Convergence in variation of solutions of nonlinear Fokker–Planck–Kolmogorov equations to stationary measures[1]

V.I. Bogachev[a2], M. Röckner[b], S.V. Shaponshnikov[a]

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

Abstract We study convergence in variation of probability solutions of nonlinear Fokker–Planck–Kolmogorov equations to stationary solutions. We obtain sufficient conditions for the exponential convergence of solutions to the stationary solution in case of coefficients that can have an arbitrary growth at infinity and depend on the solutions through convolutions with unbounded discontinuous kernels. In addition, we study a more difficult case where the nonlinear equation has several stationary solutions and convergence to a stationary solution depends on initial data. Finally, we obtain sufficient conditions for solvability of nonlinear Fokker–Planck–Kolmogorov equations.

Keywords: Nonlinear Fokker–Planck–Kolmogorov equation, stationary measure, exponential convergence

MSC: 60J75, 47G20, 60G52

1. Introduction

We consider the Cauchy problem for a nonlinear Fokker–Planck–Kolmogorov equation

$$\partial_t \mu_t = \sum_{1\le i,j\le d} \partial_{x_i}\partial_{x_j}\bigl(a^{ij}(x)\mu_t\bigr) - \sum_{i=1}^d \partial_{x_i}\bigl(b^i(x,\mu_t)\mu_t\bigr), \quad \mu_0=\nu, \tag{1.1}$$

in which the nonlinearity originates from the dependence of the drift term b on the unknown solution. This equation is understood in the sense of the integral identity

$$\int_{\mathbb{R}^d} \varphi(x)\,\mu_t(dx) - \int_{\mathbb{R}^d}\varphi(x)\,\nu(dx)$$
$$= \int_0^t \int_{\mathbb{R}^d} [\mathrm{trace}\,(A(x)D_x^2\varphi(x)) + \langle b(x,\mu_s),\nabla\varphi(x)\rangle]\,\mu_s(dx)\,ds, \quad \varphi\in C_0^\infty(\mathbb{R}^d). \tag{1.2}$$

Solutions to the stationary equation

$$\sum_{1\le i,j\le d} \partial_{x_i}\partial_{x_j}\bigl(a^{ij}(x)\mu_t\bigr) - \sum_{i=1}^d \partial_{x_i}\bigl(b^i(x,\mu_t)\mu_t\bigr) = 0 \tag{1.3}$$

are defined similarly by means of the integral identity

$$\int_{\mathbb{R}^d} [\mathrm{trace}\,(A(x)D_x^2\varphi(x)) + \langle b(x,\mu),\nabla\varphi(x)\rangle]\,\mu(dx) = 0, \quad \varphi\in C_0^\infty(\mathbb{R}^d). \tag{1.4}$$

For a recent detailed presentation of the theory of linear Fokker–Planck–Kolmogorov equations, see [10], where also some comments on nonlinear problems can be found.

Our main result (Theorem 3.1) gives sufficient conditions for the existence of a stationary probability solution μ and the exponential convergence to this stationary solution in total variation norm. To be more precise, we obtain a bound

$$\|W\cdot(\mu_t-\mu)\|_{TV} \le \alpha_1 e^{-\alpha_2 t}$$

[1]This paper has been supported by the RFBR Grant 17-01-00662, the RF President Grant MD-207.2017.1, the Simons Foundation. Financial support by the DFG through the CRC 1283 "Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications" is acknowledged.

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

with a suitable growing function W (so that the left-hand side dominates the usual total variation norm). Informally, our results are of the following nature: the drift term b in the nonlinear equation depends on a parameter $\varepsilon \geq 0$ such that for $\varepsilon = 0$ the equation has certain nice properties, for example, becomes linear (actually, the situation is more general) with reasonable properties, then for ε small enough both the original nonlinear equation and the stationary equation are solvable and we have exponential convergence in total variation norm with a weight.

First of all we consider the following very typical example demonstrating phenomena arising in the study of convergence of solutions of a nonlinear Fokker–Planck–Kolmogorov equation to the stationary distribution.

Example 1.1. Let $d = 1$, $A = I$ and $b(x, \mu) = -x + \varepsilon B(\mu)$, where

$$B(\mu) = \int_{\mathbb{R}} x\, \mu(dx).$$

In case $\varepsilon < 1$ the unique solution of the stationary equation is the standard Gaussian measure μ (see, e.g., [9]). One can show that the transition probabilities μ_t forming the solution to the Cauchy problem (1.1), for every initial condition ν (with a finite first moment), converge exponentially to the stationary measure. This is discussed in Remark 3.7 along with the case $\varepsilon > 1$. If $\varepsilon = 1$, then every measure μ given by a density

$$\varrho_a(x) = \frac{1}{\sqrt{2\pi}} \exp\bigl(-|x-a|^2/2\bigr), \quad a \in \mathbb{R}^d,$$

satisfies the stationary equation. It is readily seen that the measures μ_t converge to that stationary measure which has the same mean as ν. Indeed, in the case under consideration the mean of μ_t does not depend on time and coincides with the mean of ν. Therefore, if the mean of ν coincides with that of μ, then the mean of μ_t coincides with the mean of μ, i.e., $B(\mu_t) = B(\mu)$, and the measures μ_t satisfy the linear Fokker–Planck–Kolmogorov equation corresponding to the Ornstein–Uhlenbeck type operator, for which convergence to the solution of the stationary equation is well known.

Thus, already in this very simple one-dimensional example we see that convergence to the stationary distribution depends not only on the form of the nonlinearity, but also on the initial condition. Moreover, an important role is played by certain quantities invariant along the trajectories of solutions to the Fokker–Planck–Kolmogorov equation. Note that the existence of a stationary solution and convergence to it are not stable under small perturbations of the coefficients. For example, if $b(x, \mu) = -x + \delta + \varepsilon B(\mu)$ with arbitrarily small $\delta > 0$, then for $\varepsilon = 1$ there are no stationary solutions, because for the solution μ we must have the equality $(1 - \varepsilon)B(\mu) = \delta$.

In this paper we show that the picture described in this special example takes place in a very general situation. Certainly, in the general case it becomes difficult to take into account specific features of concrete equations, but some typical cases can be singled out.

In the example considered above the coefficient $b(x, \mu)$ has the form $b_0(x) + \varepsilon b_1(x, \mu)$. Convergence of solutions to the Cauchy problem for nonlinear Fokker–Planck–Kolmogorov equations with drift coefficients of such a form have been studied in the paper [14], where it has been shown that convergence to the stationary distribution takes place in case of a sufficiently small number ε, provided that the coefficients are Lipschitz in x and Lipschitz in μ with respect to the Kantorovich metric. In addition, the term b_1 has been assumed to be globally bounded. In the paper [22] a similar result has been obtained with the aid of the method of coupling in the case where

$$\langle b_0(x) - b_0(y), x - y\rangle \leq -\kappa(|x-y|)|x-y|,$$

$$b_1(x, \mu) = \int_{\mathbb{R}^d} K(x, y)\, d\mu, \quad |K(x, y) - K(x', y')| \leq C(|x - x'| + |y - y'|),$$

i.e., the global boundedness of b_1 has been weakened by means of the monotonicity condition for b_0. The smallness of the parameter ε is important not only for convergence, but also for the existence of a stationary distribution, which is seen from the example above. However, the situation is actually more complicated (and this is also seen from the example above) and convergence depends on the initial condition and the stationary measure. An important role in determining conditions under which one has convergence is played by certain quantities that are invariant or unboundedly increasing along trajectories of the Fokker–Planck–Kolmogorov equation (in our example such a quantity is the centre of mass). Finding such quantities enables one to single out classes of initial conditions for which one has convergence to the stationary distribution or to prove that there is no such convergence. In the case of a nondegenerate diffusion the assumption of Lipschitzness of $b(x,\mu)$ in μ with respect to the metric employed in [14] and [22] becomes superfluous, because any solution possesses a density and it is more natural to replace the Kantorovich metric by the total variation distance.

In this paper we study convergence in variation. The next two examples illustrate our main results. Let $m \geq 1$ and

$$b_\varepsilon(x,\mu) = b_0(x) + \varepsilon \int_{\mathbb{R}^d} K(x,y)\,\mu(dy),$$

where for all $x, y \in \mathbb{R}^d$,

$$\langle b_0(x), x\rangle \leq c_1 - c_2|x|^2, \quad \langle K(x,y), x\rangle \leq c_3 + c_3|x|^2, \quad C_3 < c_2,$$

$$|b_0(x)| \leq c_4 + c_4|x|^m, \quad |K(x,y)| \leq c_5(1+|x|^m)(1+|y|^m).$$

Then, there is $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0)$ there exists a solution μ to the stationary Fokker–Planck–Kolmogorov equation (1.3) with the coefficients $A = I$ and $b_\varepsilon(x,\mu)$. Moreover, for every probability measure ν such that $(1+|x|)^{2m+1} \in L^1(\nu)$, the solutions μ_t to the Cauchy problem (1.1) with initial data ν converge to μ as $t \to +\infty$ and

$$\|(1+|x|^m)(\mu_t - \mu)\|_{TV} \leq \alpha_1 e^{-\alpha_2 t}, \quad \alpha_1, \alpha_2 > 0.$$

Our second example concerns the case where the stationary equation has several solutions. Let $d = 2$, $x = (x_1, x_2)$, $y = (y_1, y_2)$ and $b_\varepsilon(x,\mu) = (b_\varepsilon^1(x,\mu), b_\varepsilon^2(x,\mu))$, where

$$b_\varepsilon^1(x,\mu) = -2x_1 + \int_{\mathbb{R}^2} (y_1+y_2)\,\mu(dy) + \varepsilon \int_{\mathbb{R}^2} H(x,y)\,\mu(dy),$$

$$b_\varepsilon^2(x,\mu) = -2x_2 + \int_{\mathbb{R}^2} (y_1+y_2)\,\mu(dy) - \varepsilon \int_{\mathbb{R}^2} H(x,y)\,\mu(dy),$$

with some bounded Borel function $H\colon \mathbb{R}\times\mathbb{R} \to \mathbb{R}$.

Then, for every number $Q_0 > 0$, there is a number $\varepsilon_0 > 0$, depending only on Q_0, such that for any $\varepsilon \in (0, \varepsilon_0)$ and $Q \in (-Q_0, Q_0)$ there exists a solution μ to the stationary Fokker–Planck–Kolmogorov equation (1.3) with the coefficients $A = I$ and $b_\varepsilon(x,\mu)$ for which

$$\int_{\mathbb{R}^2} (y_1+y_2)\,\mu(dy) = Q.$$

Moreover, for every probability measure ν such that $|x|^2 \in L^1(\nu)$ and

$$\int_{\mathbb{R}^2} (y_1+y_2)\,\nu(dy) = Q,$$

the solutions μ_t to the Cauchy problem (1.1) with initial data ν converge to μ as $t \to +\infty$ and

$$\|(1+|x|)(\mu_t - \mu)\|_{TV} \leq \alpha_1 e^{-\alpha_2 t}, \quad \alpha_1, \alpha_2 > 0.$$

More general conditions and examples are discussed after Theorems 3.1 and 3.3.

Note that it is often simpler, and in the case of a degenerate diffusion matrix more natural, to consider convergence in the Kantorovich metric (see Remark 3.8 below). Results of this sort for non-gradient drift coefficients were apparently first obtained in [1] and have been recently generalized in [22], [39], and [41]. See also the related, but more special papers [6] and [7]. The gradient case, where $b = \nabla V$, has been studied in many papers, starting from [19], [36], [37] and further studied in many papers on the theory of gradient flows (see [2], [13], [15], and [16]). In the theory of gradient flows an important role is played by the Kantorovich 2-metric and the geometry of the space of probability measures connected with this metric.

There is a vast literature devoted to nonlinear Fokker–Planck–Kolmogorov equations (see, e.g., [23]). It should be emphasized that in this paper we study equations with nonlocal nonlinearities (of the type of the so-called McKean–Vlasov equations), the investigation of which was initiated in the well-known papers [27], [32], [33], [24] and continued by many researchers. This circumstance explains the character of our assumptions about the drift, which are quite natural for such nonlinearities. For instance, the continuity of $b(\mu, x)$ in μ in total variation norm holds when $b(\mu, x)$ depends on μ through convolution, but not when $b(\mu, x)$ depends on the value of the density of μ at x. Existence and uniqueness of solutions and properties of the distributions of stochastic McKean–Vlasov equations (distributions of such equations satisfy nonlinear Fokker–Planck–Kolmogorov equations) are discussed, e.g., in [34] and [41]. In particular, in [41], sufficient conditions (monotonicity of the coefficient b, as in [22]) for the existence and uniqueness of solutions are given and convergence in the Kantorovich 2-metric to the stationary distribution is shown.

Existence and uniqueness of solutions to nonlinear Fokker–Planck–Kolmogorov equations with irregular and rapidly growing coefficients have been discussed in the recent papers [30] and [31], which also contain some examples of non-uniqueness. The papers [11] and [29] develop an approach to nonlinear equations based on estimates of distances between solutions to linear Fokker–Planck–Kolmogorov equations with different diffusion matrices and different drift coefficients. Analogous questions for nonlinear stationary Fokker–Planck–Kolmogorov equations are studied in [8], [9], and [38], where the existence of stationary solutions is proved with the aid of fixed point theorems applied to the nonlinear mapping that maps a probability measure σ to the solution μ of the linear equation with the drift coefficient $b(x, \sigma)$. The phenomenon of nonuniqueness of a stationary measure is investigated in [26], where certain explicit nonlinear expressions for stationary measures are written out in the gradient case. In this paper we also obtain some generalizations of existence and uniqueness results for stationary and parabolic Fokker–Planck–Kolmogorov equations.

The problem of convergence to the stationary measure for a linear Fokker–Planck–Kolmogorov equation has been thoroughly studied, and one can single out the following three approaches:

1) the approach based on the Harris theorem or the Meyn–Tweedie approach with Lyapunov functions (see, e.g., [20] and [25]),

2) the approach based on entropy estimates and Poincaré and Sobolev inequalities (see, e.g., [3], [4], [5], [12], [17], [35], and [40]),

3) the probabilistic approach based on coupling (see, e.g., [21], [22], [18], and [28]).

In this paper we employ the first approach and for verification of the conditions of the Harris theorem we use certain estimates for transition probabilities from [10, Chapter 8]. This enables us to substantially weaken the assumptions about the regularity of coefficients, but, on the other hand, some weak points of the Meyn–Tweedie approach connected with a complicated dependence of constants on the coefficients of the equation remain also in our case.

Throughout we assume that the matrix $A(x) = (a^{ij}(x))_{1\le i,j\le d}$ is symmetric and there exist numbers $K_1 > 0$ and $K_2 > 0$ such that

$$K_1^{-1} I \le A(x) \le K_1 I, \quad |A(x) - A(y)| \le K_2 |x - y|.$$

Let $V \in C^2(\mathbb{R}^d)$, $V \ge 1$ and $\lim\limits_{|x|\to+\infty} V(x) = +\infty$. Let $\mathcal{P}_V(\mathbb{R}^d)$ denote the space of all probability measures μ on $\mathbb{R}^d$ such that

$$\int_{\mathbb{R}^d} V\, d\mu < \infty.$$

Set

$$W(x) = V(x)^\gamma, \quad \gamma \in (0, 1/2].$$

Typical examples are $V(x) = 1 + |x|^{2m}$. We recall that the total variation norm of a finite (possibly, signed) measure σ is defined by

$$\|\sigma\|_{TV} = |\sigma|(\mathbb{R}^d),$$

where $|\sigma| = \sigma^+ + \sigma^-$ and $\sigma = \sigma^+ - \sigma^-$ is the Hahn decomposition into the difference of mutually singular nonnegative measures. The symbol $W \cdot \mu$ denotes the measure given by the density W with respect to the measure μ. Set

$$\|\mu\|_W = \|W \cdot \mu\|_{TV}.$$

Suppose that for every $\mu \in \mathcal{P}_V(\mathbb{R}^d)$ we have a Borel vector field $b(x,\mu) = (b^i(x,\mu))_{1\le i\le d}$ on $\mathbb{R}^d$ such that there exists a number $C(\mu)$ for which

$$|b(x,\mu)| \le C(\mu)V(x)^{1-\gamma}.$$

It will be assumed below that b satisfies certain additional conditions.

We say that a family $\{\mu_t\}_{t\in[0,T]}$ of probability measures $\mu_t \in \mathcal{P}_V(\mathbb{R}^d)$ satisfies the Cauchy problem (1.1) on $[0,T]$, where $T>0$ is fixed, if equality (1.2) holds. A measure $\mu \in \mathcal{P}_V(\mathbb{R}^d)$ is called a solution to the stationary equation (1.3) if equality (1.4) is fulfilled.

Set

$$L_\mu\varphi(x) = \mathrm{trace}(A(x)D^2\varphi(x)) + \langle b(x,\mu), \nabla\varphi(x)\rangle.$$

Throughout for the stationary equation (1.3) and for the parabolic equation (1.1) we use the shortened equalities $L_\mu^*\mu = 0$ and $\partial_t\mu_t = L_{\mu_t}^*\mu_t$. Similarly we write linear equations $L_\sigma^*\mu = 0$ and $\partial_t\mu_t = L_\sigma^*\mu_t$ with the coefficient $b(x,\sigma)$. Solutions to linear Fokker–Planck–Kolmogorov equations are defined precisely as in the case of general nonlinear equations by means of integral equalities of the form (1.2) and (1.4).

This paper consists of the introduction and three sections. In Section 2 we discuss some classes of functions ψ on $\mathbb{R}^d$ such that the integral of ψ against μ_t for solutions $\{\mu_t\}$ to the Cauchy problem (1.1) is constant or equals a constant multiplied by a function of the form $\exp(\lambda t)$.

With the aid of such functions one can formulate simple tests to show that convergence to stationary solutions fails to hold. In addition, if we know that

$$\int_{\mathbb{R}^d} \psi\, d\mu_t \equiv \int \psi\, d\nu$$

with μ_t satisfying our Cauchy problem and

$$b(x,\mu) = b_0(x) + \int_{\mathbb{R}^d} \psi\, d\mu,$$

then

$$b(\mu_t) = b_0(x) + const$$

i.e., along solutions the drift depends only on x and actually is independent of μ_t.

At the end of Section 2 we formulate our main conditions on the coefficients of the equation. In Section 3 we formulate and prove the main results of the paper. The first main result (Theorem 3.1) enables us to determine by the initial condition and the stationary solution whether there is convergence of solutions of the Cauchy problem (1.1) with this initial condition to the stationary solution. The second main result (Theorem 3.3) gives sufficient conditions under which a stationary solution exists and for every initial condition (having a finite moment of a suitable order) the solutions of the Cauchy problem converge to this stationary solution. In Section 4 we discuss conditions for the existence of solutions to the stationary equation and the Cauchy problem.

2. Invariant and subinvariant functions

Here we consider certain conservation laws for solutions.

For a function W as above, let I_0^W denote the set of functions $\psi \in C^2(\mathbb{R}^d)$ such that

$$\sup_x \Bigl(|\psi(x)| + |\nabla\psi(x)| + |D^2\psi(x)|\Bigr)W(x)^{-1} < \infty \tag{2.1}$$

and for every measure $\mu \in \mathcal{P}_V(\mathbb{R}^d)$ we have

$$\int_{\mathbb{R}^d} L_\mu\psi\, d\mu = 0. \tag{2.2}$$

It is clear that I_0^W is a linear space containing 1. Further for brevity we use the notation

$$\mu(\psi) := \int_{\mathbb{R}^d} \psi\, d\mu.$$

Proposition 2.1. (i) *If $\psi \in I_0^W$ and $\{\mu_t\}$ is a solution to the Cauchy problem* (1.1) *with the initial condition ν, then $\mu_t(\psi) = \nu(\psi)$.*

(ii) *If $\mu \in \mathcal{P}_V(\mathbb{R}^d)$ is a solution to the stationary equation* (1.3) *and*

$$\nu(\psi) \neq \mu(\psi)$$

for some function $\psi \in I_0^W$, then the solutions μ_t to the Cauchy problem (1.1) *with the initial condition ν do not converge to μ with respect to the norm $\|\cdot\|_W$.*

Proof. Assertion (ii) follows from (i). For justifying (i) it suffices to observe that

$$\mu_t(\psi) - \nu(\psi) = \int_0^t \int_{\mathbb{R}^d} L_{\mu_s}\psi\, d\mu_s\, ds = 0,$$

which follows from the equation. □

Let us consider an important example where $A = I$ and

$$b(x,\mu) = -\int_{\mathbb{R}^d} K(x,y)\,\mu(dy),$$

where K is a vector-valued mapping.

Proposition 2.2. *A function ψ satisfying* (2.1) *belongs to I_0^W if and only if for all x, y we have*

$$\Delta\psi(x) + \Delta\psi(y) - \langle K(x,y), \nabla\psi(x)\rangle - \langle K(y,x), \nabla\psi(y)\rangle = 0.$$

In particular, if $\psi \in I_0^W$, then $\Delta\psi(x) - \langle K(x,x), \nabla\psi(x)\rangle = 0$. Moreover, if

$$(Q, K(x,y)) = -(Q, K(y,x))$$

for some constant vector Q and $W(x)$ is growing not more slowly than $|x|$, then I_0^W contains all functions of the form $\psi(x) = (Q,x) + g$, where g is a constant number.

Proof. We observe that (2.2) is equivalent to the equality

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \Bigl[\Delta\psi(x) + \Delta\psi(y) - \langle K(x,y), \nabla\psi(x)\rangle - \langle K(y,x), \nabla\psi(y)\rangle\Bigr]\, \mu\otimes\mu(dx\,dy) = 0,$$

which holds for every probability measure $\mu \in \mathcal{P}_V(\mathbb{R}^d)$ if and only if the expression under the integral sign is skew symmetric in x and y. Since we have a symmetric function there, it must vanish. □

Example 2.3. (i) Let $d = 1$. The space of solutions to the equation $\psi'' - K(x,x)\psi' = 0$ is the linear span of 1 and the primitive of the function

$$\exp\Bigl(\int_0^x K(s,s)\,ds\Bigr).$$

A nonconstant function ψ satisfies the condition of Proposition 2.2 if and only if for all x, y we have

$$K(x,x) + K(y,y) = K(x,y) + K(y,x).$$

The latter relation is satisfied, for example, for the functions $K(x,y) = H(x-y)$, where H is an odd function. In this case I_0^W is the linear span of 1 and x.

(ii) Let $d \geq 1$ and

$$K(x,y) = -Rx + \langle v, y\rangle h + H(x,y),$$

where R is a constant matrix, v and h are constant vectors and

$$R^*v = \lambda v, \quad \langle v, h\rangle = \lambda, \quad \langle H(x,y), v\rangle = 0.$$

Then the function $x \to \langle v, x\rangle$ belongs to I_0^W. Indeed, we have

$$\langle v, K(x,y)\rangle = -\lambda\langle v, x\rangle + \lambda\langle v, y\rangle = -\langle v, K(y,x)\rangle.$$

Let I_+^W denote the set of all functions $\psi \in C^2(\mathbb{R}^d)$ such that ψ satisfies condition (2.1) and there exists a number $\lambda = \lambda(\psi) > 0$ such that

$$\int_{\mathbb{R}^d} L_\mu\psi\, d\mu = \lambda \int_{\mathbb{R}^d} \psi\, d\mu \quad \forall\, \mu \in \mathcal{P}_V(\mathbb{R}^d). \tag{2.3}$$

Proposition 2.4. (i) *If $\psi \in I_+^W$ and μ_t is a solution to the Cauchy problem* (1.1) *with the initial condition ν, then $\mu_t(\psi) = \nu(\psi)e^{\lambda(\psi)t}$.*

(ii) *If $\mu \in \mathcal{P}_V(\mathbb{R}^d)$ is a solution to the stationary equation* (1.3) *and $\psi \in I_+^W$, then $\mu(\psi) = 0$.*

(iii) *If $\nu(\psi) \neq 0$ for some $\psi \in I_+^W$, then the solutions μ_t to the Cauchy problem* (1.1) *do not converge to the stationary solution with respect to the norm $\|\cdot\|_W$.*

Proof. Assertion (ii) is obvious. Assertion (iii) follows from (i), and (i) is deduced from the equality

$$\mu_t(\psi) - \nu(\psi) = \int_0^t \int_{\mathbb{R}^d} L_{\mu_s}\psi \, d\mu_s \, ds = \lambda \int_0^t \mu_s(\psi)\, ds,$$

satisfied by the solution. □

As above, let us consider the case where $A = I$ and

$$b(x,\mu) = -\int_{\mathbb{R}^d} K(x,y)\, \mu(dy)$$

with a vector-valued mapping K.

Proposition 2.5. *A function ψ satisfying* (2.1) *belongs to I_+^W if and only if for some $\lambda > 0$ and for all x, y we have*

$$\Delta\psi(x) + \Delta\psi(y) - (K(x,y), \nabla\psi(x)) - (K(y,x), \nabla\psi(y)) = \lambda(\psi(x) + \psi(y)).$$

In particular, if $\psi \in I_+^W$, then $\Delta\psi(x) - (K(x,x), \nabla\psi(x)) = \lambda\psi(x)$.

Proof. The same reasoning as in the case of I_0^W works. □

Let $d = 1$ and $K(x,x) = -qx$, where q is a positive constant. Then the equation on the function ψ takes the form

$$\psi'' + qx\psi' = \lambda\psi.$$

If $\lambda = q$, then $\psi(x) = x$ is a solution. If $K(x,x) = 0$, then the equation takes the form $\psi'' = \lambda\psi$ and linear combinations of the exponents $e^{\sqrt{\lambda}x}$ and $e^{-\sqrt{\lambda}x}$ are all solutions.

Proposition 2.6. *Suppose that for some function $\psi \in I_0^W$ there exists a continuous function h such that $\sup_x |h(x)|/V(x) < \infty$ and for every probability measure $\sigma \in \mathcal{P}_V(\mathbb{R}^d)$ we have*

$$L_\sigma\psi(x) = C_1(\sigma)h(x) + C_2(\sigma), \quad C_1(\sigma) \neq 0$$

with some numbers $C_1(\sigma)$ and $C_2(\sigma)$. Suppose that $\mu \in \mathcal{P}_V(\mathbb{R}^d)$ satisfies the stationary equation $L_\sigma^*\mu = 0$. Then $\mu(h) = \sigma(h)$. The analogous assertion is true if $\psi \in I_+^W$ and $\sigma(\psi) = 0$.

Proof. By the definition of I_0^W and the fact that μ is a solution to the stationary equation $L_\sigma^*\mu = 0$ we have the equalities

$$C_1(\sigma)\int_{\mathbb{R}^d} h\, d\sigma + C_2(\sigma) = \int_{\mathbb{R}^d} L_\sigma\psi\, d\sigma = 0 = \int_{\mathbb{R}^d} L_\sigma\psi\, d\mu = C_1(\sigma)\int_{\mathbb{R}^d} h\, d\mu + C_2(\sigma).$$

Since $C_1(\sigma) \neq 0$, we obtain $\mu(h) = \sigma(h)$. □

Example 2.7. (i) Let $d = 1$, $A = 1$ and

$$b(x,\mu) = f(x) - \int_{\mathbb{R}^d} f(y)\, \mu(dy),$$

where f is a reasonable function. Then the function $\psi(x) = x$ belongs to I_0^W and for every σ one has

$$L_\sigma x = f(x) - \int_{\mathbb{R}^d} f(y)\, \sigma(dy) = f(x) + C_2(\sigma).$$

Therefore, for the stationary solution μ of the equation with the operator L_σ the equality $\mu(f) = \sigma(f)$ holds.

(ii) Let $d \geq 1$, $A = I$,

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

and

$$K(x,y) = -Rx + \langle v, y\rangle h + H(x,y),$$

where R is a constant matrix, v and h are constant vectors and

$$R^* v = \lambda v, \quad \lambda \neq 0, \quad \langle v, h\rangle = \lambda, \quad \langle H(x,y), v\rangle = 0.$$

Then the function $x \to \langle v, x\rangle$ belongs to I_0^W and

$$L_\sigma \langle v, x\rangle = -\lambda \langle v, x\rangle + \lambda \int_{\mathbb{R}^d} \langle v, x\rangle \, \sigma(dy).$$

Therefore, for the stationary solution μ to the equation with the operator L_σ we have that $\mu(h) = \sigma(h)$ with $h(x) = \langle v, x\rangle$. In particular, if $d = 2$, $x = (x_1, x_2)$, $y = (y_1, y_2)$ and

$$K_1(x,y) = -2x_1 + (y_1 + y_2) + H(x,y), \quad K_2(x,y) = -2x_2 + (y_1 + y_2) - H(x,y),$$

then $\mu(x_1 + x_2) = \sigma(x_1 + x_2)$. Here $R = 2I$, $v = (1,1)$ and $h = (1,1)$.

Proposition 2.6 differs from Propositions 2.1 and 2.4 in which we studied the dynamics of certain quantities along trajectories of solutions. Proposition 2.6 will play the key role in constructing stationary solutions to nonlinear Fokker–Planck–Kolmogorov equations in Section 4 (see Proposition 4.1). Note that the observations above are rather rough and in special situations more refined considerations are possible (see, for example, [26]), but it seems reasonable to begin the study of convergence of solutions of the Cauchy problem to the solution of the stationary equation from finding quantities a priori invariant or subinvariant along trajectories of solutions.

Closing this section we formulate our conditions on the coefficients in terms of the sets I_0^W and I_+^W. It is reasonable (with regards towards convergence) to consider only measures $\mu \in \mathcal{P}_V(\mathbb{R}^d)$ such that $\mu(\psi) = 0$ for every function $\psi \in I_+^W$. We observe that if ν equals zero on all functions from I_+^W (i.e., assigns zero integrals to such functions), then the same is true for the solution μ_t to the Cauchy problem.

Set

$$\mathcal{M}_\alpha(V) = \Bigl\{\mu \in \mathcal{P}_V(\mathbb{R}^d) \colon \int_{\mathbb{R}^d} V \, d\mu \le \alpha\Bigr\}.$$

Recall that $W = V^\gamma$, where $\gamma \in (0, 1/2]$, and $\|\mu\|_W = \|W\mu\|_{TV}$.

Suppose that for every $\varepsilon \in [0,1)$ we are given a mapping

$$b_\varepsilon(\,\cdot\,,\,\cdot\,) \colon \mathbb{R}^d \times \mathcal{P}_V(\mathbb{R}^d) \to \mathbb{R}^d$$

such that for every $\mu \in \mathcal{P}_V(\mathbb{R}^d)$ the mapping $x \mapsto b_\varepsilon(x,\mu)$ is Borel. Let

$$L_{\mu,\varepsilon} u(x) = \operatorname{trace}\,(A(x) D_x^2 u(x)) + \langle b_\varepsilon(x,\mu), \nabla u(x)\rangle.$$

Suppose that for every measure $\nu \in \mathcal{P}_V(\mathbb{R}^d)$ with $\nu|_{I_+^W} = 0$ there exist numbers $C > 0$, $\Lambda > 0$ and $\delta \in [0,1]$ and a positive function N_1 on $[0, +\infty)$ (thus for different ν these objects can be different) such that

($\mathbf{H_1}$) for all $\varepsilon \in [0,1)$, $\alpha \ge 1$ and $\mu \in \mathcal{M}_\alpha(V)$ satisfying the conditions $\mu|_{I_+^W} = 0$ and $\mu|_{I_0^W} = \nu|_{I_0^W}$, we have

$$L_{\mu,\varepsilon} V(x) \le (1-\delta)C + \Lambda(\delta\alpha - V(x)) \quad \forall x \in \mathbb{R}^d,$$

($\mathbf{H_2}$) for all $\varepsilon \in [0,1)$, α and $\mu \in \mathcal{M}_\alpha(V)$ satisfying the conditions $\mu|_{I_+^W} = 0$ and $\mu|_{I_0^W} = \nu|_{I_0^W}$, we have

$$|b_\varepsilon(x,\mu)| \le N_1(\alpha) V(x)^{\frac{1}{2}-\gamma} \quad \forall x \in \mathbb{R}^d.$$

Suppose that there exists a positive function N_2 on $[0, +\infty)$ such that

($\mathbf{H_3}$) for all $\varepsilon \in [0,1)$, $\alpha \ge 0$ and $\mu, \sigma \in \mathcal{M}_\alpha(V)$ satisfying the conditions $\mu|_{I_+^W} = \sigma|_{I_+^W} = 0$ and $\mu|_{I_0^W} = \sigma|_{I_0^W}$, we have

$$|b_\varepsilon(x,\mu) - b_\varepsilon(x,\sigma)| \le \varepsilon N_2(\alpha) V(x)^{\frac{1}{2}-\gamma} \|\mu - \sigma\|_W \quad \forall x \in \mathbb{R}^d.$$

Note that if

$$b_\varepsilon(x,\mu) = \int_{\mathbb{R}^d} K_\varepsilon(x,y)\,\mu(dy) + \widetilde{b_\varepsilon}(x,\mu)$$

and for every x the function $y \mapsto K_\varepsilon(x,y)$ belongs to I_0^W, then condition (H$_3$) refers only to $\widetilde{b_\varepsilon}$, since the difference of the integrals of $K_\varepsilon(x,y)$ with respect to two measures μ and σ with $\mu|_{I_0^W} = \sigma|_{I_0^W}$ is zero.

For example, this is the case where $d=1$, $A=I$ and

$$b_\varepsilon(x,\mu) = -x + \int_{\mathbb{R}} y\,\mu(dy).$$

Here b_ε does not depend on ε. The function $x \mapsto x$ belongs to I_0^W and for every measure μ satisfying the equality

$$\int_{\mathbb{R}} x\,\mu(dx) = \int_{\mathbb{R}} x\,\nu(dx) = Q$$

we have

$$L_{\mu,\varepsilon}(1+|x|^2) \le 3 + Q^2 - (1+|x|^2), \quad |b(x,\mu)| \le |Q| + |x| \le (1+|Q|)(1+|x|^2)^{1/2},$$

i.e., conditions (H$_1$), (H$_2$) and (H$_3$) are fulfilled with

$$C = 2+Q^2, \quad \delta = 0, \quad \Lambda = 1, \quad \gamma = 1/2, \quad N_1 = 1+|Q|, \quad N_2 = 0.$$

Note that (H$_3$) obviously holds at $\varepsilon = 0$ if b_0 does not depend on μ, i.e., the equation at $\varepsilon = 0$ becomes linear, but the previous example shows that this condition can hold also in case of a nontrivial dependence on μ.

The main result of this paper (presented in the next section) states that, for all sufficiently small ε, the listed conditions ensure the exponential convergence to the stationary distribution.

3. Convergence to stationary solutions

Suppose that for every $\nu \in \mathcal{P}_V(\mathbb{R}^d)$ there is a solution $\{\mu_t\}$ to the problem (1.1) on $[0,+\infty)$. Sufficient conditions for the existence of solutions to parabolic and stationary Fokker–Planck–Kolmogorov equations are discussed in the last section. It is immediate that $\mu_t|_{I_0^W} = \nu|_{I_0^W}$ according to Proposition 2.1. Moreover, if $\nu|_{I_+^W} = 0$, then $\mu_t|_{I_+^W} = 0$ according to Proposition 2.4. We shall now use conditions (H$_1$) – (H$_3$) introduced at the end of the previous section.

Theorem 3.1. *Suppose that conditions* (H$_1$), (H$_2$) *and* (H$_3$) *are fulfilled. Let* $\nu \in \mathcal{P}_V(\mathbb{R}^d)$, $\nu|_{I_+^W} = 0$ *and* $\alpha > 0$. *Then there exist positive numbers* ε_0, α_1 *and* α_2 (*depending on* ν *and* α) *such that, whenever* $\varepsilon \in [0,\varepsilon_0)$, *for the solution* μ_t *to the Cauchy problem* (1.1) *with coefficients* A *and* b_ε *and initial data* ν *and the stationary solution* μ *to equation* (1.3) *with coefficients* A *and* b_ε *such that*

$$\mu|_{I_0^W} = \nu|_{I_0^W} \quad \textit{and} \quad \int_{\mathbb{R}^d} V\,d\mu \le \alpha,$$

we have

$$\|\mu_t - \mu\|_W \le \alpha_1 e^{-\alpha_2 t} \quad \forall\, t \ge 0.$$

Example 3.2. Let $d \ge 1$, $A = I$,

$$b_\varepsilon(x,\mu) = -Rx + \int_{\mathbb{R}^d} \langle v,y\rangle\,\mu(dy) h + \varepsilon \int_{\mathbb{R}^d} H(x,y)\,\mu(dy),$$

where R is a constant matrix, v and h are constant vectors and

$$R^* v = \lambda v, \quad \langle v,h\rangle = \lambda, \quad \langle H(x,y), v\rangle = 0.$$

Suppose also that

$$\langle Rx,x\rangle \ge q|x|^2, \quad q>0, \quad \sup_{x,y} |H(x,y)| < \infty.$$

For example, for $d=2$ one can take $R = 2I$, $v = (1,1)$ and $h = (1,1)$:

$$b_\varepsilon^1(x,\mu) = -2x_1 + \int_{\mathbb{R}^2} (y_1+y_2)\,\mu(dy) + \varepsilon \int_{\mathbb{R}^2} H(x,y)\,\mu(dy),$$

$$b_\varepsilon^2(x,\mu) = -2x_2 + \int_{\mathbb{R}^2} (y_1+y_2)\,\mu(dy) - \varepsilon \int_{\mathbb{R}^2} H(x,y)\,\mu(dy).$$

Let us show that all conditions of Theorem 3.1 are fulfilled. The function $x \mapsto \langle v, x\rangle$ belongs to I_0^W. Let ν be a probability measure with $|x|^2 \in L^1(\nu)$. Set

$$Q = \int_{\mathbb{R}^2} \langle v, y\rangle\,\nu(dy).$$

For all measures σ that coincide with ν on I_0^W we have

$$L_{\sigma,\varepsilon}(1+|x|^2) \le 2d + q + q^{-1}(|h||Q| + \sup_{x,y}|H(x,y)|)^2 - q(1+|x|^2),$$

$$|b_\varepsilon(x,\mu)| \le (\|R\| + |h||Q| + \sup_{x,y}|H(x,y)|)(1+|x|^2)^{1/2}.$$

Finally, for every two measures μ and σ that coincide on I_0^W we have

$$|b_\varepsilon(x,\mu) - b_\varepsilon(x,\sigma)| \le \varepsilon \sup_{x,y}|H(x,y)|\|(\mu-\sigma)(1+|x|^2)^{1/2}\|_{TV}.$$

Thus, conditions (H$_1$), (H$_2$) and (H$_3$) are fulfilled with $\gamma = 1/2$, $W(x) = (1+|x|^2)^{1/2}$ $V(x) = 1+|x|^2$, $\delta = 0$, $\Lambda = 1$, $N_2 = \sup_{x,y}|H(x,y)|$ and

$$C = 2d + q + q^{-1}(|h||Q| + \sup_{x,y}|H(x,y)|)^2, \quad N_1 = \|R\| + |h||Q| + \sup_{x,y}|H(x,y)|.$$

Moreover, it will be shown in Section 4 (see Example 4.2) that for every $\varepsilon \in (0,1)$ and every number Q there is a stationary solution μ such that

$$Q = \int_{\mathbb{R}^2} \langle v, y\rangle\,\mu(dy).$$

In addition, for this solution μ we have

$$\int_{\mathbb{R}^2} (1+|x|^2)\,\mu(dx) \le 2dq^{-1} + 1 + q^{-2}(|h||Q| + \sup_{x,y}|H(x,y)|)^2.$$

Thus, for every number $Q_0 > 0$ there is a number $\varepsilon_0 > 0$, depending only on Q_0, such that, for any $\varepsilon \in [0,\varepsilon_0)$, $Q \in (-Q_0, Q_0)$ and a probability measure ν such that the integral of $\langle v, y\rangle$ with respect to ν equals Q and $|x|^2 \in L^1(\nu)$, the solutions μ_t to the Cauchy problem (1.1) with initial data ν converge to the stationary solution μ and

$$\|(1+|x|)(\mu_t - \mu)\|_{TV} \le \alpha_1 e^{-\alpha_2 t},$$

where α_1 and α_2 depend only on Q_0 and $\|(1+|x|)^2\|_{L^1(\nu)}$.

Theorem 3.1 is of a somewhat conditional nature: 1) we assume that there exists a stationary solution μ such that $\nu|_{I_0^W} = \mu|_{I_0^W}$, 2) ε_0 depends on ν and, what is worse, also on μ. Dependence on μ arises in connection with dependence of our conditions on α, i.e., due to nonlinearity (see the discussion in the proof of Lemma 3.5 and before Lemma 3.6). If we require stronger restrictions on the coefficients, then we can avoid such dependence.

Theorem 3.3. *Suppose that in place of conditions* (H$_1$), (H$_2$) *and* (H$_3$) *there exist positive numbers C_1, C_2 and positive functions N_1 and N_2 such that for all $\varepsilon \in [0,1)$, $\alpha > 0$ and $\mu, \sigma \in \mathcal{M}_\alpha(V)$ we have $L_{\mu,\varepsilon}V \le C_1 - C_2 V$ and*

$$|b_\varepsilon(x,\mu)| \le N_1(\alpha)V^{1/2-\gamma}(x), \quad |b_\varepsilon(x,\mu) - b_\varepsilon(x,\sigma)| \le \varepsilon N_2(\alpha)V(x)^{1/2-\gamma}\|\mu-\sigma\|_W.$$

Then there exists $\varepsilon_0 > 0$, such that, for each $\varepsilon \in [0,\varepsilon_0)$ there exists a stationary solution μ and, for every measure $\nu \in \mathcal{P}_V(\mathbb{R}^d)$, for the solution $\{\mu_t\}$ to the Cauchy problem (1.1) *with the initial condition ν one has*

$$\|\mu_t - \mu\|_W \le \alpha_1 e^{-\alpha_2 t} \quad \forall\, t \ge 0,$$

where α_1, α_2 are positive numbers such that α_2 does not depend on ν.

Example 3.4. Suppose that $A = I$ and there exist numbers $m \ge 1$, $\gamma_1 > 0$, $\gamma_2 > 0$ and positive functions N_1, N_2 such that

$$\langle b_\varepsilon(x,\mu), x\rangle \le \gamma_1 - \gamma_2|x|^2, \quad |b_\varepsilon(x,\mu)| \le N_1(\alpha)(1+|x|)^m,$$
$$|b_\varepsilon(x,\mu) - b_\varepsilon(x,\sigma)| \le \varepsilon N_2(\alpha)(1+|x|)^m \|(1+|y|)^m(\mu-\sigma)\|_{TV}$$

for all $\varepsilon \in [0,1)$, $\alpha > 0$ and $\mu, \sigma \in \mathcal{M}_\alpha((1+|x|^{2m+1})$. Hence all conditions of Theorem 3.3 are fulfilled with $V(x) = (1+|x|^2)^{m+1/2})$ and $W(x) = (1+|x|^2)^{m/2}$. In particular, the listed conditions are fulfilled if

$$b_\varepsilon(x,\mu) = b_0(x) + \varepsilon \int_{\mathbb{R}^d} K(x,y)\,\mu(dy),$$

where

$$\langle b_0(x), x\rangle \le c_1 - c_2|x|^2, \quad \langle K(x,y), x\rangle \le c_3 + c_3|x|^2, \quad c_3 < c_2,$$
$$|b_0(x)| \le c_4 + c_4|x|^m, \quad |K(x,y)| \le c_5(1+|x|^m)(1+|y|^m)$$

with some positive numbers c_1, c_2, c_3, c_4 and c_5.

The existence of a stationary solution μ under the conditions of Theorem 3.3 will be established in the next section in Proposition 4.1. As it will be explained in Remark 4.3, under the conditions of Theorem 3.3 there exists a stationary solution μ with

$$\|V\|_{L^1(\mu)} \le C_1/C_2.$$

It is the stationary solution that we need in the proof of Theorem 3.3.

We only give the proof of Theorem 3.1, since the proof of Theorem 3.3 differs by minor technical details that will be discussed in the course of the proof.

Below for shortening notation and reducing the number of indices we omit the index ε and in place of $b_\varepsilon(x,\mu)$ and $L_{\mu,\varepsilon}$ we write $b(x,\mu)$ and L_μ.

The plan of the proof is this: 1) we verify that convergence holds for solutions η_t to the linear equation with the coefficient $b(x,\mu)$, in which we substitute the stationary solution μ, 2) we obtain an estimate on the distance $\|\eta_t - \mu_t\|_W$, 3) we prove that for some $T > 0$ one has a contraction $\|\mu_T - \mu\|_W \le q\|\nu - \mu\|_W$ with $q < 1$.

Lemma 3.5. *Suppose that we are in the situation of Theorem* 3.1 *or* 3.3 *with the corresponding μ and ν. Then there exist numbers $N > 0$ and $\lambda > 0$ such that for every $t > 0$ we have*

$$\|\eta_t - \mu\|_W \le Ne^{-\lambda t}\|\mu_0 - \mu\|_W,$$

where $\mu_0 \in \mathcal{P}_V(\mathbb{R}^d)$ and $\{\eta_t\}$ is the solution to the Cauchy problem

$$\partial_t \eta_t = L_\mu^* \eta_t, \quad \eta_0 = \mu_0.$$

The numbers N and λ depend on $\|V\|_{L^1(\mu)}$ and ν, and if the condition of Theorem 3.3 *is fulfilled, then N and λ depend on C_1 and C_2, but not on μ and ν.*

Proof. Let $\{T_t\}_{t\ge 0}$ be the Markov semigroup on $L^1(\mu)$ with generator

$$L\varphi(x) = \mathrm{trace}(A(x)D^2\varphi(x)) + \langle b(x,\mu), \nabla\varphi(x)\rangle$$

on $C_0^\infty(\mathbb{R}^d)$. This semigroup exists and is unique under our assumptions, see [10, Theorem 5.2.2, Proposition 5.2.5 and Example 5.5.1]. Moreover, $\eta_t = T_t^*\mu_0$. By [10, Theorem 6.4.7] there exists a positive continuous function $\varrho(x,y,t)$ such that

$$T_t f(x) = \int_{\mathbb{R}^d} \varrho(x,y,t) f(y)\,dy.$$

Moreover, $\varrho(x,y,t)$ satisfies the Cauchy problem $\partial_t\varrho = L^*\varrho$ with respect to (t,y) with the initial condition δ_x. Set $\alpha = \|V\|_{L^1(\mu)}$ (or $\alpha = C_1/C_2$ in the case of Theorem 3.3) and $\Lambda_1 = (1-\delta)C + \Lambda\delta\alpha$ (or $\Lambda_1 = C_1$ and $\Lambda = C_2$ in the case of Theorem 3.3). Recall that $LV \le \Lambda_1 - \Lambda V$. Since

$$LW = \gamma V^{\gamma-1}LV + \gamma(\gamma-1)V^{\gamma-2}|\nabla V|^2 \le \gamma\Lambda_1 - \gamma\Lambda W,$$

we have

$$\partial_t(We^{tH_3}) + L(We^{tH_3}) \le \gamma\Lambda_1 e^{t\Lambda}.$$

By [10, Theorem 7.1.1]

$$\int_{\mathbb{R}^d} W(y)\varrho(x,y,t)\,dy \le e^{-\Lambda t}W(x) + \gamma\Lambda_1\Lambda^{-1}(1-e^{-\Lambda t}).$$

Let us fix a number $\tau > 0$ such that $\gamma\Lambda_1\Lambda^{-1}(1-e^{-\Lambda\tau}) < 1$. Note that τ depends on C_1 and C_2 in the case of Theorem 3.3. Then

$$T_\tau W(x) \le e^{-\Lambda\tau}W(x) + 1.$$

The function $Q(r) = \max_{|x|\le 2r} V(x)^{\frac{1}{2}-\gamma}$ is continuous and increasing on $[0,+\infty)$. The condition $|b(x,\mu)| \le N_1(\alpha)V(x)^{\frac{1}{2}-\gamma} \le N_1(\alpha)Q(|x|/2)$ and Harnack's inequality (see [10, Theorem 8.2.1]) imply that, for every $x \in B(0,R)$ and $y \in \mathbb{R}^d$, we have

$$\varrho(x,y,\tau) \ge \varrho(x,0,\tau/2)e^{-K(\tau)(1+Q^2(|y|)+|y|^2)} \ge m_1(R)e^{-K(\tau)(1+Q^2(|y|)+|y|^2)},$$

where $m_1(R) = \min_{x\in B(0,R)} \varrho(x,0,\tau/2)$. The number $K(\tau)$ depends only on the matrix A, τ and the dimension d, and there is an explicit expression for $K(\tau)$ in [10, Theorem 8.2.1]. Note that so far m_1 depends in a very complicated way on the stationary measure μ, since L depends on μ and ϱ defines the operator L. We would like to have dependence only on N_1, Λ_1 and Λ, which in turn depend on $\alpha = \|V\|_{L^1(\mu)}$ and ν, and in the case of Theorem 3.3 depend on C_1 and C_2 and are independent of ν and μ. Thus, we have to estimate m_1 from below. Let $\psi \in C_0^\infty(\mathbb{R}^d)$, $\psi(x) = 1$ if $y \in B(0,2R)$ and $\psi(y) = 0$ if $y \notin B(0,3R)$. Let $x \in B(0,R)$. Then

$$\int_{\mathbb{R}^d} \psi(y)\varrho(x,y,t)\,dy = \psi(x) + \int_0^t \int_{\mathbb{R}^d} L\psi(y)\varrho(x,y,s)\,dy\,ds.$$

Therefore,

$$\int_{\mathbb{R}^d} \psi(y)\varrho(x,y,t)\,dy \ge 1 - t\sup_y |L\psi(y)|.$$

Choosing t so small that the right-hand side is estimated from below by $1/2$, we obtain

$$\sup_{y\in B(0,3R)} \varrho(x,y,t) \ge 1/2 \quad \forall\, x \in B(0,R).$$

Decreasing τ if necessary, we can assume that $t = \tau/4$. Applying again Harnack's inequality from [10, Theorem 8.1.3] we obtain the estimate

$$1/2 \le C\varrho(x,0,\tau/2),$$

where C depends only on R, Q, and τ. Thus,

$$\varrho(x,y,\tau) \ge m(R)e^{-K(\tau)(1+Q^2(|y|)+|y|^2)} \quad \forall\, x \in B(0,R)$$

where $m(R)$ and τ depend only on N_1, Λ_1 and Λ.

Let us now recall the Harris ergodic theorem (see [25]). Let $\mathcal{P}(\,\cdot\,,\,\cdot\,)$ be a Markov transition kernel defined on a measurable space $(X,\mathcal{B})$, i.e., for each $x \in X$, the function $B \mapsto \mathcal{P}(x,B)$ is a probability measure on $\mathcal{B}$, and, for each $B \in \mathcal{B}$, the function $x \mapsto \mathcal{P}(x,B)$ is $\mathcal{B}$-measurable. The transition kernel defines operators on functions and measures by setting

$$\mathcal{P}f(x) = \int_X f(y)\,\mathcal{P}(x,dy),$$

$$\mathcal{P}\sigma(B) = \int_X \mathcal{P}(x,B)\,\sigma(dx).$$

Let us assume that

(i) there exist a function $U\colon X \to [0,+\infty)$ and numbers $\delta \in (0,1)$ and K such that

$$\mathcal{P}U(x) \le \delta U(x) + K \quad \forall\, x \in X;$$

(ii) there exist a number $q \in (0,1)$ and a probability measure σ such that

$$\inf_{x\colon U(x)\le R} \mathcal{P}(x,\,\cdot\,) \ge q\sigma(\,\cdot\,),$$

for some $R > 2K/(1-\delta)$.

According to [25, Theorem 1.3], there exist numbers $\beta_0 \in (0,1)$ and $\beta > 0$ such that

$$\|\mathcal{P}\mu_1 - \mathcal{P}\mu_2\|_{1+\beta U} \le \beta_0 \|\mu_1 - \mu_2\|_{1+\beta U}$$

for every pair of probability measures μ_1 and μ_2 on X. From this estimate one can derive the bound

$$\|\mathcal{P}^n \nu - \mu\|_{1+\beta U} \le \beta_0^n \|\nu - \mu\|_{1+\beta U}$$

for the stationary measure μ (that is, $\mathcal{P}\mu = \mu$) and every measure ν.

We can now apply this assertion to the Markov transition kernel $\varrho(x, y, \tau)\, dy$ with $U(x) = W(x)$, $K = 1$, $\delta = e^{-\Lambda\tau}$ and

$$q\sigma(dy) = m(R)e^{-K(\tau)(1+Q^2(|y|)+|y|^2)}\, dy,$$

where the number R is larger than $2/(1 - e^{-\Lambda\tau})$. Therefore, we have

$$\|T^*_{n\tau}\mu_0 - \mu\|_W \le N_1 \beta_0^n \|\nu - \mu\|_W. \tag{3.1}$$

Note that for every $\varphi \in C_0^\infty(\mathbb{R}^d)$ such that $|\varphi(x)| \le W(x)$ for all x we have

$$|T_t\varphi| \le T_t|\varphi| \le T_tW \le 2W \quad \forall t \in (0, \tau).$$

Hence

$$\int_{\mathbb{R}^d} \varphi\, d(T^*_t\mu_0 - \mu) = \int_{\mathbb{R}^d} T_t\varphi\, d(\mu_0 - \mu) \le 2\|\mu_0 - \mu\|_W$$

and we obtain the estimate

$$\sup_{t\in(0,\tau)} \|T^*_t\mu_0 - \mu\|_W \le 2\|\mu_0 - \mu\|_W.$$

Summing this estimate and (3.1), we complete the proof. □

Suppose that as above μ_t is the solution to the Cauchy problem (1.1) with initial data ν. Before estimating the distance between μ_t and η_t we estimate $\|V\|_{L^1(\mu_t)}$. Now let $\delta < 1$. With the aid of condition (H$_1$) we deduce that

$$\int_{\mathbb{R}^d} V\, d\mu_t \le \int_{\mathbb{R}^d} V\, d\nu + (1 - \delta)Ct - (1 - \delta) \int_0^t \int_{\mathbb{R}^d} V\, d\mu_s\, ds.$$

If $\delta = 1$, then $\|V\|_{L^1(\mu_t)} \le \|V\|_{L^1(\nu)}$. If $\delta < 1$, then by Gronwall's inequality we obtain

$$\int_{\mathbb{R}^d} V\, d\mu_t \le \Bigl(\int_{\mathbb{R}^d} V\, d\nu - \frac{C}{\Lambda}\Bigr) e^{-\Lambda(1-\delta)t} + \frac{C}{\Lambda}.$$

If $\delta < 1$, then, starting from some $\tau_0 > 0$, we can assume that $\|V\|_{L^1(\mu_t)} \le C\Lambda^{-1} + 1$ for all $t \ge \tau_0$. Since we are interested in convergence as $t \to \infty$, we can always consider the Cauchy problem for $t > \tau_0$ and with the initial condition μ_{τ_0} in place of ν. Hence we assume further that for $\delta = 1$ we have $\|V\|_{L^1(\mu_t)} \le \|V\|_{L^1(\nu)}$, and for $\delta < 1$ we have $\|V\|_{L^1(\mu_t)} \le C\Lambda^{-1} + 1$. Let

$$\theta = \max\{\|V\|_{L^1(\nu)}, C\Lambda^{-1} + 1, \|V\|_\mu\},$$

and let $\theta = 1 + C_1/C_2$ in the conditions of Theorem 3.3 . We observe that by the uniqueness of solutions to the Cauchy problem (1.1) under our assumptions about the coefficients (see Proposition 4.4) the solution $\mu_{\tau+t}$ to the Cauchy problem with the initial condition ν coincides for $t \ge 0$ with the solution μ_t to the Cauchy with the initial condition μ_τ.

Lemma 3.6. *Let $\tau \ge \tau_0$. Let $\{\mu_t\}$ be the solution to the Cauchy problem* (1.1) *and let $\{\eta_t\}$ be the solution to the Cauchy problem $\partial_t\eta_t = L^*_\mu\eta_t$, $\eta_0 = \mu_\tau$. Then*

$$\|\mu_{\tau+t} - \eta_t\|_W \le \varepsilon C(\theta)\Bigl(\int_0^t \|\mu_{\tau+t} - \mu\|_W^2\, dt\Bigr)^{1/2}, \quad C(\theta) = \theta\bigl(8K_1^2 N_2(\theta)\theta^2 t + \theta\bigr).$$

Proof. Let $\widetilde{\mu}_t = \mu_{\tau+t}$. Let $\{T_t\}_{t\ge 0}$ be the semigroup with the generator L from the previous proof. Set

$$u(x,t) = T_{s-t}\psi(x),$$

where $\psi \in C_0^\infty(\mathbb{R}^d)$ and $|\psi(x)| \le W(x)$ for all x. Then

$$\int_{\mathbb{R}^d} \psi\, d(\widetilde{\mu}_t - \eta_t) = \int_0^t \int_{\mathbb{R}^d} \langle b(x,\widetilde{\mu}_t) - b(x,\mu), \nabla_x u\rangle\, \widetilde{\mu}_t(dx)\, dt.$$

We need a bound on $|\nabla_x u|$. We have

$$\int_{\mathbb{R}^d} \psi(x)^2\, \widetilde{\mu}_t(dx) - \int_{\mathbb{R}^d} u(x,0)^2\, \mu_\tau(dx)$$
$$= \int_0^t \int_{\mathbb{R}^d} \Bigl[2|\sqrt{A}\nabla u(x)|^2 + 2\langle b(x,\widetilde{\mu}_t) - b(x,\mu), \nabla_x u(x)\rangle u(x)\Bigr]\, \widetilde{\mu}_t(dx)\, dt.$$

Recall that $|u| \le 2W$ and

$$|b(x,\widetilde{\mu}_t) - b(x,\mu)| \le \varepsilon N_2(\theta) V(x)^{\frac{1}{2}-\gamma} \|\widetilde{\mu}_t - \mu\|_W,$$

where $\|\widetilde{\mu}_t - \mu\|_W \le 2\theta^{1/2}$. Since

$$2|\langle b(x,\widetilde{\mu}_t) - b(x,\mu), \nabla_x u(x)\rangle u(x)| \le 4K_1 N_2(\theta)^2 \theta V(x) + 2^{-1}K_1^{-1}|\nabla_x u|^2,$$

we obtain

$$\int_0^t \int_{\mathbb{R}^d} |\nabla u|^2\, d\widetilde{\mu}_t\, dt \le 8K_1^2 N_2(\theta)\theta^2 t + \theta.$$

Let us observe that

$$\int_0^t \int_{\mathbb{R}^d} \langle b(x,\widetilde{\mu}_t) - b(x,\mu), \nabla_x u(x)\rangle\, \widetilde{\mu}_t(dx)\, dt$$
$$\le \varepsilon\theta\Bigl(8K_1^2 N_2(\theta)\theta^2 t + \theta\Bigr)\Bigl(\int_0^t \|\widetilde{\mu}_t - \mu\|_W^2\, dt\Bigr)^{1/2}.$$

We obtain

$$\int_{\mathbb{R}^d} \psi\, d(\widetilde{\mu}_t - \eta_t) \le \varepsilon\theta\Bigl(8K_1^2 N_2(\theta)\theta^2 t + \theta\Bigr)\Bigl(\int_0^t \|\widetilde{\mu}_t - \mu\|_W^2\, dt\Bigr)^{1/2}.$$

Since $|\psi| \le W$, we have

$$\|\widetilde{\mu}_t - \eta_t\|_W \le \varepsilon\theta\Bigl(8K_1^2 N_2(\theta)\theta^2 t + \theta\Bigr)\Bigl(\int_0^t \|\widetilde{\mu}_t - \mu\|_W^2\, dt\Bigr)^{1/2}.$$

which completes the proof. $\square$

We are now ready to prove our main theorems.

Proof of Theorem 3.1. Let $\tau \ge \tau_0$. We recall that $\mu_{\tau+t}$ with $t > 0$ solves the Cauchy problem (1.1) with the initial condition μ_τ. Let η_t be the solution to the linear Cauchy problem $\partial_t \eta_t = L_\mu^* \eta_t$, $\eta_0 = \mu_\tau$. Using Lemma 3.5 and Lemma 3.6 we obtain

$$\|\mu_{\tau+t} - \mu\|_W \le \|\eta_t - \mu\|_W + \|\mu_{\tau+t} - \eta_t\|_W$$
$$\le Ne^{-\lambda t}\|\mu_\tau - \mu\|_W + \varepsilon\theta\Bigl(8K_1^2 N_2(\theta)\theta^2 t + \theta\Bigr)\Bigl(\int_0^t \|\mu_{\tau+t} - \mu\|_W^2\, dt\Bigr)^{1/2}.$$

Let $T > 0$ be such that $Ne^{-\lambda T} < 1/2$. Set

$$M = \theta\Bigl(8K_1^2 N_2(\theta)\theta^2 T + \theta\Bigr).$$

For any $t \in [0,T]$ we have

$$\|\mu_{\tau+t} - \mu\|_W \le N\|\mu_\tau - \mu\|_W + \varepsilon M\Bigl(\int_0^t \|\mu_{\tau+t} - \mu\|_W^2\, dt\Bigr)^{1/2}.$$

By Gronwall's inequality
$$\|\mu_{\tau+t}-\mu\|_W \le 2Ne^{2\varepsilon^2M^2t}\|\mu_\tau-\mu\|_W \le 2Ne^{2\varepsilon^2M^2T}\|\mu_\tau-\mu\|_W. \tag{3.2}$$
Therefore,
$$\|\mu_{\tau+T}-\mu\|_W \le \Bigl(\frac{1}{2}+2N\varepsilon MT^{1/2}e^{2\varepsilon^2M^2T^2}\Bigr)\|\mu_\tau-\mu\|_W.$$
Let us take $\varepsilon_0>0$ such that
$$q=\frac{1}{2}+2N\varepsilon MT^{1/2}e^{2\varepsilon^2M^2T^2}<1 \quad \forall\varepsilon\in(0,\varepsilon_0).$$
Then
$$\|\mu_{\tau+T}-\mu\|_W \le q\|\mu_\tau-\mu\|_W \quad \forall\tau\ge\tau_0.$$
We have $\|\mu_{\tau_0+nT}-\mu\|_W \le \theta^n\|\mu_{\tau_0}-\mu\|_W$. Using this estimate and (3.2), we obtain
$$\|\mu_t-\mu\|_W \le \alpha_1e^{-\alpha_2t} \quad \forall t\ge 0,$$
which completes the proof. □

It is seen from the proof above that ε_0 depends on the quantities N and λ from Lemma 3.5 and on the number θ defined before Lemma 3.6. Thus, under the conditions of Theorem 3.3 the number ε_0 depends only on C_1, C_2 and the functions N_1 and N_2.

Remark 3.7. Let us discuss in more detail Example 1.1 from the introduction. Let $d=1$, $A=1$ and
$$b_\varepsilon(x,\mu)=-x+\varepsilon B(\mu), \quad B(\mu)=\int x\,\mu(dx), \quad \varepsilon\ge 0.$$
In the case where $0\le\varepsilon<1$, the standard Gaussian measure μ is the unique probability solution to the stationary equation. Let us show that for every initial condition ν with a finite first moment the measures μ_t from the solution to the Cauchy problem converge to μ. The justification repeats the main steps of the proof of Theorem 3.1. We observe that
$$\frac{d}{dt}B(\mu_t)=-(1-\varepsilon)B(\mu_t), \quad B(\mu_{t+s})=e^{-(1-\varepsilon)t}B(\mu_s).$$
Moreover,
$$\int |x|^2\,d\mu_t \le \int |x|^2\,\nu(dx)+(1-\varepsilon)^{-1}\int |x|\,\nu(dx).$$
Let $\tau>0$ and let $\{\eta_t\}$ be the solution to the Cauchy problem
$$\partial_t\eta_t=\eta_t''+(x\eta_t)', \quad \eta_0=\mu_\tau.$$
By Lemma 3.5 with $W(x)=(1+|x|^2)^{1/2}$ and $V(x)=1+|x|^2$ we have
$$\|\eta_t-\mu\|_W \le C_1e^{-C_2\tau}\|\mu_\tau-\mu\|_W \le C_3e^{-C_2\tau}.$$
Since
$$b(x,\mu_{t+\tau})-b(x,\mu)=\varepsilon B(\mu_{t+\tau})=\varepsilon e^{-(1-\varepsilon)t}B(\mu_\tau),$$
repeating the reasoning from Lemma 3.6 one can readily obtain the bound
$$\|\mu_{t+\tau}-\eta_t\|_W \le C(\varepsilon)B(\mu_\tau) \le C_4e^{-(1-\varepsilon)\tau}, \quad t\in[0,\tau].$$
Combining the obtained estimates for $t=\tau$, we conclude that
$$\|\mu_{2\tau}-\mu\|_W \le \|\mu_{2\tau}-\eta_\tau\|_W+\|\eta_\tau-\mu\|_W \le C_4e^{-(1-\varepsilon)\tau}+C_3e^{-C_2\tau}.$$
Thus,
$$\|\mu_t-\mu\|_W \le \alpha_1e^{-\alpha_2t}.$$
Note that in this case the specific form of $b_\varepsilon(x,\mu)$ has enabled us to use the exponential convergence of $B(\mu_t)$ to zero in place of Condition (H3) (Lipschitzness in μ).

The case where $\varepsilon=1$ has been considered in the introduction. In addition, it is covered by Theorem 3.1.

We now consider the case where $\varepsilon>1$. Then the unique stationary solution is the standard Gaussian measure and the solutions to the Cauchy problem converge in total variation norm to this measure only if the initial condition ν has zero mean. In case of a nonzero mean of ν it is easy to see that there is no convergence in total variation norm with weight $(1+|x|)$, because in this case the means of μ_t must converge to the mean of the stationary distribution, while in our example with $\varepsilon>1$ the mean of μ_t equals $B(\nu)e^{(\varepsilon-1)t}$ and tends to infinity. However, one still might hope that there is convergence in some weaker sense, for example, weak convergence. However, weak convergence also fails in case of a nonzero mean of ν. Indeed, let $\varphi\in C_0^\infty(\mathbb{R})$. Then

$$\int \varphi\, d\mu_t - \int \varphi\, d\nu = \int_0^t \Bigl[\int (\varphi'' - x\varphi')\, d\mu_s + B(\nu)e^{(\varepsilon-1)s}\int \varphi'\, d\mu_s\Bigr]\, ds.$$

If μ_t converges to the stationary distribution μ in the sense of weak convergence, then all integrals with φ'', $x\varphi'$, φ' and φ converge to some constants. Let

$$\int \varphi'\, d\mu \neq 0.$$

Then the right-hand side of the integral equality above is unbounded as $t\to\infty$, which contradicts the boundedness of the left-hand side.

Remark 3.8. Let us explain why convergence in the Kantorovich distance can be more easily verified. We consider the following example (see also [9, Remark 4.2]). Let $A=I$, $\sup_{x,\mu}|b(x,\mu)|(1+|x|)^{-m}<\infty$ and suppose that there exist numbers $\kappa>0$ and $C>0$ such that

$$|b(x,\mu)-b(x,\sigma)|\le CW_1(\mu,\sigma), \quad \langle x-y, b(x,\mu)-b(y,\mu)\rangle \le -\kappa|x-y|^2,$$

where $W_1(\mu,\sigma)$ is the Kantorovich metric defined as the supremum of the quantities

$$\int_{\mathbb{R}^d}\varphi\, d(\mu-\sigma)$$

over all 1-Lipschitz functions φ. Suppose that $C<\kappa$. Then it is not difficult to show that the solution $\{\mu_t\}$ to the Cauchy problem (1.1) with the initial condition ν converges to the stationary solution μ (the existence of which follows by [9, Theorem 4.1]), moreover,

$$W_1(\mu_t,\mu)\le e^{-(\kappa-C)t}W_1(\nu,\mu).$$

Let us note the remarkable sharpness of this result: in Example 1.1 we have $\kappa=1$, $C=\varepsilon$, and already for $C=\kappa$ ($\varepsilon=1$) the assertion about convergence fails. A justification of this result is not difficult. Let $\{T_t\}$ be the semigroup on $L^1(\mu)$ generated by the operator $Lf=\Delta f+\langle b(x,\mu),\nabla f\rangle$ (see [10, Chapter 5]). The measure μ is invariant with respect to T_t. For every function $f\in C_0^\infty(\mathbb{R}^d)$ with $|\nabla f|\le 1$ we have $|\nabla T_t f|\le e^{-\kappa t}$ (see [10, Theorem 5.6.41]). Multiplying the Fokker–Planck–Kolmogorov equation by the function $u(x,t)=T_{\tau-t}f(x)$ and integrating by parts (which is possible by the stated properties of u) we derive that

$$\int_{\mathbb{R}^d} f\, d(\mu_\tau-\mu) = \int_{\mathbb{R}^d} u(x,0)\, d(\nu-\mu) + \int_0^\tau\int_{\mathbb{R}^d}\langle b(x,\mu_t)-b(x,\mu),\nabla u\rangle\, d\mu_t\, dt,$$

which yields the estimate

$$W_1(\mu_t,\mu)\le W_1(\nu,\mu)e^{-\kappa\tau} + C\int_0^\tau W_1(\mu_t,\mu)e^{-\kappa(\tau-t)}\, dt.$$

It remains to apply Gronwall's inequality.

4. Solvability of nonlinear Fokker–Planck–Kolmogorov equations

In this section we discuss conditions under which the stationary equation and the Cauchy problem for the Fokker–Planck–Kolmogorov equation have solutions.

Let $\mathcal{H}$ be the set of all functions $h\in C(\mathbb{R}^d)$ such that $\sup_x|h(x)|/W(x)<\infty$ and

$$L_\sigma\psi(x)=C_1(\sigma)h(x)+C_2(\sigma)$$

for some function $\psi\in I_0^W$ and all $\sigma\in\mathcal{P}_V(\mathbb{R}^d)$, where $C_1(\sigma)$ and $C_2(\sigma)$ are numbers depending on σ. According to Proposition 2.4, if μ is a solution to the linear stationary Fokker–Planck–Kolmogorov equation with the drift coefficient $b(x,\sigma)$, then $\mu(h)=\sigma(h)$ for every function $h\in\mathcal{H}$.

Proposition 4.1. *Suppose that for every measure $\nu \in \mathcal{P}_V(\mathbb{R}^d)$ and every ball U there exist numbers $\delta \in (0,1)$, $C > 0$, $\Lambda > 0$ and for every ball U there exists a positive function $\alpha \mapsto N(\alpha, U)$, which depends also on ν, such that for all $\alpha > 0$ and $\mu \in \mathcal{M}_\alpha(V)$ satisfying the condition $\mu|_{\mathcal{H}} = \nu|_{\mathcal{H}}$ we have*

$$L_\mu V \le (1-\delta)C + \Lambda(\delta\alpha - V), \quad \sup_{x\in U} |b(x,\mu)| \le N(\alpha, U).$$

Suppose also that for all $\alpha > 0$ and $\mu_n, \mu \in \mathcal{M}_\alpha(V)$ convergence $\|\mu_n - \mu\|_W \to 0$ yields that $b(x,\mu_n)$ converges to $b(x,\mu)$ uniformly on every ball. Then for every measure $\nu \in \mathcal{P}_V(\mathbb{R}^d)$ there exists a solution μ to the stationary equation (1.3) such that $\nu|_{\mathcal{H}} = \mu|_{\mathcal{H}}$.

Proof. Let $\sigma \in \mathcal{P}_V(\mathbb{R}^d)$ and $\sigma|_{\mathcal{H}} = \nu|_{\mathcal{H}}$. It is well-known (see [10, Corollary 2.4.2 and Theorem 4.1.6]) that there is a unique probability solution μ to the linear equation

$$L^*_\sigma \mu = 0.$$

According to [10, Theorem 2.3.2] we have

$$\int_{\mathbb{R}^d} V\,d\mu \le (1-\delta)C\Lambda^{-1} + \delta \int_{\mathbb{R}^d} V\,d\sigma.$$

Set

$$\alpha = \max\{C\Lambda^{-1}, \|V\|_{L^1(\sigma)}\}.$$

Thus we have $\|V\|_{L^1(\mu)} \le \alpha$ and $\mu|_{\mathcal{H}} = \sigma|_{\mathcal{H}} = \nu|_{\mathcal{H}}$. Since $\sup_U |b(x,\mu)| \le N(\alpha, U)$, the measure μ has a density ϱ and, for every ball U and for some $\delta \in (0,1)$, one has

$$\|\varrho\|_{C^\delta(U)} \le C(U),$$

where $C(U)$ depends only on U, d, $N(\alpha, U)$, and A (see [10, Corollary 1.6.7]). Set

$$\mathcal{K} = \Bigl\{\mu \in \mathcal{M}_\alpha(V)\colon \mu|_{\mathcal{H}} = \nu|_{\mathcal{H}}, \quad \mu = \varrho\,dx, \quad \|\varrho\|_{C^\delta(U)} \le C(U)\Bigr\},$$

where $C^\delta(U)$ is the space of δ-Hölder functions with its natural Hölder norm

$$\|g\|_{C^\delta(U)} = \sup_x |g(x)| + \sup_{x\ne y} |g(x) - g(y)|/|x-y|^\delta.$$

The set $\mathcal{K}$ is convex and compact in $L^1(\mathbb{R}^d)$ and $T\colon \mathcal{K} \to \mathcal{K}$. Compactness follows from the fact that, for any sequence $\{\varrho_n\} \in \mathcal{K}$, the measures $\varrho_n\,dx$ are uniformly tight, hence there is a weakly convergent subsequence. The uniform local Hölder continuity enables us to select a further subsequence in $\{\varrho_n\}$ that converges uniformly on balls. Along with weak convergence this yields convergence in variation.

Next, if $\varrho_n \in \mathcal{K}$ and $\varrho_n \to \varrho$ in $L^1(\mathbb{R}^d)$, then

$$\|\varrho_n - \varrho\|_W \le \max_{|x|\le R} W(x)\|\varrho_n - \varrho\|_{L^1(\mathbb{R}^d)} + 2\alpha(\min_{|x|\ge R} W(x))^{-1}.$$

It follows that $\|\varrho_n - \varrho\|_W \to 0$ and $b(x, \varrho_n\,dx) \to b(x, \varrho\,dx)$ uniformly on every ball. For every $h \in \mathcal{H}$ we obtain $\displaystyle\int h\varrho_n\,dx \to \int h\varrho\,dx$. Moreover, $\sigma_n = T(\varrho_n)$ has a subsequence that converges uniformly on every ball and in $L^1(\mathbb{R}^d)$. Therefore, $\sigma_n \to \sigma$, where σ is a unique probability solution to the equation

$$L^*_{\varrho\,dx}\sigma = 0.$$

Consequently, T is a continuous mapping. By Schauder's fixed point theorem there exists $\mu \in \mathcal{K}$ such that $T(\mu) = \mu$. ☐

Example 4.2. Let $d \ge 1$, $A = I$,

$$b(x,\mu) = -Rx + \biggl(\int_{\mathbb{R}^d} \langle v, y\rangle\,\mu(dy)\biggr)h + \int_{\mathbb{R}^d} H(x,y)\,\mu(dy),$$

where R is a constant matrix, v and h are constant vectors and

$$R^*v = \lambda v, \quad \langle v, h\rangle = \lambda, \quad \langle H(x,y), v\rangle = 0.$$

Suppose also that

$$\langle Rx, x\rangle \geq q|x|^2, \quad q>0, \quad \sup_{x,y}|H(x,y)|<\infty.$$

We show that all conditions in Proposition 4.1 are fulfilled with $V(x)=(1+|x|)^2$ and $W(x)=(1+|x|)$. The function $x\to\langle v,x\rangle$ belongs to I_0^W. In addition, according to Example 2.7 the function $x\to\langle v,x\rangle$ belongs to $\mathcal{H}$. Let μ be a probability measure with

$$\int_{\mathbb{R}^d}\langle v,y\rangle\,\mu(dy)=Q.$$

Then

$$L_\mu(1+|x|^2)\leq 2d+q+q^{-1}(|h||Q|+\sup_{x,y}|H(x,y)|)^2-q(1+|x|^2),$$
$$|b(x,\mu)|\leq(\|R\|+|h||Q|+\sup_{x,y}|H(x,y)|)(1+|x|^2)^{1/2}.$$

Thus, the conditions of the theorem are fulfilled and for every number Q there exists a solution to the stationary equation with such coefficients.

Remark 4.3. Suppose that in place of conditions of Proposition 4.1 the following stronger conditions hold: there are positive numbers C_1, C_2 such that, for every ball U, there is a positive function $\alpha\mapsto N(\alpha,U)$ such that for all $\alpha>0$ and all $\mu\in\mathcal{M}_\alpha(V)$ we have

$$L_\mu V\leq C_1-C_2V, \quad \sup_{x\in U}|b(x,\mu)|\leq N(\alpha,U).$$

Suppose also that for all $\alpha>0$ and $\mu_n,\mu\in\mathcal{M}_\alpha(V)$ convergence $\|\mu_n-\mu\|_W\to 0$ yields that $b(x,\mu_n)$ converges to $b(x,\mu)$ uniformly on every ball. Then there exists a stationary solution μ such that

$$\int_{\mathbb{R}^d}V\,d\mu\leq\frac{C_1}{C_2}.$$

The proof repeats the reasoning given above. We only observe that for every $\sigma\in\mathcal{P}_V(\mathbb{R}^d)$ the solution μ to the equation $L_\sigma^*\mu=0$ satisfies the inequality $\|V\|_{L^1(\mu)}<C_1/C_2$ (see [10, Theorem 2.3.2]) and the mapping T from the proof of Proposition 4.1 maps $\mathcal{M}_\alpha(V)$ with $\alpha=C_1/C_2$ into the same set.

We now discuss the Cauchy problem (1.1). The next proposition gives conditions that guarantee the existence of a solution μ_t on $[0,+\infty)$ such that for every $T>0$ one has

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

Proposition 4.4. *Suppose that there exist continuous positive functions N_1, N_2, N_3 such that for all $\alpha>0$ and $\mu,\sigma\in\mathcal{M}_\alpha(V)$ we have*

$$L_\mu V(x)\leq N_1(\alpha), \quad |b(x,\mu)|\leq N_2(\alpha)V^{\frac{1}{2}-\gamma}, \quad |b(x,\mu)-b(x,\sigma)|\leq N_3(\alpha)V(x)^{\frac{1}{2}}\|\mu-\sigma\|_W.$$

If

$$\int_0^{+\infty}\frac{d\alpha}{N_1(\alpha)}=+\infty, \tag{4.1}$$

then for every initial condition $\nu\in\mathcal{P}_V(\mathbb{R}^d)$ there exists a solution $\{\mu_t\}$ to the Cauchy problem (1.1) *on $[0,+\infty)$ such that for every $T>0$*

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

Proof. We first prove the existence of a solution on $[0,T]$ for every fixed $T>0$. Let $\alpha(t)$ be a positive continuous function on $[0,T]$. If $\sigma_t\in\mathcal{M}_{\alpha(t)}(V)$, then

$$|b(x,\sigma_t)|\leq MV(x)^{1/2-\gamma}, \quad M=\max_{t\in[0,T]}N_2(\alpha(t)).$$

Since $L_{\sigma_t}V(x)\leq N_1(\alpha(t))\leq\max_{t\in[0,T]}N_1(\alpha(t))$, by the standard existence condition involving a Lyapunov function (see [10]) there exists a unique solution $\{\mu_t\}$ to the Cauchy problem

$$\partial_t\mu_t=L_{\sigma_t}^*\mu_t, \quad \mu_0=\nu.$$

Moreover, $\mu_t(dx) = \varrho(x,t)\,dx$ and for some $\delta \in (0,1)$ we have

$$\|\varrho\|_{C^\delta(U\times J)} \le C(U,J)$$

for every ball $U \subset \mathbb{R}^d$ and every interval $J \subset (0,T)$. Note also that by [10, Theorem 7.1.1]

$$\int_{\mathbb{R}^d} V(x)\,\mu_t(dx) \le \int_0^t N_1(\alpha(s))\,ds + \int_{\mathbb{R}^d} V(x)\,\nu(dx).$$

Let us define $\alpha(t)$ by means of the following expression:

$$\int_{\alpha_0}^{\alpha(t)} \frac{du}{N_1(u)} = t, \quad \alpha_0 = \int_{\mathbb{R}^d} V(x)\,\nu(dx).$$

By (4.1) the function α is defined on $[0,+\infty)$. If we take $\sigma_t \in \mathcal{M}_{\alpha(t)}(V)$, then the corresponding solution μ_t will belong to $\mathcal{M}_{\alpha(t)}(V)$. Let $\mathcal{K}_1$ be the set of all functions ϱ on $\mathbb{R}^d \times [0,T]$ such that $\varrho \in C(\mathbb{R}^d \times (0,T))$, $\varrho \ge 0$ and

$$\int_{\mathbb{R}^d} \varrho(x,t)\,dx = 1, \quad \int_{\mathbb{R}^d} \varrho(x,t)V(x)\,dx \le \alpha(t), \quad \|\varrho\|_{C^\delta(U\times J)} \le C(U,J).$$

Note that $\mathcal{K}_1$ is convex and compact in $L^1(\mathbb{R}^d \times [0,T])$. Let us define $T\colon \mathcal{K}_1 \to \mathcal{K}_1$ as follows: to each $\sigma_t = v(x,t)\,dx$ with $v \in \mathcal{K}_1$, the mapping T associates the solution $\mu_t = \varrho(x,t)\,dx$. If $v_n, v \in \mathcal{K}_1$ and $v_n \to v$ in $L^1(\mathbb{R}^d \times [0,T])$, then $\|v_n(y,t)dy - v(y,t)dy\|_W \to 0$. It follows that $b(x, v_n(y,t)\,dy) \to b(x, v(y,t)\,dy)$ for all (x,t). Thus, the corresponding solutions ϱ_n converge to the solution ϱ, hence the mapping T is continuous. By Schauder's fixed point theorem there exists $\varrho \in \mathcal{K}_1$ such that $T(\varrho) = \varrho$. By [30, Theorem 3.1] the constructed solution to the Cauchy problem (1.1) is unique on $[0,T]$ for every $T > 0$. This yields the existence and uniqueness on the whole half-line $[0,+\infty)$. $\square$

References

[1] N.U. Ahmed and X. Ding, On invariant measures of nonlinear Markov processes, J. Appl. Math. Stoch. Anal., 6 (1993), N 4, 385–406.

[2] L. Ambrosio, N. Gigli and G. Savaré, Gradient Flows in Metric Spaces and in the Space of Probability Measures, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel (2005)

[3] A. Arnold, E. Carlen, Q. Ju, Large-time behavior of non-symmetric Fokker–Planck type equations, Comm. Stoch. Anal., 2 (2008), N 1, 153–175.

[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] D. Bakry, P. Cattiaux and A. Guillin, Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré, J. Funct. Anal., 254 (2008), 727–759.

[6] S. Benachour, B. Roynette, D. Talay and P. Vallois, Nonlinear self-stabilizing processes. I: Existence, invariant probability, propagation of chaos, Stochastic Process. Appl., 75 (1998), N 2, 173–201.

[7] S. Benachour, B. Roynette and P. Vallois, Nonlinear self-stabilizing processes. II: Convergence to invariant probability, Stochastic Process. Appl., 75 (1998), N 2, 203–224.

[8] V.I. Bogachev, A.I. Kirillov and S.V. Shaposhnikov, The Kantorovich and variation distances between invariant measures of diffusions and nonlinear stationary Fokker–Planck–Kolmogorov equations, Math. Notes, 96 (2014), N 6, 855–863.

[9] V.I. Bogachev, A.I. Kirillov and S.V. Shaposhnikov, Distances between stationary distributions of diffusions and solvability of nonlinear Fokker–Planck–Kolmogorov equations, Teor. Veroyatn. Primen., 62 (2017), N 1, 16–43 (in Russian); English transl.: Theory Probab. Appl. 62 (2018), N 1.

[10] V.I. Bogachev, N.V. Krylov, M. Röckner and S.V. Shaposhnikov, Fokker–Planck–Kolmogorov Equations, Amer. Math. Soc., Providence, Rhode Island, 2015.

[11] V.I. Bogachev, M. Röckner and S.V. Shaposhnikov, Distances between transition probabilities of diffusions and applications to nonlinear Fokker–Planck–Kolmogorov equations, J. Funct. Anal., 271 (2016), 1262–1300.

[12] F. Bolley and I. Gentil, Phi-entropy inequalities for diffusion semigroups, J. Math. Pures Appl. (9), 93 (2010), 449–473.

[13] F. Bolley, I. Gentil and A. Guillin, Convergence to equilibrium in Wasserstein distance for Fokker–Planck equations, J. Funct. Anal., 263 (2012), 2430–2457.

[14] O.A. Butkovsky, On ergodic properties of nonlinear Markov chains and stochastic McKean–Vlasov equations, Theory Probab. Appl., 58 (2014), 661–674.

[15] J.A. Carrillo, R.J. McCann and C. Villani, Contractions in the 2-Wasserstein length space and thermalization of granular media, Arch. Ration. Mech. Anal., 179 (2006), 217–263.

[16] J.A. Carrillo, M. Difrancesco, A. Figalli, T. Laurent and D. Slepcev, Global-in-time weak measure solutions and finite-time aggregation for non-local interaction equations, Duke Math. J., 156 (2011), 229–271.

[17] P. Cattiaux, Long time behavior of Markov processes, ESAIM: Proceedings, 44 (2014), 110–128.

[18] M.F. Chen and S.F. Li, Coupling methods for multidimensional diffusion processes, Ann. Probab., 17 (1989), 151–177.

[19] D.A. Dawson and J. Gärtner, Large deviations, free energy functional and quasipotential for a mean-field model of interacting diffusions, Mem. Amer. Math. Soc., 78 (1989), no. 398, iv+94 pp.

[20] D. Down, S.P. Meyn and R.L. Tweedie, Exponential and uniform ergodicity of Markov processes, Ann. Probab., 23 (1996), 1671–1691.

[21] A. Eberle, Reflection couplings and contraction rates for diffusions, Probab. Theory Related Fields, 166 (2016), 851–886.

[22] A. Eberle, A. Guillin and R. Zimmer, Quantitative Harris type theorems for diffusions and McKean–Vlasov processes, ArXiv:1606.06012 (2016).

[23] T.D. Frank, Nonlinear Fokker–Planck Equations, Fundamentals and Applications, Springer-Verlag, Berlin, 2005.

[24] T. Funaki, : A certain class of duffusion processes associated with nonlinear parabolic equations, Z. Wahrscheinlichkeitstheorie verw. Geb., 67 (1984), 331–348.

[25] M. Hairer and J.C. Mattingly, Yet another look at Harris' ergodic theorem for Markov chains, In: Seminar on Stochastic Analysis, Random Fields and Applications VI, 2011, Progress in Probab., V. 63, pp. 109–117.

[26] S. Herrmann and J. Tugaut, Non-uniqueness of stationary measures for self-stabilizing processes, Stoch. Processes Appl., 120 (2010), 1215–1246.

[27] M. Kac, Foundations of kinetic theory, In: Proc. 3rd Berkeley Sympos. Math. Statist. Probability, V. 3, 1956, pp. 171–197.

[28] D. Luo and J. Wang, Exponential convergence in L^p-Wasserstein distance for diffusion processes without uniformly dissipative drift. Math. Nachr., 289 (2016), N 14-15, 1909–1926.

[29] O.A. Manita, Estimates for the Kantorovich distances between solutions to the nonlinear Fokker–Planck–Kolmogorov equation with monotone drift, ArXiv:1507.04014 (2015).

[30] O.A. Manita, M.S. Romanov and S.V. Shaposhnikov, On uniqueness of solutions to nonlinear Fokker–Planck–Kolmogorov equations, Nonlin. Anal., 128 (2015), 199–226.

[31] O.A. Manita and S.V. Shaposhnikov, Nonlinear parabolic equations for measures, St. Petersburg Math. J., 25 (2014), N 1, 43–62.

[32] H.P. McKean, A class of Markov processes associated with nonlinear parabolic equations, Proc. Natl. Acad. Sci. USA, 56 (1966), 1907–1911.

[33] H.P. McKean, Propagation of chaos for a class of non-linear parabolic equations, In: Lecture Series in Differential Equations, session 7, pp. 177–194. Catholic Univ. (1967).

[34] Yu.S. Mishura and A.Yu. Veretennikov, Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations, ArXiv:1603.02212 (2016).

[35] M. Röckner and F.-Y. Wang, Harnack and functional inequalities for generalized Mehler semigroups, J. Funct. Anal., 203 (2003), N 1, 237–261.

[36] Y. Tamura, On asymptotic behaviors of the solution of a nonlinear diffusion equation, J. Fac. Sci. Univ. Tokyo, Sect. 1a, Math., 31 (1984), 195–221.

[37] Y. Tamura, Free energy and the convergence of distributions of diffusion processes of McKean type, J. Fac. Sci. Univ. Tokyo, Sect. 1a, Math., 34 (1987), 443–484.

[38] L.G. Tonoyan, Nonlinear elliptic equations for measures, Dokl. Ross. Akad. Nauk, 439 (2011), N 2, 174–177 (in Russian); English transl.: Dokl. Math., 84 (2011), N 1, 558–561.

[39] A.Yu. Veretennikov, On ergodic measures for McKean–Vlasov stochastic equations, In: Niederreiter H., Talay D. (eds.) Monte Carlo and Quasi-Monte Carlo Methods, pp. 471–486. Springer, Berlin, 2004.

[40] F.-Y. Wang, Functional Inequalities, Markov Semigroups and Spectral Theory, Elsevier, Beijing, 2006.

[41] F.-Y. Wang, Distribution dependent SDEs for Landau type equations, Stoch. Processes Appl. (2017).