]. The principal novelty of our estimates is that they compare diffusions with different drifts or even different diffusion matrices, not with different initial distributions.

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 an application to differentiability of solutions with respect to a parameter is 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 lemma below; the corollary then follows from the next estimate established in [19]: 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 [19]:

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

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 [19] gives the extra factor 2 on the right as compared to the Pinsker–Csiszár–Kullback inequality.

Lemma 2.1. (i) *Let $f \in C_b^2(0, +\infty)$. Then, for any function $\psi \in C_0^\infty(\mathbb{R}^d)$ and any compact interval $[\tau, t] \subset (0, T)$, we have*

$$\begin{aligned}
&\int_{\mathbb{R}^d} f(v(x,t))\psi(x)\varrho_\mu(x,t)\,dx + \int_\tau^t \int_{\mathbb{R}^d} |A_\mu^{1/2} \nabla v|^2 f''(v)\psi \varrho_\mu \, dx \\
&\qquad = \int_{\mathbb{R}^d} f(v(x,\tau))\psi(x)\, \varrho_\mu(x,\tau) dx + \int_\tau^t \int_{\mathbb{R}^d} f(v) L_{A_\mu, b_\mu} \psi \varrho_\mu \, dx\, ds \\
&\qquad\qquad + \int_\tau^t \int_{\mathbb{R}^d} \bigl[\langle \Phi, \nabla v\rangle f''(v)\psi + \langle \Phi, \nabla \psi\rangle f'(v)\bigr] v \varrho_\mu \, dx\, ds.
\end{aligned} \tag{2.3}$$

(ii) *If $f'' \ge 0$ and $\psi \ge 0$, then*

$$\begin{aligned}
&\int_{\mathbb{R}^d} f(v(x,t))\psi(x)\varrho_\mu(x,t)\,dx + \frac{1}{2}\int_0^t \int_{\mathbb{R}^d} |A_\mu^{1/2} \nabla v|^2 f''(v)\psi \varrho_\mu \, dx \\
&\qquad \le f(1) \int_{\mathbb{R}^d} \psi \, d\nu + \int_0^t \int_{\mathbb{R}^d} f(v) L_{A_\mu, b_\mu} \psi \varrho_\mu \, dx\, ds \\
&\qquad + \frac{1}{2} \int_0^t \int_{\mathbb{R}^d} |A_\mu^{-1/2} \Phi|^2 f''(v) \psi v \varrho_\mu \, dx\, ds + \int_0^t \int_{\mathbb{R}^d} \langle \Phi, \nabla \psi\rangle f'(v) v \varrho_\mu \, dx\, ds.
\end{aligned} \tag{2.4}$$

In particular,

$$\int_{\mathbb{R}^d} f(v(x,t))\psi(x)\varrho_\mu(x,t)\,dx \le f(1)\int_{\mathbb{R}^d}\psi\,d\nu + \int_0^t\int_{\mathbb{R}^d} f(v)L_{A_\mu,b_\mu}\psi\varrho_\mu\,dx\,ds$$
$$+\frac{1}{2}\int_0^t\int_{\mathbb{R}^d}|A_\mu^{-1/2}\Phi|^2 f''(v)\psi v\varrho_\mu\,dx\,ds + \int_0^t\int_{\mathbb{R}^d}\langle\Phi,\nabla\psi\rangle f'(v)v\varrho_\mu\,dx\,ds. \quad (2.5)$$

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 $\int_{\mathbb{R}^d}\psi f\Bigl(\frac{d\nu_\sigma}{d\nu_\mu}\Bigr)d\nu_\mu$.

Proof. (i) Suppose first that $a_\mu^{ij}, a_\sigma^{ij}, b_\mu^i, b_\sigma^i$ 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 $\{x\colon |x|\le R\}\times(0,T)$, 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 $\partial_{x_i}\partial_{x_j}(a_\mu^{ij}\xi) - \partial_{x_i}(b_\mu^i\xi)$ 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

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

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

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

Therefore, we obtain the following equality:

$$\begin{aligned}\int_{\mathbb{R}^d} f(v(x,t))\varrho_\mu(x,t)\psi(x)\,dx &+ \int_\tau^t \int_{\mathbb{R}^d} \varrho_\mu\langle A_\mu\nabla v,\nabla v\rangle f''(v)\psi\,dx\,d\tau \\ &= \int_{\mathbb{R}^d} f(v(x,\tau))\psi(x)\varrho_\mu(x,\tau)\,dx + \int_\tau^t \int_{\mathbb{R}^d} \varrho_\mu f(v) L_{A_\mu,b_\mu}\psi\,dx\,ds \\ &\quad + \int_\tau^t \int_{\mathbb{R}^d} \bigl[\langle\Phi,\nabla v\rangle f''(v)\varrho_\sigma + f'(v)\langle\Phi,\nabla\psi\rangle\varrho_\sigma\bigr]\,dx\,ds.\end{aligned}$$

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. The general case, where the coefficients are not smooth, will be justified below.

(ii) We first show (2.4) in the situation of smooth coefficients considered so far in assertion (i) and assuming, in addition, that $\|\varrho_\mu(\,\cdot\,,\tau) - \varrho_\sigma(\,\cdot\,,\tau)\|_{L^1(\mathbb{R}^d)} \to 0$ as $\tau\to 0$. Applying in (i) the inequality

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

and taking into account that $f''\ge 0$ and $\psi\ge 0$, we obtain the inequality similar to (2.4), 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 $[0,t]$ if we let $\tau\to 0$. So we have to show that

$$\lim_{\tau\to 0}\int_{\mathbb{R}^d} f(v(x,\tau))\psi(x)\varrho_\mu(x,\tau)\,dx = f(1)\int_{\mathbb{R}^d}\psi(x)\,\nu(dx). \tag{2.6}$$

It follows from (1.3) that

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

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

$$\int_{\mathbb{R}^d}|f(v(x,\tau)) - f(1)|\varrho_\mu(x,\tau)\,dx \le C\int_{\mathbb{R}^d}|\varrho_\sigma(x,\tau)-\varrho_\mu(x,\tau)|\,dx \to 0 \quad \text{as } \tau\to 0,$$

which yields (2.6).

In the general case it suffices to show that our solutions can be obtained as limits of solutions to the Cauchy problems with coefficients and data of the respective classes in such a way that the right-sides of (2.4) for these approximations will converge to the right-side of (2.4) for the original solution.

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

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

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

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

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^ε_μ and b^ε_μ, because $A^\varepsilon_\mu(\varrho_\mu)_\varepsilon = (A_\mu\varrho_\mu)_\varepsilon$ and $b^\varepsilon_\mu(\varrho_\mu)_\varepsilon = (b_\mu\varrho_\mu)_\varepsilon$.

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 $\mathbb{R}^d$. In addition, for every fixed $\varepsilon > 0$, we have $(\varrho_\mu)_\varepsilon(x,t) \to \omega_\varepsilon * \nu(x)$ and $(\varrho_\sigma)_\varepsilon(x,t) \to \omega_\varepsilon * \nu(x)$ as $t \to 0$. Since these functions are probability densities, we obtain convergence in $L^1(\mathbb{R}^d)$, which yields that

$$\lim_{t\to 0} \int_{\mathbb{R}^d} |(\varrho_\sigma)_\varepsilon(x,t) - (\varrho_\mu)_\varepsilon(x,t)|\,dx \to 0.$$

Thus, our approximations satisfy the condition under which (2.4) has been verified above in the case of smooth coefficients. Since identity (2.3) holds for these approximations, it remains valid for the original solutions, because, as $\varepsilon \to 0$, we have $(\varrho_\mu)_\varepsilon \to \varrho_\mu$, $\partial_{x_i}(\varrho_\mu)_\varepsilon \to \partial_{x_i}\varrho_\mu$ in $L^2(U \times [\tau, t])$ and the same for σ. Thus, (2.3) is proved completely. In addition, this shows convergence of the second terms in the right-hand side of (2.4). Convergence of the first terms is trivial, the only problem is to show convergence of the third terms involving Φ. Let us show that

$$\lim_{\varepsilon\to 0} \int_0^T \int_U \frac{|(A^\varepsilon_\mu - A^\varepsilon_\sigma)\nabla(\varrho_\sigma)_\varepsilon|^2}{(\varrho_\sigma)_\varepsilon}\,dx = \int_0^T \int_U \frac{|(A_\mu - A_\sigma)\nabla\varrho_\sigma|^2}{\varrho_\sigma}\,dx. \tag{2.7}$$

For almost every fixed $t \in (0,T)$, we have $(\varrho_\mu)_\varepsilon(x,t) \to \varrho_\mu(x,t)$ in $W^{2,1}(U)$, since we have $\varrho_\mu(x,t) \in W^{2,1}(U)$. Therefore, we have convergence of the inner integrals in (2.7) for almost all $t \in (0,T)$. Let us show that the integrals over U admit a majorant integrable on $[0,T]$.

Writing

$$|\omega_\varepsilon * ((A_\mu - A_\sigma)\nabla\varrho_\sigma)| = \Bigl|\omega_\varepsilon * \Bigl(\frac{(A_\mu - A_\sigma)\nabla\varrho_\sigma}{\sqrt{\varrho_\sigma}}\sqrt{\varrho_\sigma}\Bigr)\Bigr|$$

and using the Cauchy inequality in the convolution we obtain

$$\frac{|\omega_\varepsilon * ((A_\mu - A_\sigma)\nabla\varrho_\sigma)|^2}{(\varrho_\sigma)_\varepsilon} \le \omega_\varepsilon * \biggl(\frac{|(A_\mu - A_\sigma)\nabla\varrho_\sigma|^2}{\varrho_\sigma}\biggr).$$

Therefore,

$$\int_U \frac{|\omega_\varepsilon * ((A_\mu - A_\sigma)\nabla\varrho_\sigma)(x,t)|^2}{(\varrho_\sigma)_\varepsilon(x,t)}\,dx \le \int_{U'} \frac{|(A_\mu - A_\sigma)\nabla\varrho_\sigma(x,t)|^2}{\varrho_\sigma(x,t)}\,dx,$$

where U' is the ball with the same center as U and radius increased by 1. On U we have

$$\frac{|(A^\varepsilon_\mu - A^\varepsilon_\sigma)\nabla(\varrho_\sigma)_\varepsilon|^2}{(\varrho_\sigma)_\varepsilon} \le 2I + 2J + 2K + 2\omega_\varepsilon * \biggl(\frac{|(A_\mu - A_\sigma)\nabla\varrho_\sigma|^2}{\varrho_\sigma}\biggr),$$

where

$$I = \frac{|(A^\varepsilon_\mu - A_\mu)\nabla(\varrho_\sigma)_\varepsilon|^2}{(\varrho_\sigma)_\varepsilon}, \quad J = \frac{|(A^\varepsilon_\sigma - A_\sigma)\nabla(\varrho_\sigma)_\varepsilon|^2}{(\varrho_\sigma)_\varepsilon},$$

$$K(x,t) = \frac{|\omega_\varepsilon * ((A_\mu(x,t) - A_\mu(\,\cdot\,,t) - A_\sigma(x,t) + A_\sigma(\,\cdot\,,t))\nabla\varrho_\sigma)(x,t)|^2}{(\varrho_\sigma)_\varepsilon(x,t)}$$

Since $\|A_\mu(x+y,t)-A_\mu(x,t)\|\le C|y|$ for all $x\in U$ and y with $|y|\le 1$, the terms I and J are dominated by $C^2\eta^\varepsilon * \varrho_\sigma$, where

$$\eta^\varepsilon(x)=\varepsilon^{-d}\eta(x/\varepsilon),\quad \eta=\frac{|\nabla\omega|^2}{\omega}.$$

Indeed, for all $x\in U$ we have

$$\|A^\varepsilon_\mu(x,t)-A_\mu(x,t)\|\le C\varepsilon,$$

and the Cauchy inequality yields

$$|\nabla(\varrho_\sigma)_\varepsilon(x,t)|^2=|\nabla\omega_\varepsilon*\varrho_\sigma(x,t)|^2\le\varepsilon^{-2}\Bigl(\int_{\mathbb{R}^d}|\nabla\omega(z)|\varrho_\sigma(x-\varepsilon z,t)\,dz\Bigr)^2$$
$$\le\varepsilon^{-2}(\varrho_\sigma)_\varepsilon(x,t)\int_{\mathbb{R}^d}\frac{|\nabla\omega(z)|^2}{\omega(z)}\varrho_\sigma(x-\varepsilon z,t)\,dz.$$

The same is true for J. The term K is similarly estimated by $4d^2C^2\eta_\varepsilon*\varrho_\sigma+4d^2C^2$, because

$$\int_{\mathbb{R}^d}(a^{ij}_\mu(x,t)-a^{ij}_\mu(x+y,t))\partial_{x_j}\varrho_\sigma(x+y,t)\omega_\varepsilon(y)\,dy$$
$$=\int_{\mathbb{R}^d}\partial_{y_j}a^{ij}_\mu(x+y,t)\varrho_\sigma(x+y,t)\omega_\varepsilon(y)\,dy$$
$$-\int_{\mathbb{R}^d}(a^{ij}_\mu(x,t)-a^{ij}_\mu(x+y,t))\varrho_\sigma(x+y,t)\partial_{y_j}\omega_\varepsilon(y)\,dy,$$

where the square of the first integral does not exceed $C^2(\varrho_\sigma)_\varepsilon$ and the square of the second integral does not exceed $C^2\varrho_\sigma*\eta^\varepsilon$. Indeed, $|a^{ij}_\mu(x,t)-a^{ij}_\mu(x+y,t)|\le C\varepsilon$ if $x\in U$, $|y|\le 1$, $\partial_{y_j}\omega_\varepsilon(y)=\varepsilon^{-1}\partial_{y_j}\omega(y/\varepsilon)$, and the square of the integral of $\varrho_\sigma(x+y,t)|\nabla\omega(y/\varepsilon)|$ with respect to y is estimated as above.

Therefore, we arrive at the estimate

$$\int_U\frac{|(A^\varepsilon_\mu-A^\varepsilon_\sigma)\nabla(\varrho_\sigma)_\varepsilon(x,t)|^2}{(\varrho_\sigma)_\varepsilon(x,t)}\,dx\le\widetilde{C}\int_{\mathbb{R}^d}\frac{|\nabla\omega(z)|^2}{\omega(z)}\,dz+2\int_{U'}\frac{|(A_\mu-A_\sigma)\nabla\varrho_\sigma(x,t)|^2}{\varrho_\sigma(x,t)}\,dx,$$

where $\widetilde{C}$ is a constant and the right-hand side is integrable over $[0,T]$. Now the Lebesgue dominated convergence theorem yields (2.7). Moreover, this also yields convergence of the third terms in (2.4), because there is no problem with the component in Φ corresponding to $h_\mu-h_\sigma$. Convergence of the last terms on the right in (2.4) follows from what we have just proved. □

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

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

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 (2.5) and the equality $f''(v)=1/v$, we arrive at the desired estimate. However, this function f is not of class C_b^∞ 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 $\psi\in C_0^\infty(\mathbb{R}^d)$, $\psi\ge 0$ and $\psi(x)=1$ whenever $|x|<1$. Let us substitute in (2.5) 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 $|f_m(v)| \le C(m,k)v$ and

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

Since $|L_{A_\mu,b_\mu}\psi| \le N^{-2}|a_\mu^{ij}|\,|\partial^2_{x_i x_j}\psi| + N^{-1}|b_\mu|\,|\nabla\psi|$, 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,

$$\int_0^t \int_{\mathbb{R}^d} \langle \Phi, \nabla\psi_N\rangle \log\big((k^{-1}\vee v)\wedge m\big)\,d\sigma_s\,ds$$
$$\le (\log k + \log m)\max|\nabla\psi| \int_0^t \int_{\mathbb{R}^d} (1+|x|)^{-1}|\Phi|\,d\sigma_s\,ds.$$

Finally, note that

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

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 N, m and k in (2.5) 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 (2.5) the functions f_m 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

$$\int_0^t \int_{\mathbb{R}^d} \varrho_\mu(f(v) - \alpha v - \beta)L_{A_\mu,b_\mu}\psi\,dx\,ds$$
$$= \int_0^t \int_{\mathbb{R}^d} \varrho_\mu f(v)L_{A_\mu,b_\mu}\psi\,dx\,ds - \alpha\int_0^t \int_{\mathbb{R}^d} \langle\Phi,\nabla\psi\rangle\varrho_\sigma\,dx\,ds$$
$$+ \beta\int_{\mathbb{R}^d} \psi(x)\varrho_\mu(x,t)\,dx - \beta\int_{\mathbb{R}^d}\psi\,d\nu$$
$$+ \alpha\int_{\mathbb{R}^d}\psi(x)v(x,t)\varrho_\mu(x,t)\,dx - \alpha\int_{\mathbb{R}^d}\psi\,d\nu. \quad (2.8)$$

Applying (2.8) with $\alpha = \log m$ and $\beta = k^{-1}$ we have

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

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

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

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 $\delta \in (0,1)$

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

As above, we conclude that

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

Thus, passing to the limit in N, m and k we obtain the desired bound. ☐

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

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

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

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

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 $V(x) = 1 + |x|^2$ 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 μ and σ. The latter can be verified by using Lyapunov functions (see, e.g., [5], [8], and [14]). 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 particular, for $\varphi = 1$ we obtain that

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

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., [35])

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

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 $T, m, \lambda_i, \gamma_i, \int |x|^{2m} \, d\nu$, and $\|\varrho_0 \ln \varrho_0\|_{L^1(\mathbb{R}^d)}$.

Proof. This follows easily from Corollary 3.1 combined with estimate (1.7) and Gronwall's inequality. □

Remark 3.6. We observe that passing from (2.4) to (2.5) 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 [10]), 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. The previous notation b_μ 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. The 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 $\alpha \in C^+[0,\tau_0]$ and $\tau \in (0,T]$ we set

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

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

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

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

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

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

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

According to Remark 3.4 we have

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

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

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

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

Applying Corollary 1.2 and Remark 3.4 we obtain that

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

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

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

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

Note that under different assumptions a solution to a nonlinear Vlasov equation was constructed in [22] by employing the contraction mapping theorem for the Kantorovich norm. The existence of not necessarily unique solutions has been proved in [12] and [29] 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 $\tau \in (0,\tau_0]$ 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 [7]), 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 $\lim\limits_{|x|\to\infty} \Psi_\mu(x) = +\infty$, $|\nabla\Psi_\mu|\Psi_\mu^{-1}V^{-1}$ 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 $\|W(\mu_t - \sigma_t)\|_{TV} = 0$. □

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 [32], [36] and the recent paper [15].

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 $\alpha_0 \in [0, 1]$

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

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

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

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

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

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 [7], 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

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

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

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

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] 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.

[3] V.I. Bogachev, Measure Theory, V. 1,2, Springer, Berlin – New York, 2007.

[4] V.I. Bogachev, Differentiable Measures and the Malliavin Calculus, Amer. Math. Soc., Providence, Rhode Island, 2010.

[5] V.I. Bogachev, G. Da Prato, M. Röckner, On parabolic equations for measures, Comm. Partial Differential Equations 33 (1-3) (2008) 397–418.

[6] V.I. Bogachev, A.V. Kolesnikov, The Monge–Kantorovich problem: achievements, connections, and perspectives, Uspehi Matem. Nauk 67 (5) (2012) 3–110 (in Russian); English transl.: Russian Math. Surveys 67 (5) (2012) 785–890.

[7] 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.

[8] V.I. Bogachev, N.V. Krylov, M. Röckner, Elliptic and parabolic equations for measures, Russian Math. Surveys 64 (6) (2009) 973–1078.

[9] V.I. Bogachev, N.V. Krylov, M. Röckner, S.V. Shaposhnikov, Fokker–Planck–Kolmogorov equations, Institute of Computer Sciences, Moscow – Izhevsk, 2013 (in Russian); English translation: Amer. Math. Soc. (to appear).

[10] V.I. Bogachev, M. Röckner, S.V. Shaposhnikov, Global regularity and bounds for solutions of parabolic equations for probability measures, Teor. Verojatn. Primen. 50 (4) (2005) 652–674 (in Russian); English transl.: Theory Probab. Appl. 50 (4) (2006) 561–581.

[11] V.I. Bogachev, M. Röckner, S.V. Shaposhnikov, Estimates of densities of stationary distributions and transition probabilities of diffusion processes, Teor. Verojatn. i Primen. 52 (2) (2007) 240–270 (in Russian); English transl.: Theory Probab. Appl. 52 (2) (2008) 209–236.

[12] 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.

[13] V.I. Bogachev, M. Röckner, S.V. Shaposhnikov, On uniqueness problems related to elliptic equations for measures, J. Math. Sci. (New York) 176 (6) (2011) 759–773.

[14] V.I. Bogachev, M. Röckner, S.V. Shaposhnikov, On uniqueness problems related to the Fokker–Planck–Kolmogorov equation for measures, J. Math. Sci. (New York) 179 (1) (2011) 7–47.

[15] V.I. Bogachev, A.Yu. Veretennikov, S.V. Shaposhnikov, Differentiability of invariant measures of diffusions with respect to a parameter, Doklady Akad. Nauk 460 (5) (2015) 507–511 (in Russian); English transl.: Dokl. Math. 91 (1) (2015) 76–79.

[16] F. Bolley, I. Gentil, Phi-entropy inequalities for diffusion semigroups, J. Math. Pures Appl. 93 (5) (2010) 449–473.

[17] F. Bolley, I. Gentil, A. Guillin, Dimensional contraction via Markov transportation distance, J. London Math. Soc. (2) 90 (1) (2014) 309–332.

[18] 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.

[19] 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.

[20] J.A. Carrillo, R.J. McCann, C. Villani, Kinetic equilibration rates for granular media and related equations: entropy, dissipation and mass transportation estimates, Rev. Math. Iberoamer. 19 (2003) 971–1018.

[21] J.A. Carrillo, G. Toscani, Exponential convergence toward equilibrium for homogeneous Fokker–Planck–type equations, Math. Methods Appl. Sci. 21 (13) (1998) 1269–1286.
[22] R.L. Dobrushin, Vlasov equations, Funk. Anal. i Prilozhen. 13 (2) (1979) 48–58 (in Russian); English transl.: Functional Anal. Appl. 13 (2) (1979) 115–123.
[23] 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.
[24] N. Gozlan, Integral criteria for transportation-cost inequalities, Electron. Commun. Probab. 11 (2006) 64–77.
[25] Yu.M. Kabanov, R.Sh. Liptser, A.N. Shiryaev, On the variation distance for probability measures defined on a filtered space, Probab. Theory Relat. Fields 71 (1) (1986) 19–35.
[26] F. Liese, Hellinger integrals of diffusion processes, Statistics 17 (1) (1986) 63–78.
[27] 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.
[28] F. Liese, W. Schmidt, On the strong convergence, contiguity and entire separation of diffusion processes, Stochastics Stochastics Rep. 50 (3-4) (1994) 185–203.
[29] 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.
[30] 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.
[31] F. Otto, M. Westdickenberg, Eulerian calculus for the contraction in the Wasserstein distance, SIAM J. Math. Anal. 37 (2005) 1227–1255.
[32] E. Pardoux, A.Yu. Veretennikov, On the Poisson equation and diffusion approximation. II, Ann. Probab. 31 (3) (2003) 1166–1192.
[33] M.-K. von Renesse, K.-T. Sturm, Transport inequalities, gradient estimates, entropy and Ricci curvature, Comm. Pure Appl. Math. 68 (2005) 923–940.
[34] 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.
[35] S.V. Shaposhnikov, The Fokker–Planck–Kolmogorov equations with a potential and a non-uniformly elliptic diffusion matrix, Trudy Mosk. Matem. Ob. 74 (1) (2013) 17–34 (in Russian); English transl.: Trans. Moscow Math. Soc. (2013) 15–29.
[36] A.Yu. Veretennikov, On Sobolev solutions of Poisson equations in $\mathbb{R}^d$ with a parameter, J. Math. Sci. (New York) 179 (1) (2011) 48–79.
[37] C. Villani, Optimal Transport. Old and New, Springer-Verlag, Berlin, 2009.
[38] F.-Y. Wang, From super Poincaré to weighted log-Sobolev and entropy-cost inequalities, J. Math. Pures Appl. 90 (3) (2008) 270–285.