SIAM J. NUMER. ANAL.
Vol. 56, No. 4, pp. 2148–2177

ON THE DISCRETIZATION OF SOME NONLINEAR FOKKER–PLANCK–KOLMOGOROV EQUATIONS AND APPLICATIONS*

ELISABETTA CARLINI† AND FRANCISCO J. SILVA‡

Abstract. In this work, we consider the discretization of some nonlinear Fokker–Planck–Kolmogorov equations. The scheme we propose preserves the nonnegativity of the solution, conserves the mass, and, as the discretization parameters tend to zero, has limit measure-valued trajectories which are shown to solve the equation. The main assumptions to obtain a convergence result are that the coefficients are continuous and satisfy a suitable linear growth property with respect to the space variable. In particular, we obtain a new proof of existence of solutions for such equations. We apply our results to some nonlinear examples, including Mean Field Games systems and variations of the Hughes model for pedestrian dynamics.

Key words. nonlinear Fokker–Planck–Kolmogorov equations, numerical analysis, semi-Lagrangian schemes, Markov chain approximation, Mean Field Games

AMS subject classifications. 35Q84, 65N12, 65N75

DOI. 10.1137/17M1143022

1. Introduction.

In this article we consider the nonlinear Fokker–Planck–Kolmogorov (FPK) equation

$$
\text{(FPK)} \qquad \begin{aligned} \partial_t m - \tfrac{1}{2} \sum_{1 \le i,j \le d} \partial^2_{x_i, x_j} \left(a_{i,j}(m,x,t) m \right) + \operatorname{div}\left(b(m,x,t) m \right) &= 0, \\ m(0) &= \bar{m}_0, \end{aligned}
$$

where, denoting by $\mathcal{P}_1(\mathbb{R}^d)$ (respectively, $\mathcal{P}_2(\mathbb{R}^d)$) the space of probability measures on $\mathbb{R}^d$ with first (respectively, second) bounded moments, $\bar{m}_0 \in \mathcal{P}_2(\mathbb{R}^d)$ and

$$
\begin{aligned}
& b : C([0,T]; \mathcal{P}_1(\mathbb{R}^d)) \times \mathbb{R}^d \times [0,T] \to \mathbb{R}^d, \\
& a_{i,j}(m,x,t) := \textstyle\sum_{k=1}^r \sigma_{ik}(m,x,t) \sigma_{jk}(m,x,t) \ \ \forall\, i,\ j = 1, \dots, d, \\
& \ \text{with} \quad \sigma_{i,j} : C([0,T]; \mathcal{P}_1(\mathbb{R}^d)) \times \mathbb{R}^d \times [0,T] \to \mathbb{R} \ \ \forall\, i = 1, \dots, d, \ \ j = 1, \dots, r.
\end{aligned}
$$

Equation (FPK) is understood as an equation for measures in the sense that we seek a solution m in the space $C([0,T]; \mathcal{P}_1(\mathbb{R}^d))$. Note that the coefficients b and $a_{i,j}$ depend, a priori, on the values $m(t) \in \mathcal{P}_1(\mathbb{R}^d)$ in the entire time interval $[0,T]$. The notion of weak solution to this equation, as well as the assumptions we impose on the coefficients b and $\sigma_{i,j}$, will be detailed in section 2.

---

*Received by the editors August 10, 2017; accepted for publication (in revised form) May 16, 2018; published electronically July 17, 2018.
http://www.siam.org/journals/sinum/56-4/M114302.html
Funding: The first author acknowledges financial support by the INdAM GNCS project "Metodi numerici per equazioni iperboliche e cinetiche e applicazioni." The second author is partially supported by the ANR project MFG ANR-16-CE40-0015-01 and the PEPS-INSMI Jeunes project "Some open problems in Mean Field Games" for the years 2016 and 2017. Both authors acknowledge financial support by the PGMO project VarPDEMFG.

†Dipartimento di Matematica Guido Castelnuovo, Sapienza Università di Roma, 00185 Rome, Italy (carlini@mat.uniroma1.it).
‡TSE-R, UMR-CNRS 5314, Université Toulouse I Capitole, 31015 Toulouse, France, and Institut de recherche XLIM-DMI, UMR-CNRS 7252, Faculté des Sciences et Techniques, Université de Limoges, 87060 Limoges, France (francisco.silva@unilim.fr).

Equation (FPK) has been studied mostly in the linear case, i.e., when $b(m,x,t) = b(x,t)$ and $\sigma_{i,j}(m,x,t) = \sigma_{i,j}(x,t)$ for all $i = 1, \dots, d$ and $j = 1, \dots, r$. This is due in part to the close relation between solutions to (FPK) and solutions to the standard stochastic differential equation (SDE)

$$\mathrm{d}X(t) = b(X(t),t)\mathrm{d}t + \sigma(X(t),t)\mathrm{d}W(t), \quad X(0) = x, \tag{1.1}$$

where σ is the matrix $d \times r$ matrix whose (i,j) entry is $\sigma_{i,j}$, W is an r-dimensional Brownian motion, and $x \in \mathbb{R}^d$. Indeed, under some assumptions on b and $\sigma_{i,j}$, it is possible to show a correspondence of solutions to (FPK) and the time marginal laws of weak solutions to (1.1) for almost every $x \in \mathbb{R}^d$ with respect to (w.r.t.) $\bar{m}_0$ (see, e.g., [46, 32, 10] and the references therein). We refer the reader to [10] for a systematic account of the theory of linear FPK equations and their probabilistic interpretation. When $b(m,x,t) = b(m(t),x,t)$ and $\sigma_{i,j}(m,x,t) = \sigma_{i,j}(m(t),x,t)$ the associated FPK equation is often called the McKean–Vlasov equation, and several results exist concerning the well-posedness of the equation and its probabilistic interpretation (see, e.g., [34, 51]). In the case of general nonlinear coefficients, the article [11] provides an existence result when $\sigma_{i,j} \equiv 0$ and in the articles [49, 50] sufficient conditions on the coefficients defining (FPK) are given in order to ensure the existence of solutions in the second order case. The uniqueness of solutions to (FPK) is a difficult matter. The reader is referred to [46, 32] for the analysis in the linear case with rough coefficients, which borrow some ideas from [30, 4] dealing with the analogous problem when $\sigma_{i,j} = 0$, and to [47, 48, 12] for the nonlinear case.

Let us now comment on the numerical approximation of FPK equations. One of the most popular numerical schemes in the linear case is the one introduced by Chang and Cooper in [24]. An interesting feature of this finite difference scheme is that the discrete solution preserves some intrinsic properties of the analytical one such as nonnegativity and conservation of the initial mass (see, e.g., [53]). Starting from this article, several improvements have been obtained in subsequent works; see, for instance, [62, 31], where high order finite difference schemes have been proposed also for the nonlinear case. Let us also mention [6], which deals with the application of this scheme in the context of stochastic optimal control problems. Finally, finite element approximations have also been discussed in [59].

In the 1970s, Kushner provided a systematic procedure to discretize the solution of an SDE by a discrete time Markov chain in a countable state space. The method the author proposes induces finite difference schemes for the associated Kolmogorov backward and forward equations (see, e.g., [39, 40, 41]) and so a finite difference discretization of (FPK) in the linear case. A proof of convergence of the scheme by using probabilistic tools (weak convergence of probability measures) is provided under the assumption that the coefficients of the SDE are bounded and uniformly continuous. More recently, in the context of Mean Field Games (MFGs) systems (see [45, 36]), Achdou and Capuzzo-Dolcetta introduced in [2] a semi-implicit finite difference scheme for a linear FPK equation, which can be seen as a dual scheme for the scheme associated to the dual equation, the linear Kolmogorov backward equation. The scheme is obtained by computing the adjoint scheme of a monotone and consistent discretization of the corresponding dual equation, i.e., the Kolmogorov backward equation. Finally, in the first order case $\sigma_{i,j} = 0$, we refer the reader to the recent articles [28, 58] dealing with explicit upwind finite volume schemes for the linear equation and to [42] for a similar scheme in the nonlinear and nonlocal case. Let us underline that all the schemes mentioned above share some of the good features of the Chang–Cooper scheme. Indeed, the approximated solutions are nonnegative and conserve the initial mass. On the other hand, the main drawback of finite difference and finite element schemes is that, when implemented in their explicit form, they have to satisfy a CFL condition, which implies a strong restriction on the size of the time steps.

A different class of methods in the linear case is the so-called path integration method, introduced in [54]. These are explicit schemes where the marginal laws of the solution of (1.1) are approximated via an Euler–Maruyama discretization of (1.1) using Gaussian one step transition kernels. Recently, in [25], a convergence result for the discrete time marginal laws in the L^1 strong topology was proved in the framework of a linear and uniformly elliptic FPK equation with unbounded coefficients.

Inspired by the papers [21, 22], dealing with the approximation of MFGs, our aim in this article is to provide a discretization of the general FPK equation (FPK) and to establish some convergence results. In the linear case, the scheme that we propose can be seen as a particular Markov chain approximation of (1.1), and, similarly to [2] in the context of finite difference discretizations, it can be obtained as the dual scheme to the semi-Lagrangian (SL) scheme proposed in [14] for the associated linear Kolmogorov backward equation. In this sense, our discretization is related to the one proposed by Kushner in [39] but uses a different Markov chain approximation that allows us to avoid the CFL condition and hence consider large time steps. For this reason, we find that "semi-Lagrangian scheme" is a good appellation for our discretization. More importantly, our scheme naturally adapts to the general (FPK) equation, also preserves the positivity, conserves the total mass, and allows us to obtain convergence results under rather general assumptions on b and $\sigma_{i,j}$. Namely, in Theorem 4.1 we prove that local Lipschitzianity and sublinear growth with respect to the space variable x, uniformly w.r.t. m and t, are sufficient conditions to prove that if the time step h and space step ρ tend to zero and satisfy that $\rho^2/h \to 0$, then every limit point of the approximated solutions (there exists at least one) solves (FPK). Under a suitable modification of the scheme, a similar convergence result is obtained in Theorem 4.2 when the local Lipschitzianity property of b and $\sigma_{i,j}$ is relaxed to merely continuity. Naturally, if the (FPK) equation admits a unique solution, then we get the convergence of the whole sequence of approximated solutions. As a by-product of this result, we obtain a new proof of existence of solutions to (FPK).

Note also that the initial condition $\bar{m}_0$ is rather general; we can consider, for instance, singular measures (e.g., Dirac masses) as initial distributions. Moreover, as we will see in section 5, we can also construct our scheme by using suitable approximations of the coefficients b and $\sigma_{i,j}$ in the case where such coefficients do not have an explicit form and have to be approximated, and the convergence result remains valid.

Let us point out that a different SL scheme for the (FPK) equation has been proposed in [38] in the linear case. In that article, the advection part and the diffusion reaction term are approximated separately by using two fractional steps. Furthermore, in order to obtain a conservative scheme, the SL method applied to the advection part needs to be adjusted. Since our scheme is derived directly from the probabilistic interpretation of (FPK), it has the advantage that the advection and diffusion terms can be treated together and the conservation of the mass is automatically verified (see also the paper [13], where a conservative SL scheme for a parabolic equation in divergence form is studied).

We study in this work the application of the scheme to two nonlinear models (see [23] for some applications in the linear case). In the first one, we apply our scheme to a particular nondegenerate FPK arising in MFGs. The resulting approximation is similar to the one proposed in [21, 22], the main difference being that the nondegeneracy of the system allows us to prove the convergence of the approximation in general dimensions. In the second model, we propose a variation of the Hughes model for pedestrian dynamics (see [37]), where, differently from MFGs, agents do not forecast the evolution of the crowd in order to choose their optimal trajectories. We prove an existence result for the associated FPK, as well as the convergence of the proposed discretization.

The article is organized as follows. In section 2 we introduce the main notation and recall some fundamental results about the space $C([0,T];\mathcal{P}_1(\mathbb{R}^d))$, which are the keys to establish the convergence results. Section 3 presents the scheme, first in the linear case, for pedagogical reasons, and then in the general nonlinear case. In section 4 we prove our main results, concerning the convergence of the discretization. Finally, in section 5 we consider the application of the scheme to the models described in the previous paragraph.

2. Preliminaries. We denote by $\mathcal{P}(\mathbb{R}^d)$ the space of probability measures on $\mathbb{R}^d$. Given a Borel measurable function $\Psi:\mathbb{R}^d\to\mathbb{R}^{d'}$ and $\mu\in\mathcal{P}(\mathbb{R}^d)$, we denote by $\Psi\sharp\mu\in\mathcal{P}(\mathbb{R}^{d'})$ the probability measure defined as $\Psi\sharp\mu(A):=\mu(\Psi^{-1}(A))$ for all $A\in\mathcal{B}(\mathbb{R}^{d'})$. Given $p\in[1,\infty[$, the set $\mathcal{P}_p(\mathbb{R}^d)$ denotes the subset of $\mathcal{P}(\mathbb{R}^d)$ with bounded p moments, i.e.,

$$\mathcal{P}_p(\mathbb{R}^d):=\left\{\mu\in\mathcal{P}(\mathbb{R}^d)\ ;\ \int_{\mathbb{R}^d}|x|^p\mathrm{d}\mu(x)<\infty\right\}.$$

For every $\mu_1,\ \mu_2\in\mathcal{P}_p(\mathbb{R}^d)$, define

$$d_p(\mu_1,\mu_2):=\inf\left\{\left(\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^p\mathrm{d}\gamma(x,y)\right)^{\frac{1}{p}}\ ;\ \gamma\in\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d),\ \pi_1\sharp\gamma=\mu_1,\ \ \pi_2\sharp\gamma=\mu_2\right\},$$

where $\pi_i:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}^d$ $(i=1,2)$ is defined as $\pi_i(x_1,x_2)=x_i$. It is well known that d_p is a distance in $\mathcal{P}_p(\mathbb{R}^d)$ (see, e.g., [61, Theorem 7.3]) and that $\big(\mathcal{P}_p(\mathbb{R}^d),d_p\big)$ is a separable complete metric space (see, e.g., [5, Proposition 7.1.5]). Moreover,

$$d_p(\mu_1,\mu_2)^p\leq\inf\left\{\int_{\mathbb{R}^d}|x-T(x)|^p\,\mathrm{d}\mu_1(x)\ ;\ T:\mathbb{R}^d\to\mathbb{R}^d\text{ is Borel measurable and }\ T\sharp\mu_1=\mu_2\right\}\tag{2.1}$$

with equality if μ_1 has no atoms (see [3, Theorem 2.1]). Finally, let us mention an important result that says that d_1 corresponds to the Kantorovic–Rubinstein metric, i.e.,

$$d_1(\mu_1,\mu_2)=\sup\left\{\int_{\mathbb{R}^d}f(x)\mathrm{d}(\mu_1-\mu_2)(x)\ ;\ f\in\mathrm{Lip}_1(\mathbb{R}^d)\right\},\tag{2.2}$$

where $\mathrm{Lip}_1(\mathbb{R}^d)$ denotes the set of Lipschitz functions defined in $\mathbb{R}^d$ with Lipschitz constant less than or equal to 1 (see, e.g., [61]).

Now, let $\mathcal{C}\subseteq C([0,T];\mathcal{P}_1(\mathbb{R}^d))$, and suppose that there exists a modulus of continuity $\bar\omega:[0,T]\to\mathbb{R}$, i.e., $\bar\omega\geq 0$, $\bar\omega$ is continuous, and $\bar\omega(0)=0$, such that

$$\sup_{\mu\in\mathcal{C}}d_1(\mu(t_1),\mu(t_2))\leq\bar\omega(|t_1-t_2|)\qquad\forall\,t_1,\ t_2\in[0,T].\tag{2.3}$$

Assume in addition that there exists $C>0$ such that

$$\sup_{t\in[0,T]}\int_{\mathbb{R}^d}|x|^2\mathrm{d}\mu(t)(x)\leq C\qquad\forall\,\mu\in\mathcal{C}.\tag{2.4}$$

Since the set $\left\{\mu \in \mathcal{P}_1(\mathbb{R}^d)\ ;\ \int_{\mathbb{R}^d} |x|^2 \mathrm{d}\mu(x) \leq C\right\}$ is compact in $\mathcal{P}_1(\mathbb{R}^d)$ (see [5, Proposition 7.1.5]), (2.4), (2.3), and the Arzelá–Ascoli theorem yield the following result.

LEMMA 2.1. *Under the above assumptions, $\mathcal{C}$ is a relatively compact subset of $C([0,T];\mathcal{P}_1(\mathbb{R}^d))$.*

For notational convenience, for $\psi = b,\ \sigma_{i,j}$ we set $\psi[\mu](x,t) := \psi(\mu,x,t)$. We say that $m \in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ solves (FPK) if for all $t \in [0,T]$ and $\varphi \in C_0^\infty(\mathbb{R}^d)$, the space of C^∞-functions with compact support, we have

$$
\begin{aligned}
\int_{\mathbb{R}^d} \varphi(x)\mathrm{d}m(t)(x) = \quad & \int_{\mathbb{R}^d} \varphi(x)\mathrm{d}\bar{m}_0(x) + \int_0^t \int_{\mathbb{R}^d} \left[b[m](x,s)\cdot\nabla\varphi(x)\right]\mathrm{d}m(s)(x)\mathrm{d}s \\
& + \int_0^t \int_{\mathbb{R}^d} \left[\tfrac{1}{2}\sum_{i,j} a_{i,j}[m](x,s)\partial^2_{x_i,x_j}\varphi(x)\right]\mathrm{d}m(s)(x)\mathrm{d}s.
\end{aligned} \tag{2.5}
$$

The following assumption will be the principal one in the remainder of this paper.

(H) We will suppose that

(i) the maps b and σ are continuous;

(ii) there exists $C > 0$ such that

$$
|b[\mu](x,t)| + |\sigma[\mu](x,t)| \leq C(1+|x|) \ \ \forall\, \mu \in C([0,T];\mathcal{P}_1(\mathbb{R}^d)),\ \ x \in \mathbb{R}^d,\ \ t \in [0,T]. \tag{2.6}
$$

The aim of this article is to study convergent numerical schemes for solutions to (FPK) (if they exist). As can be guessed from the references [46, 32, 10] in the linear case, i.e., when b and $\sigma_{i,j}$ do not depend on m, the existence of solutions to (FPK) should be related to the existence of (weak) solutions to the "extended" McKean–Vlasov equation

$$
\mathrm{d}X(t) = b[m](X(t),t)\mathrm{d}t + \sigma[m](X(t),t)\mathrm{d}W(t), \quad X(0) = X_0. \tag{2.7}
$$

In (2.7), W is an r-dimensional Brownian motion defined on a probability space $(\Omega,\mathcal{F},\mathbb{P})$, m belongs to $C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ and satisfies $m(t) = \mathrm{Law}(X(t))$ for all $t \in [0,T]$, where we have denoted by $\mathrm{Law}(Y)$ the law induced in $\mathbb{R}^d$ by a d-valued random variable Y, and X_0 is a random variable, independent of W, and such that $\mathrm{Law}(X_0) = m_0$.

This observation, relating formally solutions of (FPK) and (2.7), leads us naturally to studying the laws of discrete approximations of (2.7), for which existence is not difficult to show, and then to studying their limit behavior. This strategy will be followed in the next sections.

Remark 2.1. In this article we do not tackle the uniqueness of solutions to (FPK). As can be seen in [46, 32, 10], in the linear case, the study of uniqueness is already quite complicated in the absence of further assumptions on b and $a_{i,j}$ (see, e.g., [32, Theorem 1.3]). We refer the reader to [47, 48, 12] for some recent and interesting results in the general nonlinear case.

3. The fully discrete scheme.

In this section we describe the scheme we propose and study its main properties. In order to introduce the main ideas we will start by considering first the (FPK) equation with $\sigma = 0$ and b independent of m, i.e., the first order linear FPK equation, also called *continuity equation*. Then, we will consider the more general case (i.e., with σ not necessarily identically zero) but still with coefficients b and σ independent of m. Finally, the scheme for the general (FPK) will easily follow by freezing the m dependence of b and σ. We motivate the schemes by assuming stronger assumptions on b and σ, which will imply uniqueness of solutions of the underlying SDEs, in order to take advantage of the semigroup properties of the solutions and somehow guess a consistent approximation.

We assume first that $\sigma \equiv 0$ and that b does not depend on m, i.e., $b[m](x,t) = b(x,t)$. In addition to **(H)**, assume that b is Lipschitz w.r.t. x, uniformly in $t \in [0,T]$. For any $0 \le s \le t \le T$ and $x \in \mathbb{R}^d$, we set $\Phi(x,s,t) = X(t)$, where X is the unique solution of

$$\dot{X}(t') = b(X(t'), t') \quad \text{for } t' \in]s,T[, \quad X(s) = x. \tag{3.1}$$

We know that Φ defines a measurable function of (x,s,t) (if $t \le s$ we simply set $\Phi(x,s,t) = x$). Then, $m \in C([0,T]; \mathcal{P}_1(\mathbb{R}^d))$ defined as

$$m(t)(A) := \Phi(\cdot, 0, t)\sharp \bar{m}_0(A) \quad \forall\, A \in \mathcal{B}(\mathbb{R}^d), \quad t \in [0,T], \tag{3.2}$$

is the unique solution of (FPK) (see, e.g., [5, Chapter 8] for a proof of this classical result in a more general framework). We also have that for all $t \in [0,T]$ and $h \in [0, T-t]$

$$m(t+h)(A) = \Phi(\cdot, t, t+h)\sharp m(t)(A) \quad \forall\, A \in \mathcal{B}(\mathbb{R}^d). \tag{3.3}$$

Given $N \in \mathbb{N}$ we set $h := T/N$ and $t_k := kh$ $(k = 0, \dots, N)$. Let us consider the following *explicit* time discretization of (3.2), based on a standard explicit Euler approximation of (3.1) and property (3.3):

$$m_0 = \bar{m}_0, \quad m_{k+1} := \Phi_k \sharp m_k, \quad \text{where} \quad \Phi_k(x) := x + hb(x, t_k) \quad \forall\, k = 0, \dots, N-1. \tag{3.4}$$

The sequences m_k and Φ_k $(k = 0, \dots, N)$ depend of course on h but we have omitted this dependence in order to ease the reading. Let us now introduce some standard notation that will be used for the space discretization. Let $\rho > 0$ be a given *space step*, and consider a uniform space grid

$$\mathcal{G}_\rho := \{x_i = i\rho\,;\ i \in \mathbb{Z}^d\}.$$

Given a uniform Cartesian grid of $\mathbb{R}^d$, with vertices belonging to $\mathcal{G}_\rho$, we consider a $\mathbb{Q}_1$ finite element basis $(\beta_i)_{i \in \mathbb{Z}^d}$; i.e., for all $i \in \mathbb{Z}^d$, β_i is a continuous function whose restriction to each d-cube of the grid is a polynomial of degree less than or equal to 1 and satisfies that $\beta_i(x_j) = 1$ if $i = j$ and $\beta_i(x_j) = 0$ otherwise. Moreover, the support $\text{supp}(\beta_i)$ of β_i is compact and

$$0 \le \beta_i \le 1 \quad \forall\, i \in \mathbb{Z}^d, \quad \text{and} \quad \sum_{i \in \mathbb{Z}^d} \beta_i(x) = 1 \quad \forall\, x \in \mathbb{R}^d.$$

We look for a discretization of (3.4) taking the form

$$m_k = \sum_{i \in \mathbb{Z}^d} m_{i,k}\delta_{x_i} \quad \forall\, k = 0, \dots, N-1. \tag{3.5}$$

For all $i \in \mathbb{Z}^d$, let us define

$$E_i := \left\{ x \in \mathbb{R}^d\,;\ |x - x_i|_\infty \le \frac{\rho}{2} \right\}.$$

In section 4 we will let $\rho \downarrow 0$; thus, without loss of generality, we can assume that $\bar{m}_0(\partial E_i) = 0$ for all $i \in \mathbb{Z}^d$. We define the weights $m_{i,k}$ of the Dirac masses in (3.5) inductively as

$$m_{i,0} = \bar{m}_0(E_i), \quad m_{i,k+1} = \sum_{j \in \mathbb{Z}^d} \beta_i(\Phi_{j,k}) m_{j,k} \quad \forall\, k = 0, \dots, N-1, \ \ i \in \mathbb{Z}^d, \tag{3.6}$$

where

$$\Phi_{i,k} := \Phi_k(x_i) = x_i + hb(x_i, t_k) \quad \forall\, i \in \mathbb{Z}^d. \tag{3.7}$$

The sequences of weights in (3.6) depend on (ρ, h), but, for notational convenience, we have omitted this dependence.

Remark 3.1. (i) In order to understand the intuitive meaning of (3.6), take $d = 1$, $\rho = 1$, and $\beta_i(x) := \max\{1 - |x - x_i|, 0\}$ for all $i \in \mathbb{Z}$, $x \in \mathbb{R}$. Then, the mass $m_{i,k+1}$, at x_i at time t_{k+1}, is obtained by first considering the set $\mathcal{A}_{i,k}$ of j's such that $\Phi_{j,k} \in \text{supp}(\beta_i)$ and then adding the masses $m_{j,k}$ ($j \in \mathcal{A}_{i,k}$) weighted by $1-|\Phi_{j,k}-x_i|$. For instance, if $\Phi_{j,k} = x_i + 1/2$, then, at the discrete time $k+1$, half of the mass $m_{j,k}$ will be in x_i and the other half will be in x_{i+1}.

(ii) In this deterministic setting if $d = 1$ it is easy to check that (3.6) coincides with the scheme proposed in [55].

Remark 3.2. In the case of bounded space domains, the scheme (3.6) can be extended to triangular meshes by considering a $\mathbb{P}_1$ basis $(\hat{\beta}_i)$ (see, e.g., [56, Chapter IV]) and modifying the sets E_i, which are used to construct the approximation of the initial measure $\bar{m}_0$. The triangular mesh can be chosen as general as long as the interpolation estimate (4.3) in the next section holds true. We refer the reader to [18] for an example of this extension applied to a particular instance of (FPK).

Now, if $\sigma[m](x,t) = \sigma(x,t)$ is not identically zero, we can consider the same type of scheme, taking into account that the characteristics curves are stochastic. Indeed, consider a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, an r-dimensional Brownian motion W defined in this probability space and adapted to the filtration $\mathbb{F} := \{\mathcal{F}_t\}_{t\in[0,T]}$. Define $\Phi : \Omega \times \mathbb{R}^d \times [0,T] \times [0,T] \to \mathbb{R}^d$ as $\Phi(\omega, x, s, t) = x$ if $t \le s$ and, for $s < t$, $\Phi(\omega, x, s, t) = X(t, \omega)$, where X solves

$$\mathrm{d}X(t') = b(X(t'), t')\mathrm{d}t' + \sigma(X(t'), t')\mathrm{d}W(t') \quad \text{for } t' \in]s, T[, \quad X(s) = x. \tag{3.8}$$

Then, assuming that b and σ are Lipschitz with respect to x, uniformly in $t \in [0,T]$, we have that (see, e.g., the classical monograph [60])

$$m(t)(A) := \int_\Omega \Phi(\omega, \cdot, 0, t)\sharp\bar{m}_0(A)\mathrm{d}\mathbb{P}(\omega) = \mathbb{E}\left(\Phi(\cdot, 0, t)\sharp\bar{m}_0(A)\right) \quad \forall\, A \in \mathcal{B}(\mathbb{R}^d), \quad t \in [0,T], \tag{3.9}$$

where, as usual, we have omitted the dependence of Φ on ω inside the expectation. Analogously to (3.3), we have that

$$m(t+h)(A) = \int_\Omega \Phi(\omega, \cdot, t, t+h)\sharp m(t)(A)\mathrm{d}\mathbb{P}(\omega) = \mathbb{E}\left(\Phi(\cdot, t, t+h)\sharp m(t)(A)\right) \quad \forall\, A \in \mathcal{B}(\mathbb{R}^d). \tag{3.10}$$

Therefore, if we discretize the Brownian motion W by an r-dimensional random walk with N time steps, the stochastic characteristic

$$X(t+h) = X(t) + \int_t^{t+h} b(X(t'), t')\mathrm{d}t' + \int_t^{t+h} \sigma(X(t'), t')\mathrm{d}W(t')$$

can be approximated with an explicit Euler scheme by

$$X(t+h) = X(t) + hb(X(t), t) + \sqrt{rh}\,\sigma(X(t), t)Z, \tag{3.11}$$

where Z is an r-valued random variable, independent of $X(t)$, satisfying that for all $\ell = 1, \dots, r$,

(3.12)
$$\mathbb{P}(Z^\ell = 1) = \mathbb{P}(Z^\ell = -1) = \frac{1}{2r} \quad \text{and} \quad \mathbb{P}\left(\bigcup_{1 \le \ell_1 < \ell_2 \le r} \{Z^{\ell_1} \neq 0\} \cap \{Z^{\ell_2} \neq 0\} \right) = 0.$$

Relations (3.11)–(3.12) motivate the following extensions of $\Phi_{i,k}$, defined in (3.7):

(3.13)
$$\begin{array}{rcl} \Phi_{i,k}^{\ell,+} & := & x_i + hb(x_i, t_k) + \sqrt{rh}\sigma_\ell(x_i, t_k) \quad \forall\, i \in \mathbb{Z}^d, \;\; k = 0, \dots, N-1, \;\; \ell = 1, \dots, r, \\ \Phi_{i,k}^{\ell,-} & := & x_i + hb(x_i, t_k) - \sqrt{rh}\sigma_\ell(x_i, t_k) \quad \forall\, i \in \mathbb{Z}^d, \;\; k = 0, \dots, N-1, \;\; \ell = 1, \dots, r. \end{array}$$

Inspired by (3.6), relation (3.10) induces the following *explicit* scheme:

(3.14)
$$\begin{array}{rcl} m_{i,0} & := & \bar{m}_0(E_i) \quad \forall\, i \in \mathbb{Z}^d, \\ m_{i,k+1} & := & \frac{1}{2r} \sum_{\ell=1}^{r} \sum_{j \in \mathbb{Z}^d} \left[\beta_i(\Phi_{j,k}^{\ell,+}) + \beta_i(\Phi_{j,k}^{\ell,-}) \right] m_{j,k} \quad \forall\, i \in \mathbb{Z}^d, \;\; k = 0, \dots, N-1. \end{array}$$

Remark 3.3. Note that the previous scheme is conservative. Indeed, for all $k = 0, \dots, N-1$,

$$\sum_{i \in \mathbb{Z}^d} m_{i,k+1} = \sum_{j \in \mathbb{Z}^d} \frac{m_{j,k}}{2r} \sum_{\ell=1}^{r} \sum_{i \in \mathbb{Z}^d} \left[\beta_i(\Phi_{j,k}^{\ell,+}) + \beta_i(\Phi_{j,k}^{\ell,-}) \right] = \sum_{i \in \mathbb{Z}^d} m_{i,k},$$

and so $\sum_{i \in \mathbb{Z}^d} m_{i,k+1} = \sum_{i \in \mathbb{Z}^d} m_{i,0} = 1$.

Markov chain interpretation. Note that (3.14) can be interpreted in terms of a discrete time Markov chain in a countable state space. Indeed, given the initial law $m_{\cdot,0}$ on $\mathcal{G}_\rho$, consider the nonhomogeneous Markov chain $\{X_k \; ; \; k = 0, \dots, N\}$ with values in $\mathcal{G}_\rho$ defined by the previous initial law and the transition probabilities

$$p_{ji}^{(k)} := \mathbb{P}\left(X_{k+1} = x_i \mid X_k = x_j \right) := \frac{1}{2r} \sum_{\ell=1}^{r} \left[\beta_i(\Phi_{j,k}^{\ell,+}) + \beta_i(\Phi_{j,k}^{\ell,-}) \right] \; \forall\, i, j \in \mathbb{Z}^d, \;\; k = 0, \dots, N-1.$$

Then, (3.14) gives the distribution of X_k for all $k = 0, \dots, N$.

Remark 3.4. (i) Note that if $\sigma \equiv 0$, we recover the scheme (3.6).

(ii) As we will see in section 4, the Markov chain $(X_k)_{k=0}^N$ is a consistent approximation, in the sense of Kushner (see [40]), of the diffusion in (3.8) with $s = 0$ and with $\mathrm{Law}(X_0) = \bar{m}_0$. It is easily seen that, as a function of $\bar{m}_0$, scheme (3.14) can be formally understood as the dual scheme associated to the SL scheme (see [52]) for the Kolmogorov backward equation

$$\begin{array}{rcl} \partial_t u - \frac{1}{2} \sum_{1 \le i,j \le d} a_{i,j} \partial^2_{x_i, x_j} u + b \cdot \nabla u & = & 0, \\ u(\cdot, T) & = & g(\cdot), \end{array}$$

as a function of $g \in C_b(\mathbb{R}^d)$ (where $C_b(\mathbb{R}^d)$ is the space of bounded continuous functions in $\mathbb{R}^d$).

(iii) In [19, section 3.1], it is shown that scheme (3.14) can also be constructed from the weak formulation of (FPK) (when b and σ are independent of m).

In the general nonlinear case, as we have explained at the end of section 2, formally, m solves (FPK) iff for all $t \in [0,T]$, we have that $m(t) = \text{Law}(X(t))$, where X solves (2.7) (assuming that (2.7) admits a solution in a weak sense). On the other hand, even in the particular case of regular coefficients and local in time dependence on m, i.e., $b[m](x,t) = b(m(t),x,t)$ and $\sigma[m](x,t) = \sigma(m(t),x,t)$, with b and σ regular w.r.t. x, X is not a Markov process. Nevertheless, loosely speaking again, X solves (2.7) iff $\text{Law}(X(\cdot)) \in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ is a fixed point of the application

$$\mu \in C([0,T];\mathcal{P}_1(\mathbb{R}^d)) \mapsto \mathcal{F}(\mu) := \text{Law}(X[\mu](\cdot)) \in C([0,T];\mathcal{P}_1(\mathbb{R}^d)), \tag{3.15}$$

where $X[\mu](\cdot)$ solves

$$\mathrm{d}X(t) = b[\mu](X(t),t)\mathrm{d}t + \sigma[\mu](X(t),t)\mathrm{d}W(t) \;\text{ for } t \in]0,T[, \quad X(0) = X_0. \tag{3.16}$$

Since for every fixed μ, $X[\mu]$ defines a Markov diffusion, we can apply (3.14) to approximate its law.

Even if the previous discussion is purely formal, it provides the idea of how to construct a natural discretization of (FPK) by considering a discrete version of the fixed-point problem (3.15), which will be constructed using (3.14). However, since $b[\cdot](x,t)$ and $\sigma[\cdot](x,t)$ act on $C([0,T];\mathcal{P}_1(\mathbb{R}^d))$, given ρ and h we first need to map elements of the set

$$\mathcal{S}^{\rho,h} := \left\{ (\mu_{i,k})_{i\in\mathbb{Z}^d,\, k=0,\dots,N} \;;\; \mu_{i,k} \geq 0 \;\; \forall\, i \in \mathbb{Z}^d, \sum_{i\in\mathbb{Z}^d} \mu_{i,k} = 1, \;\; \sum_{i\in\mathbb{Z}^d} |x_i|\mu_{i,k} < \infty \;\; \forall\, k = 0,\dots,N \right\}$$

to elements of $C([0,T];\mathcal{P}_1(\mathbb{R}^d))$. This can be naturally done by using time interpolation. Given $\mu \in \mathcal{S}^{\rho,h}$, we still denote by μ the element of $C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ defined by

$$\mu(t) := \left(\frac{t-t_k}{h}\right) \sum_{i\in\mathbb{Z}^d} \mu_{i,k+1}\delta_{x_i} + \left(\frac{t_{k+1}-t}{h}\right) \sum_{i\in\mathbb{Z}^d} \mu_{i,k}\delta_{x_i} \quad \text{if } t \in [t_k, t_{k+1}[, \tag{3.17}$$

for all $k = 0,\dots,N-1$. Using this notation, define

$$\mu \in \mathcal{S}^{\rho,h} \mapsto \mathcal{F}^{\rho,h}(\mu) := (\text{Law}(X_k[\mu]))_{k=0,\dots,N} \in \mathcal{S}^{\rho,h}, \tag{3.18}$$

where we compute $\mathbb{P}(X_k[\mu] = x_i) := m_{i,k}[\mu]$ recursively with (3.14) with $\Phi^{\ell,+}_{j,k}$ and $\Phi^{\ell,-}_{j,k}$ replaced by

$$\Phi^{\ell,+}_{i,k}[\mu] := x_i + hb[\mu](x_i,t_k) + \sqrt{rh}\sigma_\ell[\mu](x_i,t_k), \quad \Phi^{\ell,-}_{i,k}[\mu] := x_i + hb[\mu](x_i,t_k) - \sqrt{rh}\sigma_\ell[\mu](x_i,t_k), \tag{3.19}$$

respectively. For $\mu \in \mathcal{S}^{\rho,h}$ let us set $\nu_k[\mu] := \left(\mathcal{F}^{\rho,h}(\mu)\right)_k$ $(k = 0,\dots,N)$. By definition of the scheme, using that $\bar{m}_0 \in \mathcal{P}_2(\mathbb{R}^d)$ and that $\bar{m}_0(\partial E_i) = 0$, we have

$$\begin{aligned}
\int_{\mathbb{R}^d} |x|^2 \mathrm{d}\nu_0[\mu](x) &= \sum_{i\in\mathbb{Z}^d} |x_i|^2 \bar{m}_0(E_i) \\
&= \sum_{i\in\mathbb{Z}^d} \int_{E_i} |x - (x - x_i)|^2 \mathrm{d}\bar{m}_0(x) \\
&\leq 2\int_{\mathbb{R}^d} |x|^2 \mathrm{d}\bar{m}_0(x) + d\rho^2/2 < +\infty.
\end{aligned}$$

Moreover, arguing exactly as in the proof of Proposition 4.1 in the next section, under **(H)**(ii) we obtain the existence of $c > 0$, independent of μ, such that

$$\int_{\mathbb{R}^d} |x|^2 \mathrm{d}\nu_k[\mu](x) = \sum_{i\in\mathbb{Z}^d} |x_i|^2 \nu_{i,k}[\mu] \leq c \quad \forall\, k = 0,\dots,N, \quad \forall\, \mu \in \mathcal{S}^{\rho,h}. \tag{3.20}$$

In particular, $\mathcal{F}^{\rho,h}$ is well defined. The discretization of (FPK) that we propose is

$$\text{find } m \in \mathcal{S}^{\rho,h} \text{ such that } \ m = \mathcal{F}^{\rho,h}(m), \tag{3.21}$$

or equivalently, find $m \in \mathcal{S}^{\rho,h}$ such that

(3.22)

$$\begin{aligned} & m_{i,0} = \bar{m}_0(E_i) \quad \forall\, i \in \mathbb{Z}^d, \\ & m_{i,k+1} = \tfrac{1}{2r}\sum_{\ell=1}^{r}\sum_{j\in\mathbb{Z}^d}\left[\beta_i(\Phi^{\ell,+}_{j,k}[m]) + \beta_i(\Phi^{\ell,-}_{j,k}[m])\right] m_{j,k} \quad \forall\, i \in \mathbb{Z}^d,\ \ k = 0,\dots,N-1. \end{aligned}$$

Now, let us prove the existence of solutions of (3.21). In the following proof we identify $\mathcal{S}^{\rho,h}$ with a subset of $\mathcal{P}_1(\mathbb{R}^d)^{N+1}$ by setting for all $\mu \in \mathcal{S}^{\rho,h}$

$$\mu_k := \sum_{i\in\mathbb{Z}^d} \mu_{i,k}\delta_{x_i} \quad \forall\, k = 0,\dots,N. \tag{3.23}$$

PROPOSITION 3.1. *There exists at least one solution $m^{\rho,h} \in \mathcal{S}^{\rho,h}$ of* (3.21).

Proof. As before, for $\mu \in \mathcal{S}^{\rho,h}$ denote by $\nu_k[\mu] := \left(\mathcal{F}^{\rho,h}(\mu)\right)_k$ $(k = 0,\dots,N)$. Let $c > 0$ be such that (3.20) holds. Then, defining

$$\mathcal{S}^{\rho,h}_c := \left\{ \mu \in \mathcal{S}^{\rho,h}\ ;\ \sum_{i\in\mathbb{Z}^d} |x_i|^2 \mu_{i,k} \leq c \qquad \forall\, k = 0,\dots,N \right\},$$

we have that $\mathcal{S}^{\rho,h}_c$ is convex and $\mathcal{F}^{\rho,h}(\mathcal{S}^{\rho,h}_c) \subseteq \mathcal{S}^{\rho,h}_c$. Moreover, by [5, Proposition 7.1.5 and Proposition 5.1.8], Fatou's lemma, and the identification (3.23), we have that $\mathcal{S}^{\rho,h}_c$ is a compact subset of $\mathcal{P}_1(\mathbb{R}^d)^{N+1}$. Finally, if $\mu_n \in \mathcal{S}^{\rho,h}$ converge to $\mu \in \mathcal{S}^{\rho,h}$, seen as elements of $\mathcal{P}_1(\mathbb{R}^d)^{N+1}$, then, using the extension (3.17), assumption **(H)**(i) implies that $\Phi^{\ell,+}_{j,k}[\mu_n]$ and $\Phi^{\ell,-}_{j,k}[\mu_n]$ converge to $\Phi^{\ell,+}_{j,k}[\mu]$ and $\Phi^{\ell,-}_{j,k}[\mu]$, respectively, which implies the continuity of $\mathcal{F}^{\rho,h}$. Since the topology of $\mathcal{P}_1(\mathbb{R}^d)$ is the restriction to $\mathcal{P}_1(\mathbb{R}^d)$ of the topology induced by the modified Kantorovic–Rubinstein norm on the linear space of all bounded Borel measures on $\mathbb{R}^d$ with respect to which all the Lipschitz functions are integrable (see the discussion before Proposition 1.1.4 in [9]), the existence of a solution of (3.21) follows from Schauder's fixed-point theorem. ∎

The computation in Remark 3.3 applies in the nonlinear case, and so the scheme is conservative.

Remark 3.5 (explicit and implicit schemes). Note that if for all $t \in [0,T]$, $b[m](x,t) = \hat{b}(m(\cdot\wedge t), x, t)$ and $\sigma[m](x,t) = \hat{\sigma}(m(\cdot\wedge t), x, t)$ for some functions $\hat{b}$ and $\hat{\sigma}$ defined in $C([0,T];\mathcal{P}_1(\mathbb{R}^d)) \times \mathbb{R}^d \times [0,T]$, then (3.19) implies that the scheme (3.22) is explicit in the time steps, and the existence of solution, as well as the uniqueness, of the scheme is straightforward. Notice that the explicit character of the scheme is a consequence of (3.19). If, instead, an implicit discretization of the underlying ODE is considered, then the scheme (3.22) would become implicit also in this case.

In the general case, i.e., when $b[m](x,t)$ and $\sigma[m](x,t)$ depend on $m(s)$ for some $s \in (t,T]$, the scheme is implicit in the time steps and, as we have seen in the proof of the previous proposition, the existence of solutions is a consequence of the Shauder fixed-point theorem. The latter situation is the one we face when we consider MFGs, as we will see in section 5.1. In the implicit cases, the uniqueness of solutions is generally not true and its fulfilment depends on the problem at hand.

4. Convergence analysis. In this section we prove our main results concerning the convergence of solutions of (3.22) to solutions of (FPK). In our first main result in Theorem 4.1, we prove the desired convergence result under an additional local Lipschitz assumption on b and σ, with respect to the space variable, and suitable conditions on the time and space steps. In Theorem 4.2, we consider a variation of the scheme in section 3, with regularized coefficients, and we prove a similar convergence result by assuming only **(H)** and some conditions on the discretization parameters.

Let us first introduce and recall some classical properties of the linear interpolation operator that we consider (see, e.g., [26, 57] for further details). Let $B(\mathcal{G}_\rho)$ be the space of bounded functions on $\mathcal{G}_\rho$ and for $f \in B(\mathcal{G}_\rho)$ set $f_i := f(x_i)$. We consider the following linear interpolation operator:

$$I[f](\cdot) := \sum_{i\in\mathbb{Z}^d} f_i \beta_i(\cdot) \quad \text{for } f \in B(\mathcal{G}_\rho). \tag{4.1}$$

Given $\phi \in C_b(\mathbb{R}^d)$, let us define $\hat{\phi} \in B(\mathcal{G}_\rho)$ by $\hat{\phi}_i := \phi(x_i)$ for all $i \in \mathbb{Z}^d$. Suppose that $\phi : \mathbb{R}^d \to \mathbb{R}$ is Lipschitz with constant L. Then,

$$I[\hat{\phi}] \quad \text{is Lipschitz with constant } \sqrt{d}L \quad \text{and} \quad \sup_{x\in\mathbb{R}^d} |I[\hat{\phi}](x) - \phi(x)| = c_0\rho \tag{4.2}$$

for some $c_0 > 0$. On the other hand, if $\phi \in \mathcal{C}^2(\mathbb{R}^d)$, with bounded second derivatives, then there exists $c_1 > 0$ such that

$$\sup_{x\in\mathbb{R}^d} |I[\hat{\phi}](x) - \phi(x)| = c_1\rho^2. \tag{4.3}$$

Now, let $\{N_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathbb{N}$ such that $N_n \to \infty$ as $n \to \infty$ and set $h_n := T/N_n$. Given a sequence of space steps ρ_n, such that $\rho_n \to 0$ as $n \to \infty$, we want to study the limit behavior of the extensions to $C([0,T];\mathcal{P}_1(\mathbb{R}^d))$, defined in (3.17), of sequences of solutions $m^n := m^{\rho_n,h_n} \in \mathcal{S}^{\rho_n,h_n}$ of (3.21), with $\rho = \rho_n$ and $h = h_n$ (by Proposition 3.1 we know that (3.21) admits at least one solution).

First note that by considering the transport plan $T(x) = x_i$ if $x \in E_i$, and arbitrarily defined in ∂E_i (because $\bar{m}_0(\partial E_i) = 0$), we have that $T\sharp\bar{m}_0 = m^n(0)$. Thus, inequality (2.1) with $p = 1$ yields

$$d_1(\bar{m}_0, m^n(0)) \leq \int_{\mathbb{R}^d} |x - T(x)| \mathrm{d}\bar{m}_0(x) = \sum_{i\in\mathbb{Z}^d} \int_{E_i} |x - x_i| \mathrm{d}\bar{m}_0(x) \leq \sqrt{d}\rho_n/2, \tag{4.4}$$

which implies that $m^n(0) \to \bar{m}_0$ in $\mathcal{P}_1(\mathbb{R}^d)$ as $n \to \infty$. We prove in this section that under suitable conditions over ρ_n and h_n the set $\mathcal{C} := \{m^n \ ; \ n \in \mathbb{N}\}$ satisfies (2.3) and (2.4). Therefore, Lemma 2.1 will imply that m^n has at least one limit point $m \in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$. In the proof of (2.3) and (2.4) we will need some properties of the Markov chain X^n, defined by the transition probabilities

$$p_{ji}^{n,k} := \mathbb{P}\left(X_{k+1}^n = x_i \mid X_k^n = x_j\right) := \frac{1}{2r}\sum_{\ell=1}^{r}\left[\beta_i(\Phi_{j,k}^{\ell,+}[m^n]) + \beta_i(\Phi_{j,k}^{\ell,-}[m^n])\right] \quad \forall\, i,j \in \mathbb{Z}^d,$$

and $k=0,\dots,N_n-1$. Note that (3.22) implies that the marginal distributions of this chain are given by m^n. Moreover, it is easy to check that (4.2) (resp., (4.3)) implies that if $\phi:\mathbb{R}^d\to\mathbb{R}$ is Lipschitz (resp., C^2 with bounded second derivatives), then

(4.5)

$$\begin{aligned}\mathbb{E}\left(\phi(X^n_{k+1})\big|X^n_k=x_i\right)&=\tfrac{1}{2r}\textstyle\sum_{\ell=1}^r\left[I[\hat\phi](\Phi^{\ell,+}_{i,k}[m^n])+I[\hat\phi](\Phi^{\ell,-}_{i,k}[m^n])\right]\\&=\tfrac{1}{2r}\textstyle\sum_{\ell=1}^r\left[\phi(\Phi^{\ell,+}_{i,k}[m^n])+\phi(\Phi^{\ell,-}_{i,k}[m^n])\right]+O(\rho_n),\end{aligned}$$

$$\text{(resp.)}\quad \mathbb{E}\left(\phi(X^n_{k+1})\big|X^n_k=x_i\right)=\tfrac{1}{2r}\textstyle\sum_{\ell=1}^r\left[\phi(\Phi^{\ell,+}_{i,k}[m^n])+\phi(\Phi^{\ell,-}_{i,k}[m^n])\right]+O(\rho_n^2).$$

PROPOSITION 4.1. *Suppose that $\rho_n^2=O(h_n)$. Then, there exists a constant $c>0$ such that*

$$\sup_{n\in\mathbb{N}}\sup_{t\in[0,T]}\int_{\mathbb{R}^d}|x|^2\mathrm{d}m^n(t)\le c. \tag{4.6}$$

Proof. By (3.17), it is enough to show that there exists $c>0$, independent of n, such that

$$\sup_{k=0,\dots,N_n}\int_{\mathbb{R}^d}|x|^2\mathrm{d}m^n(t_k)\le c. \tag{4.7}$$

For notational convenience we will omit the superscript n. By definition,

$$\int_{\mathbb{R}^d}|x|^2\mathrm{d}m(t_{k+1})(x)=\sum_{i\in\mathbb{Z}^d}|x_i|^2m_{i,k+1}=\mathbb{E}(|X_{k+1}|^2),$$

from which, using (4.5) and **(H)**(ii),

$$\begin{aligned}\mathbb{E}(|X_{k+1}|^2)&=\textstyle\sum_{i\in\mathbb{Z}^d}\mathbb{E}\left(|X_{k+1}|^2\big|X_k=x_i\right)m_{i,k},\\&=\tfrac{1}{2r}\textstyle\sum_{\ell=1}^r\sum_{i\in\mathbb{Z}^d}\left[\left|\Phi^{\ell,+}_{i,k}[m]\right|^2+\left|\Phi^{\ell,-}_{i,k}[m]\right|^2\right]m_{i,k}+O(\rho_n^2),\\&=\mathbb{E}\left[\left|X_k+h_nb[m](X_k,t_k)+\sqrt{rh_n}\sigma[m](X_k,t_k)Z_k\right|^2\right]+O(\rho_n^2),\\&=\mathbb{E}\left[|X_k|^2+h_n^2|b[m](X_k,t_k)|^2+rh_n\textstyle\sum_{\ell=1}^r|\sigma_\ell[m](X_k,t_k)|^2+2h_nX_k\cdot b[m](X_k,t_k)\right]\\&\quad+O(\rho_n^2),\\&\le(1+Ch_n)\mathbb{E}(|X_k|^2)+Ch_n+O(\rho_n^2),\end{aligned}$$

where Z_k is an r-valued random variable, independent of X_k, satisfying (3.12) and C is independent of n. Iterating, we get

$$\begin{aligned}\textstyle\int_{\mathbb{R}^d}|x|^2\mathrm{d}m(t_{k+1})(x)&\le(1+Ch_n)^{\frac{T}{h_n}}\mathbb{E}(|X_0|^2)+(Ch_n+O(\rho_n^2))\textstyle\sum_{k=0}^N(1+Ch_n)^k\\&\le e^{CT}\mathbb{E}(|X_0|^2)+\tfrac{T}{h_n}\left[Ch_n+O(\rho_n^2)\right]e^{CT}\\&\le e^{CT}\mathbb{E}(|X_0|^2)+\left(CT+O\left(\tfrac{\rho_n^2}{h_n}\right)\right)e^{CT},\end{aligned}$$

from which the result follows. ∎

Now, we prove a consistency property of the chain X^n in the spirit of Kushner [40]. For all $0\le k\le N_n-1$ let us define $\delta_kX^n:=X^n_{k+1}-X^n_k$, $Y^n_k:=\delta_kX^n-\mathbb{E}\left(\delta_kX^n|X^n_k\right)$.

LEMMA 4.1. *For all $k=0,\dots,N_n-1$ we have that*

$$\begin{aligned}\mathbb{E}(\delta_kX^n|X^n_k)&=h_nb[m^n](X^n_k,t_k),\\\mathbb{E}(|Y^n_k|^2|X^n_k)&=h_n\textstyle\sum_{\ell=1}^r|\sigma_\ell[m^n](X^n_k,t_k)|^2+O(\rho_n^2).\end{aligned}$$

Proof. By definition of $p^{n,k}_{i_k,i_{k+1}}$ we have

$$\begin{aligned}\mathbb{E}(\delta_k X^n | X^n_k = x_{i_k}) &= \textstyle\sum_{i_{k+1}} \left(x_{i_{k+1}} - x_{i_k}\right) p^{n,k}_{i_k,i_{k+1}} \\ &= \tfrac{1}{2r} \textstyle\sum_{\ell=1}^{r} \Big(I[\mathrm{id} - x_{i_k}](\Phi^{\ell,+}_{i_k,k}[m^n]) + I[\mathrm{id} - x_{i_k}](\Phi^{\ell,-}_{i_k,k}[m^n]) \Big) \\ &= h_n b[m^n](x_{i_k}, t_k),\end{aligned}$$

where $\mathrm{id}(x) = x$ and the last equality follows from the fact that $I[\mathrm{id}-x_{i_k}](y) = y-x_{i_k}$ for all $y \in \mathbb{R}^d$. Analogously,

$$\mathbb{E}(|Y^n_k|^2 | X^n_k = x_{i_k}) = \sum_{i_{k+1}} \left[x_{i_{k+1}} - x_{i_k} - \mathbb{E}(\delta_k X^n | X^n_k = x_{i_k})\right]^2 p^{n,k}_{i_k,i_{k+1}}.$$

Using (4.3) and the definition of $p^{n,k}_{i_k,i_{k+1}}$ again we get that

$$\mathbb{E}(|Y^n_k|^2 | X_k = x_{i_k}) = h_n \sum_{\ell=1}^{r} |\sigma_\ell[m^n](x_{i_k}, t_k)|^2 + O(\rho_n^2),$$

from which the result follows. ◻

Now, we prove that $\mathcal{C} := \{m^n \,;\, n \in \mathbb{N}\}$ satisfies (2.3).

PROPOSITION 4.2. *Suppose that $\rho_n^2 = O(h_n)$. Then, there exists a constant $C > 0$ such that*

$$\sup_{n\in\mathbb{N}} d_2(m^n(t), m^n(s)) \leq C|t-s|^{\frac{1}{2}} \quad \forall\, t, s \in [0,T]. \tag{4.8}$$

In particular, since $d_1 \leq d_2$, we have that $\mathcal{C}$ satisfies (2.3).

Proof. The proof is divided into two steps.

Step 1. We first show that for given N_n there exists a constant C, independent of n, such that

$$d_2(m^n(t_k), m^n(t'_k)) \leq C\sqrt{|k-k'|h_n} \qquad \forall\, k, k' = 0, \dots, N_n. \tag{4.9}$$

We assume, without loss of generality, that $k' = 0$. For notational convenience, we omit the superscript n on the sequences X^n_k, $\delta_k X^n$, and Y^n_k. By the definition of d_2 we have

$$d_2(m^n(t_k), m^n_0) \leq \left[\mathbb{E}(|X_k - X_0|^2)\right]^{\frac{1}{2}}. \tag{4.10}$$

We have that

$$\mathbb{E}\left(|X_k - X_0|^2\right) = \mathbb{E}\left|\sum_{p=0}^{k-1} \left(Y_p + \mathbb{E}\left(\delta_p X | X_p\right)\right)\right|^2 \leq 2\mathbb{E}\left|\sum_{p=0}^{k-1} Y_p\right|^2 + 2\mathbb{E}\left|\sum_{p=0}^{k-1} \mathbb{E}\left(\delta_p X | X_p\right)\right|^2. \tag{4.11}$$

Now, for $0 \leq r < l \leq k-1$ conditioning on $\mathcal{F}_l := \sigma(X_0, \dots, X_l)$ and using that, by the Markov property, $\mathbb{E}(\delta_l X | \mathcal{F}_l) = \mathbb{E}(\delta_l X | X_l)$ we get

$$\mathbb{E}(Y_l \cdot Y_r) = \mathbb{E}\left[\left(\delta_l X - \mathbb{E}\left(\delta_l X | X_l\right)\right) \cdot Y_r\right] = \mathbb{E}(\mathbb{E}\left[\left(\delta_l X - \mathbb{E}\left(\delta_l X | X_l\right)\right) \big| \mathcal{F}_l\right] \cdot Y_r) = 0,$$

and so, by Lemma 4.1,

$$
\begin{aligned}
\mathbb{E}\left|\sum_{p=0}^{k-1} Y_p\right|^2 &= \sum_{p=0}^{k-1} \mathbb{E}(\mathbb{E}(|Y_p|^2|X_p)) \\
&= h_n \sum_{p=0}^{k-1} \sum_{\ell=1}^{r} \mathbb{E}(|\sigma_\ell[m^n](X_p,t_p)|^2) + O\left(k\rho_n^2\right) \\
&\leq C h_n k \left(1 + \sup_{p=0,\dots,N} \mathbb{E}|X_p|^2\right) + O\left(k\rho_n^2\right).
\end{aligned} \tag{4.12}
$$

On the other hand, using Lemma 4.1 again,

$$
\left|\sum_{p=0}^{k-1} \mathbb{E}\left(\delta_p X|X_p\right)\right|^2 \leq C h_n^2 k \sum_{p=0}^{k-1}(1+|X_p|^2),
$$

and so

$$
\mathbb{E}\left|\sum_{p=0}^{k-1} \mathbb{E}\left(\delta_p X|X_p\right)\right|^2 \leq C h_n^2 k^2 \left(1 + \sup_{p=0,\dots,N} \mathbb{E}|X_p|^2\right). \tag{4.13}
$$

By Proposition 4.1, (4.12), (4.13), (4.11), and our assumption $\rho_n^2 = O(h_n)$, we get the existence of $C > 0$ such that (4.9) holds true.

Step 2: *Proof of* (4.8). Let $0 \leq s < t \leq T$ and k', k such that $s \in [t_{k'}, t_{k'+1}[$ and $t \in [t_k, t_{k+1}[$. Then, by the triangular inequality

$$
d_2(m^n(t), m^n(s)) \leq d_2(m^n(t), m^n(t_k)) + d_2(m^n(t_k), m^n(t_{k'+1})) + d_2(m^n(t_{k'+1}), m^n(s)). \tag{4.14}
$$

By the dual representation of $d_2^2(\cdot,\cdot)$ (see [61, Theorem 1.3]), this function is convex in $\mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d)$. Thus, relations (3.17) and (4.9) imply that

$$
d_2^2(m^n(t), m^n(t_k)) \leq \left(\frac{t-t_k}{h_n}\right) d_2^2(m^n(t_{k+1}), m^n(t_k)) \leq C^2(t-t_k),
$$

from which

$$
d_2(m^n(t), m^n(t_k)) \leq C(t-t_k)^{\frac{1}{2}}. \tag{4.15}
$$

Analogously,

$$
d_2(m^n(t_{k'+1}), m^n(s)) \leq C(t_{k'+1}-s)^{\frac{1}{2}}. \tag{4.16}
$$

Relations (4.14), (4.15), (4.16), and the Cauchy–Schwarz inequality imply the existence of $C > 0$, independent of n, such that

$$
d_2(m^n(t), m^n(s)) \leq C|t-s|^{\frac{1}{2}}.
$$

Relation (4.8) follows. □

For notational convenience, for all $\varphi \in C_0^\infty(\mathbb{R}^d)$ let us set

$$
\begin{aligned}
L_{b,\sigma,\varphi}[\mu](x,t) &:= \tfrac{1}{2} \sum_{i,j} a_{i,j}[\mu](x,t) \partial^2_{x_i,x_j} \varphi(x) + b[\mu](x,t) \cdot \nabla\varphi(x) \\
&\forall\, (\mu,x,t) \in C([0,T];\mathcal{P}_1(\mathbb{R}^d)) \times \mathbb{R}^d \times [0,T].
\end{aligned} \tag{4.17}
$$

We have now all the elements to prove our main convergence results. We consider first the case where, in addition to **(H)**, the coefficients satisfy the following local Lipschitz property.

(Lip) For any $\mu \in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ and a compact set $K \subseteq \mathbb{R}^d$, there exists a constant $C = C(\mu, K) > 0$ such that

$$|b[\mu](y,t) - b[\mu](x,t)| + |\sigma[\mu](y,t) - \sigma[\mu](x,t)| \leq C|y-x| \;\; \forall\, x,\; y \in K, \;\; t \in [0,T]. \tag{4.18}$$

The case of more general coefficients satisfying only **(H)** will be treated just after.

THEOREM 4.1. *Assume* (**H**)-(**Lip**) *and that* $\rho_n^2 = o(h_n)$. *Then, every limit point* $m \in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ *of* m^n *(there exists at least one) solves* (FPK). *In particular,* (FKP) *admits at least one solution.*

Proof. By Propositions 4.1 and 4.2 and Lemma 2.1, with $\mathcal{C} = \{m^n \; ; \; n \in \mathbb{N}\}$, the sequence m^n has at least one limit point m. We use the same superscript n to index a subsequence m^n converging to m in $C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ and we need to show that m satisfies (2.5). Let $t \in]0,T]$ and, without loss of generality, consider a sequence $t_{n'} = n'h_n$ such that $t \in]t_{n'}, t_{n'+1}]$. Then, for every $\varphi \in C_0^\infty(\mathbb{R}^d)$

$$\int_{\mathbb{R}^d} \varphi(x) \mathrm{d}m^n(t_{n'})(x) = \int_{\mathbb{R}^d} \varphi(x)\mathrm{d}m^n(0)(x) + \sum_{k=0}^{n'-1} \int_{\mathbb{R}^d} \varphi(x) \mathrm{d}\left[m^n(t_{k+1}) - m^n(t_k)\right](x). \tag{4.19}$$

For all $k = 0, \dots, n'-1$ we have that

$$\begin{aligned}
\textstyle\int_{\mathbb{R}^d} \varphi(x)\mathrm{d}m^n(t_{k+1})(x) &= \textstyle\sum_{i\in\mathbb{Z}^d} \varphi(x_i) m^n_{i,k+1} \\
&= \textstyle\sum_{j\in\mathbb{Z}^d} m^n_{j,k} \frac{1}{2r} \sum_{\ell=1}^r \sum_{i\in\mathbb{Z}^d} \varphi(x_i)\left[\beta_i(\Phi^{\ell,+}_{j,k}[m^n]) + \beta_i(\Phi^{\ell,-}_{j,k}[m^n])\right] \\
&= \textstyle\sum_{j\in\mathbb{Z}^d} m^n_{j,k} \frac{1}{2r} \sum_{\ell=1}^r \left[I[\varphi](\Phi^{\ell,+}_{j,k}[m^n]) + I[\varphi](\Phi^{\ell,-}_{j,k}[m^n])\right] \\
&= \textstyle\sum_{j\in\mathbb{Z}^d} m^n_{j,k} \frac{1}{2r} \sum_{\ell=1}^r \left[\varphi(\Phi^{\ell,+}_{j,k}[m^n]) + \varphi(\Phi^{\ell,-}_{j,k}[m^n])\right] + O(\rho_n^2) \\
&= \textstyle\sum_{j\in\mathbb{Z}^d} m^n_{j,k} \left[\varphi(x_j) + h_n L_{b,\sigma,\varphi}[m^n](x_j,t_k)\right] + O\left(\rho_n^2 + h_n^2\right) \\
&= \textstyle\int_{\mathbb{R}^d} \left[\varphi(x) + h_n L_{b,\sigma,\varphi}[m^n](x,t_k)\right] \mathrm{d}m^n(t_k)(x) + O\left(\rho_n^2 + h_n^2\right),
\end{aligned} \tag{4.20}$$

where we have used a fourth order Taylor expansion for the terms $\varphi(\Phi^{\ell,+}_{j,k}[m^n])$ and $\varphi(\Phi^{\ell,-}_{j,k}[m^n])$. As a consequence, (4.19) yields

$$\begin{aligned}
\int_{\mathbb{R}^d} \varphi(x)\mathrm{d}m^n(t_{n'})(x) &= \int_{\mathbb{R}^d} \varphi(x)\mathrm{d}m^n(0)(x) \\
&+ h_n \sum_{k=0}^{n'-1} \int_{\mathbb{R}^d} L_{b,\sigma,\varphi}[m^n](x,t_k)\mathrm{d}m^n(t_k)(x) + O\left(\frac{\rho_n^2}{h_n} + h_n\right).
\end{aligned} \tag{4.21}$$

Assumption (**H**)(i) implies the existence of a modulus of continuity $\bar{\omega}_1$, independent of k, such that

$$\int_{\mathbb{R}^d} L_{b,\sigma,\varphi}[m^n](x,t_k)\mathrm{d}m^n(t_k)(x) = \int_{\mathbb{R}^d} L_{b,\sigma,\varphi}[m](x,t_k)\mathrm{d}m^n(t_k)(x) + \bar{\omega}_1\left(\sup_{t\in[0,T]} d_1(m^n(t), m(t))\right). \tag{4.22}$$

Since ϕ has a compact support, condition **(Lip)** implies that $L_{b,\sigma,\varphi}[m](\cdot,t_k)$ is Lipschitz, uniformly in k. Thus, by (2.2) and (4.8), we have

(4.23)
$$\left|\int_{\mathbb{R}^d} L_{b,\sigma,\varphi}[m](x,t_k)\mathrm{d}\left(m^n(s)-m^n(t_k)\right)(x)\right| \leq Cd_1(m^n(s),m^n(t_k)) \leq C'\sqrt{h_n} \quad \forall\, s\in[t_k,t_{k+1}),$$

for some positive constants C and C', independent of n. This implies that

$$\left|h_n\int_{\mathbb{R}^d} L_{b,\sigma,\varphi}[m](x,t_k)\mathrm{d}m^n(t_k)(x) - \int_{t_k}^{t_{k+1}}\int_{\mathbb{R}^d} L_{b,\sigma,\varphi}[m](x,t_k)\mathrm{d}m^n(s)(x)\right| = O\left(h_n^{\frac{3}{2}}\right).$$

Therefore, by (4.21),

(4.24)
$$\begin{aligned}\int_{\mathbb{R}^d}\varphi(x)\mathrm{d}m^n(t_{n'})(x) = \quad & \int_{\mathbb{R}^d}\varphi(x)\mathrm{d}m^n(0)(x) + \int_0^{t_{n'}}\int_{\mathbb{R}^d}\hat{L}^n_{b,\sigma,\varphi}[m](x,s)\mathrm{d}m^n(s)(x)\mathrm{d}s\\ & +O\left(\frac{\rho_n^2}{h_n}+\sqrt{h_n}+\bar{\omega}_1\left(\sup_{t\in[0,T]} d_1(m^n(t),m(t))\right)\right),\end{aligned}$$

where

$$\hat{L}^n_{b,\sigma,\varphi}[m](x,s) := L_{b,\sigma,\varphi}[m](x,t_k) \quad \forall\, x\in\mathbb{R}^d,\ \ s\in[t_k,t_{k+1}).$$

By **(H)**(i), and the fact that ϕ has compact support, we have that $\hat{L}^n_{b,\sigma,\varphi}[m](\cdot,\cdot)$ is bounded, uniformly in n, and converges uniformly to $L_{b,\sigma,\varphi}[m](\cdot,\cdot)$ in $\mathbb{R}^d\times[0,T]$. As a consequence, for each $s\in[0,T]$, we have that $\int_{\mathbb{R}^d}\hat{L}_{b,\sigma,\varphi}[m](x,s)\mathrm{d}m^n(s)(x)$ is uniformly bounded and converges, as $n\to\infty$, to $\int_{\mathbb{R}^d}L_{b,\sigma,\varphi}[m](x,s)\mathrm{d}m(s)(x)$. Therefore, by Lebesgue's dominated convergence theorem, the second term on the right-hand side of (4.24) converges to

$$\int_0^t\int_{\mathbb{R}^d} L_{b,\sigma,\varphi}[m](x,s)\mathrm{d}m(s)(x)\mathrm{d}s.$$

Finally, passing to the limit in (4.24), we get that (2.5) holds true. ∎

In the remainder of this section, we consider the case where b and σ satisfy only assumption **(H)**. Since in the proof of Theorem 4.1 the local Lipchitz assumption **(Lip)** plays an important role, in the present case we need to regularize the coefficients, which will be done by convolution with a mollifier. Let $\phi\in C^\infty(\mathbb{R}^d)$ have a compact support contained in the closed unit ball $B(0,1):=\{x\in\mathbb{R}^d\ ;\ |x|\leq 1\}$, and, given a sequence ε_n, with $0<\varepsilon_n\leq 1$, set $\phi_{\varepsilon_n}(x):=\phi(x/\varepsilon_n)/(\varepsilon_n)^d$ for all $x\in\mathbb{R}^d$. Let us define

$$b_n[\mu](x,t) := \left(\phi_{\varepsilon_n} * b[\mu]\right)(x,t) \quad \text{and} \quad \sigma_n[\mu](x,t) := \left(\phi_{\varepsilon_n} * \sigma[\mu]\right)(x,t), \tag{4.25}$$

where the convolution is applied in the space variable x and componentwise for the coordinates of $b[\mu](\cdot,t)$ and $\sigma[\mu](\cdot,t)$. It is easy to check that for each $\mu\in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ and each compact set $K\subseteq\mathbb{R}^d$, we have that b_n and σ_n satisfy (4.18) with $C_K=C'_K/\varepsilon_n$, where C'_K depends only on ϕ and on

$$\sup\left\{|b[\mu](x,t)|+|\sigma[\mu](x,t)|\ |\ x\in K+B(0,1),\ t\in[0,T]\right\}<\infty.$$

We consider the approximation (3.22) of (FPK) with $\Phi^{\ell,+}_{i,k}[\mu]$ and $\Phi^{\ell,-}_{i,k}[\mu]$ replaced by

$$\begin{aligned}\Phi^{n,\ell,+}_{i,k}[\mu] &:= x_i + h_n b_n[\mu](x_i,t_k) + \sqrt{rh_n}(\sigma_n)_\ell[\mu](x_i,t_k),\\ \Phi^{n,\ell,-}_{i,k}[\mu] &:= x_i + h_n b_n[\mu](x_i,t_k) - \sqrt{rh_n}(\sigma_n)_\ell[\mu](x_i,t_k),\end{aligned}$$

respectively. Namely, find $m \in \mathcal{S}^{\rho_n,h_n}$ such that

(4.26)
$$
\begin{aligned}
&m_{i,0} = \bar{m}_0(E_i) \quad \forall\, i \in \mathbb{Z}^d,\\
&m_{i,k+1} = \tfrac{1}{2r}\sum_{\ell=1}^{r}\sum_{j\in\mathbb{Z}^d}\left[\beta_i(\Phi_{j,k}^{n,\ell,+}[m]) + \beta_i(\Phi_{j,k}^{n,\ell,-}[m])\right] m_{j,k} \quad \forall\, i \in \mathbb{Z}^d,\ \ k = 0,\dots,N-1.
\end{aligned}
$$

The coefficients b_n and σ_n satisfy **(H)** and the linear growth condition (2.6) holds with a constant C independent of n. As a consequence, for each $n \in \mathbb{N}$, problem (3.22) admits at least one solution m^n and, denoting likewise the extension of m^n in (3.17) to an element in $C([0,T];\mathcal{P}_1(\mathbb{R}^d))$, by (4.6) and (4.8), whose proofs can be reproduced without modifications and with constants independent of n, the set $\{m^n \mid n \in \mathbb{N}\}$ is relatively compact in $C([0,T];\mathcal{P}_1(\mathbb{R}^d))$.

We have the following convergence result, assuming only **(H)** and whose proof is almost identical to the previous one.

THEOREM 4.2. *Assume* (**H**) *and that* $\rho_n^2 = o(h_n)$ *and* $h_n = o(\varepsilon_n^2)$. *Then, every limit point* $m \in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ *of* m^n *(there exists at least one) solves* (FPK). *In particular,* (FKP) *admits at least one solution.*

Proof. Arguing exactly as in the proof of Theorem 4.1, and using the same notation, we have the existence of $m \in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ such that, up to some subsequence, $m^n \to m$ in $C([0,T];\mathcal{P}_1(\mathbb{R}^d))$. Moreover, for each $n \in \mathbb{N}$ we have

$$
\begin{aligned}
\int_{\mathbb{R}^d}\varphi(x)\mathrm{d}m^n(t_{n'})(x) &= \int_{\mathbb{R}^d}\varphi(x)\mathrm{d}m^n(0)(x)\\
&\quad + h_n\sum_{k=0}^{n'-1}\int_{\mathbb{R}^d} L_{b_n,\sigma_n,\varphi}[m^n](x,t_k)\mathrm{d}m^n(t_k)(x) + O\left(\frac{\rho_n^2}{h_n} + h_n\right),
\end{aligned} \tag{4.27}
$$

where $L_{b_n,\sigma_n,\varphi}$ is given by (4.17), with b and σ replaced by b_n and σ_n, respectively. Estimate (4.22) still holds and (4.23) changes to

$$
\begin{aligned}
\left|\int_{\mathbb{R}^d} L_{b_n,\sigma_n,\varphi}[m](x,t_k)\mathrm{d}\left(m^n(s) - m^n(t_k)\right)(x)\right| &\le \frac{C}{\varepsilon_n} d_1(m^n(s), m^n(t_k)) \le C'\frac{\sqrt{h_n}}{\varepsilon_n}\\
&\quad \forall\, s \in [t_k, t_{k+1}),
\end{aligned} \tag{4.28}
$$

for some constants C and C' independent of n. Relation (4.27) then gives

(4.29)
$$
\begin{aligned}
\textstyle\int_{\mathbb{R}^d}\varphi(x)\mathrm{d}m^n(t_{n'})(x) = \ &\textstyle\int_{\mathbb{R}^d}\varphi(x)\mathrm{d}m^n(0)(x) + \int_0^{t_{n'}}\int_{\mathbb{R}^d}\hat{L}_{b_n,\sigma_n,\varphi}[m](x,s)\mathrm{d}m^n(s)(x)\mathrm{d}s\\
&+O\left(\tfrac{\rho_n^2}{h_n} + \tfrac{\sqrt{h_n}}{\varepsilon_n} + \bar{\omega}_1\left(\sup_{t\in[0,T]} d_1(m^n(t), m(t))\right)\right),
\end{aligned}
$$

where $\hat{L}_{b_n,\sigma_n,\varphi}[m](x,s) := L_{b_n,\sigma_n,\varphi}[m](x,t_k)$ for all $x \in \mathbb{R}^d$, and $s \in [t_k, t_{k+1})$. By **(H)**(i) we have that $\hat{L}_{b_n,\sigma_n,\varphi}[m](\cdot,\cdot) \to L_{b,\sigma,\varphi}[m](\cdot,\cdot)$ uniformly in $\mathbb{R}^d \times [0,T]$ and, passing to the limit in (4.29), we can conclude as in the previous proof. ∎

Remark 4.1. In particular, Theorem 4.2 yields a Peano-type existence result for (FPK). We point out that more general existence results for the (FPK) equation are proven in the articles [49, 50] by using purely analytical techniques.

Remark 4.2. (i) In the deterministic case $\sigma \equiv 0$, the proof in [21, Proposition 3.9] shows that (4.8) can be replaced by

$$
\sup_{n\in\mathbb{N}} d_1(m^n(t), m^n(s)) \le C|t-s| \quad \forall\, t,s \in [0,T],
$$

and, hence, the estimate (4.28) can be improved to

$$\left|\int_{\mathbb{R}^d} L_{b_n,\sigma_n,\varphi}[m](x,t_k)\mathrm{d}\left(m^n(s)-m^n(t_k)\right)(x)\right| \leq \frac{C}{\varepsilon_n} d_1(m^n(s),m^n(t_k)) \leq C'\frac{h_n}{\varepsilon_n} \quad \forall\, s\in[t_k,t_{k+1}),$$

for some constants C and C' independent of n. As a consequence, the result in Theorem 4.2 holds true under the weaker assumption $h_n = o(\varepsilon_n)$.

(ii) The regularization of the coefficients, defined in (4.25), can also be useful in order to approximate the (FPK) equation with coefficients b and σ defined almost everywhere w.r.t. the Lebesgue measure. In this case, in order to give a meaning to a solution m of (2.5) one can require that $m(t)$ should be absolutely continuous w.r.t. the Lebesgue measure for almost every $t\in[0,T]$. One can then consider coefficients b^n and σ^n which regularize b and σ, but in general we can only expect L^1 convergence $L_{b^n,\sigma^n,\varphi}$ to $L_{b,\sigma,\varphi}$. In this case, the scheme (3.21) should be modified in order to discretize the density of m and a stronger compactness result, for example in L^∞ endowed with the weak* topology, should be proved for the constructed approximation m^n. As we will discuss in Remark 5.1(ii), this is exactly the situation in degenerate MFGs (see [21, 22]).

5. Applications and numerical simulations.

The scheme that we have proposed in section 3 has already been tested for the case of some linear FPK equations in [23]. In this section, we apply our scheme to solve two nonlinear models with $\sigma[m](x,t)\equiv\sigma I_d$ for some $\sigma\neq 0$ (where I_d is the $d\times d$ identity matrix), but where $b[m](x,t)$ does not admit an explicit expression and has to be approximated. The approximation technique is similar to the one presented at the end of the previous section, where the coefficients are supposed to satisfy **(H)** only. In the first model we consider an example of the so-called MFG system with nonlocal interactions (see [45]). In this case, the drift $b[m](x,t)$ is related to the value function of an optimal control problem starting at x at time t, having running and terminal costs depending on $\{m(s)\,;s\in]0,T[\}$ and $m(T)$, respectively. Therefore, as explained in Remark 3.5, the proposed scheme is implicit. Our approximation is similar to the one in [21, 20, 22] dealing with degenerate MFG systems and where the authors prove the convergence when the state dimension d is equal to one. In our present nondegenerate setting, the theory developed in section 4 allows us to prove the convergence of the scheme in general space dimensions. In the second nonlinear model, we consider an FPK equation where the velocity field $b[m](x,t)$ depends on the value function of an optimal control starting at x at time t with running and terminal costs depending only on the value $m(t)$. This model, which seems to be new, is inspired by the Hughes model [37] and could be used to model crowd motion in some "panic" situations. We prove that the related FPK equation admits at least one solution, and we also provide a convergence result for the associated scheme.

5.1. Mean Field Games as a nonlinear implicit model.

We consider here the MFG system

$$\begin{aligned} -\partial_t v - \tfrac{\sigma^2}{2}\Delta v + \tfrac{1}{2}|\nabla v|^2 &= F(x,m(t)) \text{ in } \mathbb{R}^d\times(0,T),\\ \partial_t m - \tfrac{\sigma^2}{2}\Delta m - \operatorname{div}\big(\nabla v m\big) &= 0 \text{ in } \mathbb{R}^d\times(0,T),\\ v(x,T) = G(x,m(t)) \text{ for } x\in\mathbb{R}^d, &\qquad m(0)=\bar{m}_0(\cdot)\in\mathcal{P}_2(\mathbb{R}^d), \end{aligned} \tag{5.1}$$

where $\sigma \neq 0$ and F, $G : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \to \mathbb{R}$ are continuous, twice differentiable w.r.t. the space variable, and satisfy that there exists a constant $c > 0$ such that for $\psi = F, G$

$$\sup_{x\in\mathbb{R}^d,\mu\in\mathcal{P}_1(\mathbb{R}^d)} \big(|\psi(x,\mu)| + |\nabla_x\psi(x,\mu)| + |\nabla^2_{xx}\psi(x,\mu)|\big) \leq c. \tag{5.2}$$

System (5.1) is a particular instance of a generic class of models introduced by Lasry and Lions in [43, 44, 45] which characterize Nash equilibria of symmetric stochastic differential games with an infinite number of players. In order to explain the intuition behind (5.1), for $m \in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ consider the HJB equation

$$\begin{array}{rcl} -\partial_t v - \frac{\sigma^2}{2}\Delta v + \frac{1}{2}|\nabla v|^2 & = & F(x, m(t)) \ \text{ in } \mathbb{R}^d \times (0,T), \\ v(x,T) & = & G(x, m(T)) \ \text{ for } x \in \mathbb{R}^d. \end{array} \tag{5.3}$$

Standard results in stochastic control (see, e.g., [33]) imply that the unique solution $v[m]$ of (5.3) can be represented as

$$v[m](x,t) := \inf_{\alpha} \mathbb{E}\left(\int_t^T \left[\tfrac{1}{2}|\alpha(s)|^2 + F(X^{x,t,\alpha}(s), m(s))\right] \mathrm{d}s + G(X^{x,t,\alpha}(T), m(T))\right), \tag{5.4}$$

where the expectation $\mathbb{E}$ is taken in a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which an r-dimensional Brownian motion W is defined, the $\mathbb{R}^d$-valued processes α are adapted to the natural filtration generated by W, completed with the $\mathbb{P}$-null sets, and they satisfy $\mathbb{E}(\int_0^T |\alpha(t)|^2 \mathrm{d}t) < \infty$, and $X^{x,t,\alpha}$ is defined as the solution of

$$\mathrm{d}X(s) = \alpha(s)\mathrm{d}s + \sigma \mathrm{d}W(s) \ \ s \in (t,T), \qquad X(t) = x. \tag{5.5}$$

The optimization problem in (5.4) can be interpreted in terms of a generic small agent whose state is x at time t and optimizes a cost depending on the future distribution of the agents $\{m(s) \ ; \ s \in]t,T]\}$. The solution $v[m]$ of (5.3) is classical (see, e.g., [16], where the proof is based upon the Hopf–Cole transformation). Note that (5.4), assumption (5.2), and standard estimates for the solutions of the controlled SDE (5.5) imply that $C([0,T];\mathcal{P}_1(\mathbb{R}^d)) \times \mathbb{R}^d \times [0,T] \ni (\mu,x,t) \mapsto v[\mu](x,t) \in \mathbb{R}$ is bounded and continuous. In addition, we have that

$$\sup_{t\in[0,T],\ \mu\in C([0,T];\mathcal{P}_1(\mathbb{R}^d))} |\nabla_x v[\mu](\cdot,t)|_\infty < \infty, \tag{5.6}$$

and $\mathbb{R}^d \ni x \mapsto \nabla_x v[\mu](x,t) \in \mathbb{R}^d$ is locally Lipschitz with local Lipschitz constants independent of μ and t. Thus, for every $(x,t) \in \mathbb{R}^d \times [0,T)$ the equation

$$\mathrm{d}X(s) = -\nabla_x v[m]\,(X(s),s)\,\mathrm{d}s + \sigma \mathrm{d}W(s) \ \ s \in (t,T), \qquad X(t) = x, \tag{5.7}$$

admits a unique solution $X^{x,t}$ and, hence, by a verification argument (see, e.g., [33]), the optimal trajectory for $v[m](x,t)$ in (5.4) is given by $X^{x,t}$ and the optimal control α is given in the feedback form $\alpha(x,t) = -\nabla_x v[m]\,(x,t)$. Therefore, if all the players, distributed as m_0 at time 0, act optimally according to this feedback law, then the evolution of m_0 will be described by the FPK equation

$$\partial_t\mu - \frac{\sigma^2}{2}\Delta\mu - \mathrm{div}\big(\nabla v[m]\mu\big) = 0 \quad \text{in } \ \mathbb{R}^d \times (0,T), \quad \mu(0) = m_0,$$

and the equilibrium condition reads $m = \mu$, i.e.,

$$\partial_t m - \frac{\sigma^2}{2}\Delta m - \operatorname{div}\big(\nabla v[m]m\big) = 0 \quad \text{in } \mathbb{R}^d \times (0,T), \quad m(0) = m_0. \tag{5.8}$$

The equilibrium equation (5.8) is a particular instance of (FKP) with $r = d$, $\sigma_{ij} = \sigma$ if $i = j$ and 0 otherwise, and

$$b[\mu](x,t) := -\nabla v[\mu](x,t) \qquad \forall\, (\mu,x,t) \in C([0,T];\mathcal{P}_1(\mathbb{R}^d)) \times \mathbb{R}^d \times [0,T], \tag{5.9}$$

which depends on μ nonlocally in time through $\{\mu(s)\ ;\ s \in (t,T]\}$ by (5.4) (with m replaced by μ).

Notice that (5.6) implies that **(H)**(ii) is satisfied. In order to check **(H)**(i), recall that v is uniformly semiconcave w.r.t. the space variable (see, e.g., [15] and [33, Chapter 4]), i.e., there exists $c > 0$, independent of $t \in [0,T]$ and $\mu \in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$, such that for all $x \in \mathbb{R}^d$, $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, and $t \in [0,T]$,

$$v[\mu](x+h,t) - 2v[\mu](x,t) + v[\mu](x-h,t) \le c|h|^2 \qquad \forall\, h \in \mathbb{R}^d, \tag{5.10}$$

or equivalently, since $v[\mu](\cdot,t)$ is differentiable, there exists a constant $c > 0$, independent of $t \in [0,T]$ and $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, such that

$$v[\mu](x+h,t) \le v[\mu](x,t) + \nabla_x v[\mu](x,t)\cdot h + c|h|^2 \quad \forall\, h \in \mathbb{R}^d,\ t \in [0,T]. \tag{5.11}$$

As a consequence, the continuity of v yields that for any $(\mu_n, x_n, t_n) \to (\mu, x, t)$ we have that any limit point p of $\nabla_x v[\mu_n](x_n,t_n)$ (there exists at least one by (5.6)) must satisfy

$$v[\mu](x+h,t) \le v[\mu](x,t) + p\cdot h + c|h|^2 \quad \forall\, h \in \mathbb{R}^d,\ t \in [0,T],$$

and, hence, $p = \nabla_x v[\mu](x,t)$ by [15, Propositions 3.3.1 and 3.1.5(c)]. Therefore, b, defined in (5.9), is continuous and, hence, **(H)**(i) holds true.

By the previous remarks, the results of sections 3 and 4 are applicable to (5.7). However, from the numerical point of view, we cannot implement the fully discrete scheme directly using b, because we do not have an explicit expression for this vector field, which depends on the value function v. To overcome this difficulty, we approximate b by a sequence of computable vector fields. We consider an SL scheme for the solution to (5.3), with m replaced by μ. Given $\rho > 0$, $h = T/N > 0$, with $N \in \mathbb{N}$, and $\mu \in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ we first define $v^{\rho,h}[\mu]$ in $\mathcal{G}_\rho \times \{0,\dots,N\}$ recursively as

(5.12)
$$\begin{aligned} v^{\rho,h}_{i,k} &= \inf_{\alpha\in\mathbb{R}^d}\left\{\tfrac{h}{2}|\alpha|^2 + \tfrac{1}{2d}\textstyle\sum_{\ell=1}^d \left(I[v^{\rho,h}_{\cdot,k+1}](x_i + h\alpha + \sigma\sqrt{hd}e_\ell) + I[v^{\rho,h}_{\cdot,k+1}](x_i + h\alpha - \sigma\sqrt{hd}e_\ell)\right)\right\} \\ &\quad + hF(x_i,\mu(t_k)) \qquad \forall\, i \in \mathbb{Z}^d,\ \ \forall\, k = 0,\dots,N-1, \\ v^{\rho,h}_{i,N} &= G(x_i,\mu(T)) \qquad \forall\, i \in \mathbb{Z}^d, \end{aligned}$$

where $\{e_\ell\ ;\ \ell = 1,\dots,d\}$ is the canonical basis of $\mathbb{R}^d$, and we have omitted the μ dependence of $v^{\rho,h}$. We then define $v^{\rho,h} : C([0,T];\mathcal{P}_1(\mathbb{R}^d)) \times \mathbb{R}^d \times [0,T] \to \mathbb{R}$ by

$$v^{\rho,h}[\mu](x,t) = I[v^{\rho,h}_{\cdot,k}[\mu]](x) \quad \text{ if } t \in [t_k, t_{k+1}[.$$

In order to get a function differentiable w.r.t. the space variable, given $\varepsilon > 0$ and $\phi \in C^\infty(\mathbb{R}^d)$, nonnegative and such that $\int_{\mathbb{R}^d}\phi(x)\mathrm{d}x = 1$, let us set $\phi_\varepsilon(x) := \frac{1}{\varepsilon^d}\phi(x/\varepsilon)$. We then define $v^{\rho,h,\varepsilon} : C([0,T];\mathcal{P}_1(\mathbb{R}^d)) \times \mathbb{R}^d \times [0,T] \to \mathbb{R}$ by

$$v^{\rho,h,\varepsilon}[\mu](\cdot,t) := \left(\phi_\varepsilon * v^{\rho,h}[\mu]\right)(\cdot,t) \quad \forall\, t \in [0,T].$$

In [22, Lemma 3.2 (i)] it is shown that $v^{\rho,h,\varepsilon}[\mu](\cdot,t)$ is Lipschitz, uniformly in $(\rho,h,\varepsilon,\mu,t)$, which shows the bound (5.6) for $v^{\rho,h,\varepsilon}$. Using that $v^{\rho,h}$ satisfies a discrete semiconcavity property (see [22, Lemma 3.1 (ii)]), by [1, Lemma 4.3 and Remark 4.4] there exists a constant $c>0$, independent of $(\rho,h,\varepsilon,\mu,t)$, such that $v^{\rho,h,\varepsilon}[\mu](\cdot,t)$ satisfies the following weak semiconcavity property:

$$\left(\nabla_x v^{\rho,h,\varepsilon}[\mu](y,t)-\nabla_x v^{\rho,h,\varepsilon}[\mu](x,t)\right)\cdot(y-x)\leq c\left(|y-x|^2+\frac{\rho^2}{\varepsilon^2}\right). \tag{5.13}$$

Using the previous ingredients, we can prove the following result.

PROPOSITION 5.1. *Consider sequences ρ_n, h_n, and ε_n of positive numbers converging to* 0 *and such that $\frac{\rho_n^2}{h_n}\to 0$ and $\rho_n=o(\varepsilon_n)$. Then, for every sequence $\mu_n\in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ converging to μ we have that $v^{\rho_n,h_n,\varepsilon_n}[\mu_n]$ and $\nabla_x v^{\rho_n,h_n,\varepsilon_n}[\mu_n]$ converge to $v[\mu]$ and $\nabla_x v[\mu]$, respectively, uniformly over compact subsets of $\mathbb{R}^d\times[0,T]$.*

Proof. The assertion on the convergence of $v^{\rho_n,h_n,\varepsilon_n}[\mu_n]$ is a consequence of the uniform convergence over compact sets of $v^{\rho_n,h_n}[\mu_n]$ to $v[\mu]$ if $\frac{\rho_n^2}{h_n}\to 0$, which is a standard result proved with the theory developed in [8] (see, e.g., [27, Theorem 4.2]). The argument to establish the uniform convergence of $\nabla_x v^{\rho_n,h_n,\varepsilon_n}[\mu_n]$ is similar to the proof of [21, Theorem 3.5]. Namely, for all $n\in\mathbb{N}$, $x_n\to x$, $t_n\to t$, and $y\neq x$ we have (for n large enough)

$$v^{\rho_n,h_n,\varepsilon_n}[\mu_n](y,t_n)-v^{\rho_n,h_n,\varepsilon_n}[\mu_n](x_n,t_n)-\nabla_x v^{\rho_n,h_n,\varepsilon_n}[\mu_n](x_n,t_n)\cdot(y-x_n)\leq r_{1n}+r_{2n},$$

where

$$r_{1,n}:=\int_0^{\frac{\rho_n}{\varepsilon_n|y-x_n|}}\left[\nabla_x v^{\rho_n,h_n,\varepsilon_n}[\mu_n](x_n+\tau(y-x_n),t_n)-\nabla_x v^{\rho_n,h_n,\varepsilon_n}[\mu_n](x_n,t_n)\right]\cdot(y-x_n)\mathrm{d}\tau,$$
$$r_{2,n}:=\int_{\frac{\rho_n}{\varepsilon_n|y-x_n|}}^1\left[\nabla_x v^{\rho_n,h_n,\varepsilon_n}[\mu_n](x_n+\tau(y-x_n),t_n)-\nabla_x v^{\rho_n,h_n,\varepsilon_n}[\mu_n](x_n,t_n)\right]\cdot(y-x_n)\mathrm{d}\tau.$$

Since $\frac{\rho_n}{\varepsilon_n}\to 0$, the uniform Lipschitz property satisfied by $v^{\rho_n,h_n,\varepsilon_n}[\mu_n](\cdot,t)$, for $t\in[0,T]$, implies that $r_{1,n}\to 0$. On the other hand, by (5.13),

$$r_{2,n}\leq\int_{\frac{\rho_n}{\varepsilon_n|y-x_n|}}^1\frac{c}{\tau}\left(\tau^2|y-x_n|^2+\left(\frac{\rho_n}{\varepsilon_n}\right)^2\right)\mathrm{d}\tau\leq 2c|y-x_n|^2\int_0^1\tau\mathrm{d}\tau=c|y-x_n|^2.$$

By the uniform convergence of $v^{\rho_n,h_n,\varepsilon_n}[\mu_n]$, we conclude that any limit point p of $\nabla_x v^{\rho_n,h_n,\varepsilon_n}[\mu_n](x_n,t_n)$ (there exists at least one because this sequence is uniformly bounded) must satisfy

$$v[\mu](y,t)\leq v[\mu](x,t)+p\cdot(y-x)+c|y-x|^2\quad\forall\, y\in\mathbb{R}^d,\ t\in[0,T],$$

which implies that $p=\nabla_x v[\mu](x,t)$ by [15, Propositions 3.3.1 and 3.1.5(c)]. Thus, if for all $i=1,\dots,d$ we denote

$$b_i^{\sup}:=\limsup_{x'\to x,t'\to t,n\to\infty}\partial_{x_i}v^{\rho_n,h_n,\varepsilon_n}[\mu_n](x',t'),\ b_i^{\inf}:=\liminf_{x'\to x,t'\to t,n\to\infty}\partial_{x_i}v^{\rho_n,h_n,\varepsilon_n}[\mu_n](x',t'),$$

we deduce that $b_i^{\sup}=b_i^{\inf}=\partial_{x_i}v[\mu](x,t)$, and so the local uniform convergence of $\nabla_x v^{\rho_n,h_n,\varepsilon_n}[\mu_n](\cdot,\cdot)$ to $\nabla_x v[\mu](\cdot,\cdot)$ follows (see, e.g., [7, Chapter V, Lemma 1.9]). ☐

Suppose that ρ_n, h_n, and ε_n satisfy the conditions in Proposition 5.1, and denote by $m^n\in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ the extension to $C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ of the solution of (3.22) computed with coefficients $b^n[\mu](x,t):=\nabla_x v^{\rho_n,h_n,\varepsilon_n}[\mu](x,t)$ and $\sigma_\ell^n=\sigma e_\ell$ ($\ell=1,\dots,d$).

PROPOSITION 5.2. *Assume that $\rho_n^2 = o(h_n)$ and that $\rho_n = o(\varepsilon_n)$. Then, every limit point $m \in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ of m^n (there exists at least one) solves* (5.8).

Proof. The proof is analogous to the proof of Theorem 4.1. Indeed, arguing exactly as in the proof of that theorem, the sequence m^n has at least one limit point $m \in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ and for every $\varphi \in C_0^\infty(\mathbb{R}^d)$ the following equality holds:

$$\begin{aligned}\int_{\mathbb{R}^d}\varphi(x)\mathrm{d}m^n(t_{n'})(x) &= \int_{\mathbb{R}^d}\varphi(x)\mathrm{d}m^n(0)(x)\\ &+h_n\sum_{k=0}^{n'-1}\int_{\mathbb{R}^d}L_{b^n,\sigma,\varphi}[m^n](x,t_k)\mathrm{d}m^n(t_k)(x) + O\left(\frac{\rho_n^2}{h_n}+h_n\right),\end{aligned}\tag{5.14}$$

where we recall that $L_{b^n,\sigma,\varphi}[m^n]$ is given by (4.17). Denoting still by m^n a subsequence of m^n converging to m, Proposition 5.1 implies that $b^n[m^n] \to b[m] := \nabla_x v[m]$ uniformly over compact subsets of $\mathbb{R}^d \times [0,T]$. Therefore,

$$\int_{\mathbb{R}^d} L_{b^n,\sigma,\varphi}[m^n](x,t_k)\mathrm{d}m^n(t_k)(x) = \int_{\mathbb{R}^d} L_{b,\sigma,\varphi}[m](x,t_k)\mathrm{d}m^n(t_k)(x) + O(\delta_n),\tag{5.15}$$

where

$$\delta_n := \sup\left\{|L_{b^n,\sigma,\varphi}[m^n](x,s) - L_{b,\sigma,\varphi}[m](x,s)| \mid x \in \operatorname{supp}(\varphi),\ \ s\in[0,T]\right\}$$

tends to 0 as $n \to \infty$. Using (5.15) and that b is locally Lipschitz w.r.t. x, we can complete the proof by following the same steps as those in the proof of Theorem 4.1. ∎

Remark 5.1. (i) If F and G satisfy the monotonicity conditions

$$\int_{\mathbb{R}^d}[F(x,m_1)-F(x,m_2]\,\mathrm{d}(m_1-m_2)(x) > 0 \quad \forall\, m_1,m_2 \in \mathcal{P}_1(\mathbb{R}^d),\ m_1\neq m_2,$$
$$\int_{\mathbb{R}^d}[G(x,m_1)-G(x,m_2]\,\mathrm{d}(m_1-m_2)(x) \geq 0 \quad \forall\, m_1,m_2 \in \mathcal{P}_1(\mathbb{R}^d),$$

then system (5.1) admits a unique solution (v,m) (see [45]). In this case the entire sequence m^n in Proposition 5.2 converges to m.

(ii) In the articles [21, 22] a very similar scheme is proposed for degenerate MFG systems when m_0 is absolutely continuous, with a compact support and with an essentially bounded density. In those frameworks, the velocity field $b[\mu](x,t)$ is only defined for a.e. $x \in \mathbb{R}^d$. Therefore (see Remark 4.2(ii)), the proposed scheme discretizes the density of m for which an L^∞ bound is proved if $d = 1$. Moreover, the authors show the L^1 convergence of the approximations of the velocity field, which is weaker than the result in Proposition 5.1. On the other hand, when $d = 1$, uniform bounds in L^∞ are shown for the approximated densities, which allows them to prove, in these degenerate cases, a version of Proposition 5.2 in the one dimensional case. In their entire analysis, the extra assumptions on m_0 play an important role.

(iii) The introduction of the additional parameter $\varepsilon_n > 0$ in the fully discrete scheme, and the corresponding assumption $\rho_n = o(\varepsilon_n)$, is a disadvantage of this discretization compared with the finite difference discretization in [2], where only the space and time step parameters are considered.

5.1.1. Numerical test. We consider the MFG system (5.1) in dimension $d = r = 1$ on the space-time domain $\mathcal{O}\times[0,T] := [-3,3]\times[0,5]$, $\sigma = 0.01$, and with running and terminal costs given respectively by

$$F(x,m) := d(x,\mathcal{P})^2 V_\delta(x,m), \quad G(x,m) := F(x,m),\tag{5.16}$$

where

$$V_\delta(x,m) := (\phi_\delta * (\phi_\delta * m))(x) \quad \text{with} \quad \phi_\delta(x) := \tfrac{1}{\delta\sqrt{2\pi}} e^{\frac{-x^2}{2\delta^2}},$$

and $d(\cdot,\mathcal{P})$ denotes the distance to the set $\mathcal{P} := [-2,-2.5] \cup [1,1.5]$. We choose as initial distribution

$$\bar{m}_0(x) = \frac{\nu(x)}{\int_{\mathcal{O}} \nu(y)\mathrm{d}y} \mathbb{I}_{\mathcal{O}}(x) \quad \text{with} \quad \nu(x) := e^{-x^2/0.2}.$$

Since we consider a bounded space domain, we complement the FPK equation (5.8) with a homogeneous Neumann boundary condition, which, in terms of the underlying characteristics, means that trajectories are reflected once they touch the boundary. As a consequence, the total mass is preserved during the evolution. Accordingly, at the level of the fully discrete scheme we reflect the discrete characteristics. This modification of the scheme is discussed in [19] in the context of the Hughes model for pedestrian flow (see [37]). Let us point out that a theoretical study of the convergence of the resulting scheme has not yet been established and remains as an interesting subject of future research.

By formula (5.4) the interpretation in this setting is that agents want to reach the meeting areas, defined by the set $\mathcal{P}$, without spending too much effort (modeled by the $|\alpha|^2$ term in (5.4)), and to avoid congestion, modeled by the coupling terms F and G. Once the players reach the meeting areas they have no incentive to leave, and they remain in $\mathcal{P}$.

We heuristically solve the implicit scheme (3.22) using the learning procedure proposed in [17] (analyzed at the continuous level). More precisely, given the discretization parameters ρ, h, and ε and an initial guess m^0 for the solution of (3.22), we compute v^0 by solving backwards (5.12) with $\mu = m^0$. The new iterate m^1 is computed using scheme (3.14) with

$$\Phi^{\pm}_{j,k} = x_j - h\tilde{\nabla}(v^0)^{\varepsilon}_{j,k} \pm \sqrt{h}\sigma,$$

where $\tilde{\nabla}(v^0)^{\varepsilon}_{j,k}$ is an approximation of $\nabla_x v^{\rho,h,\varepsilon}[m^0](x_j,t_k)$. Then, given m^p $(p \geq 1)$ we compute v^p by solving backwards (5.12) with $\mu = \frac{1}{p+1}\sum_{p'=0}^{p} m^{p'}$ and define m^{p+1} using (3.14) with

$$\Phi^{\pm}_{j,k} = x_j - h\tilde{\nabla}(v^p)^{\varepsilon}_{j,k} \pm \sqrt{h}\sigma,$$

where $\tilde{\nabla}(v^p)^{\varepsilon}_{j,k}$ is an approximation of $\nabla_x v^{\rho,h,\varepsilon}[m^p](x_j,t_k)$. We continue with these iterations until the difference between m^p and m^{p+1} is less than 0.01 in the discrete infinity norm.

Remark 5.2. Numerically, this heuristic performs rather well. The proof of convergence of this algorithm is not analyzed in this paper, and it is postponed to a future work. One could expect that the arguments in [17] apply to a discrete time, discrete space MFG (see [35]). The main issue with the approximation (3.22) is that it does not correspond exactly to a discrete MFG because the distribution of the players does not evolve according to the discrete optimal controls of the typical players (computed as the optimizers of the right-hand side of (5.12)), but it evolves according to their approximations $\nabla_x v^{\rho,h,\varepsilon}[m^p](x_j,t_k)$.

The numerical approximation of the density $\mathbf{m}^{\rho,h,\varepsilon}$ for $\rho = 0.02$, $h = \rho$, $\varepsilon = 0.15$, and $\delta = 0.02$ is depicted in Figure 1. In Figure 2, we plot the densities $\mathbf{m}^{\rho,h,\varepsilon}$ at times $t = 0$, 0.6, and 5. We observe that the density of agents divides into three groups. The largest one moves towards the right meeting area which is the closest one. The second largest group moves towards the left area. The third and smallest group waits before moving towards the meeting area. We note that, in this equilibrium configuration, the agents somehow make rational decisions based on their aversion to crowded places out of the meeting zones.

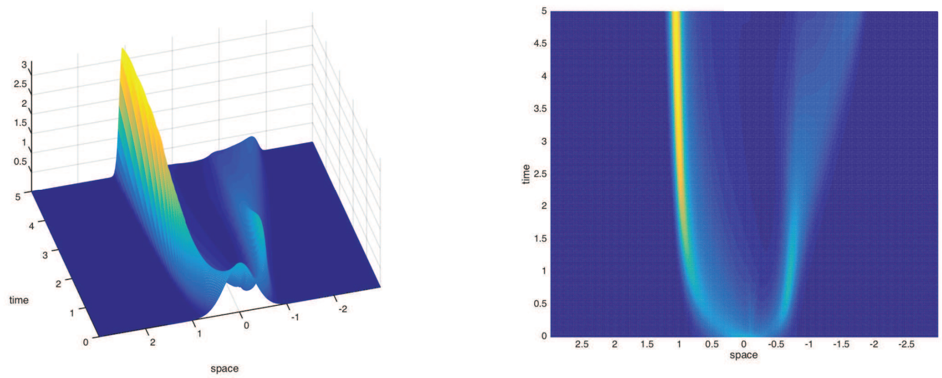

FIG. 1. *Test* 5.1.1*: 3D and 2D views in the (x,t) domain of the evolution of the density of agents.*

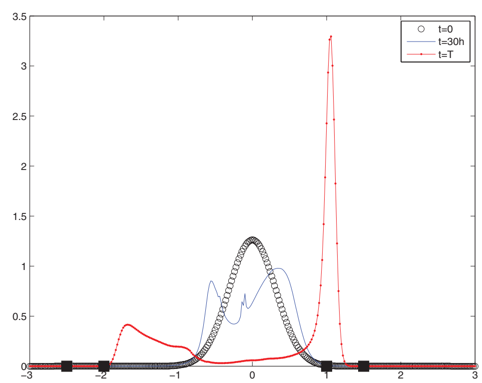

FIG. 2. *Test* 5.1.1*: Densities at times $t = 0$, 0.6, and 5 (black squares on the x axis represent the boundary of the "meeting areas").*

5.2. A nonlinear Hughes-type explicit model. In this section we consider the FPK equation

$$\partial_t m - \frac{\sigma^2}{2}\Delta m - \operatorname{div}\big(\nabla v[m]m\big) = 0 \quad \text{in } \mathbb{R}^d \times (0,T), \quad m(0) = \bar{m}_0, \tag{5.17}$$

where $v : C([0,T]; \mathcal{P}_1(\mathbb{R}^d)) \times \mathbb{R}^d \times [0,T] \to \mathbb{R}$ is given by

$$v[m](x,t) := \inf_{\alpha} \mathbb{E}\left(\int_t^T \left[\tfrac{1}{2}|\alpha(s)|^2 + F(X^{x,t,\alpha}(s), m(t))\right] \mathrm{d}s + G(X^{x,t,\alpha}(T), m(t))\right), \tag{5.18}$$

and the processes α and $X^{x,t,\alpha}$ are as in section 5.1. We also assume that F and G satisfy (5.2).

Note that the main difference with the MFG model considered in section 5.1 is that the optimal control problem solved by an agent located at point x at time t depends on the global distribution m of the agents *only* through its value at time t. In this sense, agents do not forecast, or in other words, no learning procedure has been adopted by the population of agents regarding their future behavior (see [17] for the analysis of the *fictitious play* procedure in MFGs which can explain the formation of the equilibria). This model is a variation of the one introduced by Hughes in [29], where the optimal control problem solved by the typical player is stationary of minimum time type. In terms of PDEs, at each time $t \in (0,T)$ we consider the HJB equation

$$
\begin{array}{rcl}
-\partial_s u(x,s) - \frac{\sigma^2}{2}\Delta u(x,s) + \frac{1}{2}|\nabla u(x,s)|^2 &=& F(x,m(t)) \ \text{ in } \mathbb{R}^d \times (0,T), \\
u(x,T) &=& G(x,m(t)) \ \text{ for } x \in \mathbb{R}^d,
\end{array} \tag{5.19}
$$

which admits a classical solution $u[m(t)]$. We have that $v[m](x,t) = u[m(t)](x,t)$. By the continuity of F and G, assumption (5.2), and the representation formula (5.18), we have that v is continuous. This can also be seen as a consequence of the stability of viscosity solutions with respect to continuous parameter perturbations (for (5.19) the parameter is $m(t)$). Moreover, as in the case of MFG, assumption (5.2) implies that

$$
\sup_{t\in[0,T],\ m\in C([0,T];\mathcal{P}_1(\mathbb{R}^d))} |\nabla_x v[m](\cdot,t)|_\infty < \infty, \tag{5.20}
$$

and that for all $t \in [0,T]$, $\nabla_x v[m](\cdot,t)$ is locally Lipschitz, with local Lipschitz constants which are independent of (m,t). In addition, $v[m](\cdot,t)$ is semiconcave, with a semiconcavity constant which is independent of (m,t). Using this property and arguing exactly as in section 5.1 we obtain that $(m,x,t) \in C([0,T];\mathcal{P}_1(\mathbb{R}^d)) \times \mathbb{R}^d \times [0,T] \to \nabla_x v[m](x,t) \in \mathbb{R}^d$ is continuous and so Theorem 4.1 gives the following result.

Proposition 5.3. *Equation* (5.17) *admits at least one solution.*

As in the case of MFGs, in practice we do not know explicitly the velocity vector field $-\nabla_x v[m](x,t)$, and so we have to approximate it. Given $\hat\mu \in \mathcal{P}_1(\mathbb{R}^d)$, we consider first an approximation of the solution $u[\hat\mu]$ to (5.19), where $m(t)$ is replaced by $\hat\mu$. Given $\rho > 0$, $h = T/N > 0$, with $N \in \mathbb{N}$ and $\hat\mu \in \mathcal{P}_1(\mathbb{R}^d)$, we define

$$
\begin{array}{rcl}
u^{\rho,h}_{i,k}[\hat\mu] &=& \inf_{\alpha\in\mathbb{R}^d}\left\{ \frac{h}{2}|\alpha|^2 + \frac{1}{2d}\sum_{\ell=1}^d \Big(I[u^{\rho,h}_{\cdot,k}[\hat\mu]](x_i + h\alpha + \sigma\sqrt{hd}e_\ell) \right. \\
&& \left. \qquad + I[u^{\rho,h}_{\cdot,k+1}[\hat\mu]](x_i + h\alpha - \sigma\sqrt{hd}e_\ell)\Big)\right\} \\
&& \qquad + hF(x_i,\hat\mu) \quad \forall\, i \in \mathbb{Z}^d, \quad \forall\, k = 0,\dots,N-1, \\
u^{\rho,h}_{i,N}[\hat\mu] &=& G(x_i,\hat\mu) \quad \forall\, i \in \mathbb{Z}^d,
\end{array} \tag{5.21}
$$

which is extended to $\mathcal{P}_1(\mathbb{R}^d) \times \mathbb{R}^d \times [0,T]$, by setting

$$
u^{\rho,h}[\hat\mu](x,t) := I[u^{\rho,h}_{\cdot,k}[\hat\mu]](x) \quad \text{ if } \ t \in [t_k, t_{k+1}[.
$$

Now, given $\varepsilon > 0$ and $\phi \in C^\infty(\mathbb{R}^d)$, nonnegative and such that $\int_{\mathbb{R}^d}\phi(x)\mathrm{d}x = 1$, we define $u^{\rho,h,\varepsilon} : \mathcal{P}_1(\mathbb{R}^d) \times \mathbb{R}^d \times [0,T] \to \mathbb{R}$ by

$$
u^{\rho,h,\varepsilon}[\hat\mu](\cdot,t) := \left(\phi_\varepsilon * u^{\rho,h}[\hat\mu]\right)(\cdot,t) \quad \forall\, t \in [0,T],
$$

where $\phi_\varepsilon(x) := \frac{1}{\varepsilon^d}\phi(x/\varepsilon)$. Notice that, reasoning as in the proof of [21, Theorem 3.3], if $\hat{\mu}_n \to \hat{\mu}$ in $\mathcal{P}_1(\mathbb{R}^d)$ and $\rho_n^2 = o(h_n)$, then

$$u^{\rho_n,h_n,\varepsilon_n}[\hat{\mu}_n] \to u[\hat{\mu}], \qquad \text{uniformly over compact subsets of } \mathbb{R}^d \times [0,T]. \tag{5.22}$$

The approximation that we consider for $C([0,T];\mathcal{P}_1(\mathbb{R}^d))\times\mathbb{R}^d\times[0,T] \ni (\mu,x,t) \mapsto v[\mu](x,t) \in \mathbb{R}$ is $v^{\rho,h,\varepsilon} : C([0,T];\mathcal{P}_1(\mathbb{R}^d)) \times \mathbb{R}^d \times [0,T] \to \mathbb{R}$, defined as

$$v^{\rho,h,\varepsilon}[\mu](\cdot,t) = u^{\rho,h,\varepsilon}[\mu(t_k)](\cdot,t) \quad \text{ if } \ t \in [t_k, t_{k+1}[.$$

Comparing with (5.12), where given $\mu \in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ the scheme discretizes only (5.3) (with m replaced by μ), in order to approximate v, we now need to compute $u^{\rho,h,\varepsilon}[\mu(t_k)]$, the solution to (5.21), for each $k = 0,\dots,N-1$.

By assumption (5.2), the bound (5.6) and the semiconcavity property (5.13) remain valid for $v^{\rho,h,\varepsilon}$. Now, let ρ_n, h_n, and ε_n satisfy the conditions in Proposition 5.1 and let $m^n \in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ be the extension to $C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ of the solution to (3.22) computed with coefficients $b^n[\mu](x,t) := \nabla_x v^{\rho_n,h_n,\varepsilon_n}[\mu](x,t)$ and $\sigma_\ell^n = \sigma e_\ell$ ($\ell = 1,\dots,d$). As before, using that $\nabla_x v^{\rho_n,h_n,\varepsilon_n}[m_n](\cdot,t)$ is uniformly bounded in t and n, we have that m^n has at least one limit point $m \in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$. Denoting still by m^n a sequence converging to m, setting $v^n(x,t) := v^{\rho_n,h_n,\varepsilon_n}[m^n](x,t)$, and using (5.22), for all (x,t) we get

$$\begin{aligned}
\limsup_{x'\to x, t'\to t, n\to\infty} v^n(x',t') &= \limsup_{x'\to x, t'\to t, n\to\infty} u^{\rho_n,h_n,\varepsilon_n}[m^n(t_n)](x',t') \\
&= u[m(t)](x,t) \\
&= \liminf_{x'\to x, t'\to t, n\to\infty} u^{\rho_n,h_n,\varepsilon_n}[m^n(t_n)](x',t') \\
&= \liminf_{x'\to x, t'\to t, n\to\infty} v^n(x',t'),
\end{aligned}$$

where $t_n \in \{0, h_n, 2h_n, \dots, T-h_n\}$ is such that $t' \in [t_n, t_n + h_n)$. Since $v[m](x,t) = u[m(t)](x,t)$, by [7, Chapter V, Lemma 1.9] we get that v^n converges to $v[m]$ uniformly over compact subsets of $\mathbb{R}^d \times [0,T]$. Since v^n also satisfies (5.13), arguing as in the proof of Proposition 5.1, we have that $\nabla_x v^n \to \nabla_x v[m]$ uniformly over compact subsets of $\mathbb{R}^d \times [0,T]$. Using this fact, the proof of the following result is the same as the proof of Proposition 5.2.

PROPOSITION 5.4. *Assume that $\rho_n^2 = o(h_n)$ and that $\rho_n = o(\varepsilon_n)$. Then, every limit point $m \in C([0,T];\mathcal{P}_1(\mathbb{R}^d))$ of m^n (there exists at least one) solves* (5.17).

5.2.1. Numerical test. For the sake of comparison, we consider here the same framework as the one in subsection 5.1.1; i.e., we take $d = r = 1$, we work on the domain $\mathcal{O} \times [0,T] = [-3,3] \times [0,5]$, and we impose a homogeneous Neumann boundary condition on the FPK equation (5.17). The functions F and G are also as in the previous test, as well as the initial distribution $\bar{m}_0$ of the agents.

We proceed iteratively in the following way: given the discrete measure $m_k^{\rho,h,\varepsilon}$ at time t_k ($k = 0,\dots,N-1$), we compute at each space grid point j the discrete value function $v_{j,k}$ by using (5.21) with $\mu(t_k)$ replaced by $m_k^{\rho,h,\varepsilon}$. We regularize the interpolated function $I[v_{\cdot,k}]$ by using a discrete space convolution with a mollifier ϕ_ε. We denote by $\tilde{\nabla} v_{j,k}^\varepsilon$ the approximation of its spatial gradient at x_j. Then we calculate $m_{k+1}^{\rho,h,\varepsilon}$ with scheme (3.22) by approximating the discrete trajectories by

$$\Phi_{j,k}^{\pm} = x_j - h\tilde{\nabla} v_{j,k}^\varepsilon \pm \sqrt{h}\sigma,$$

and we iterate the process until $k = N - 1$. Note that, by construction, the scheme is explicit in time.

The approximation of the density evolution in the (x, t) domain, computed with $\rho = 0.02$, $h = \rho$, $\varepsilon = 0.15$, and $\delta = 0.01$, is shown in Figure 3. In Figure 4, we plot the approximated density at times $t = 0$, 0.6, and 5. We observe that the initial density $\bar{m}_0$ divides into two parts. The first one quickly reaches the meeting area on the right, and once there it stops and begins to accumulate in this zone. The second part of the density moves in the opposite direction, trying to reach the left meeting area. In contrast to the presented MFG model, in this model the agents make their decisions based only on the current global configuration. As a consequence, we observe faster and higher accumulation of agents in the meeting zones.

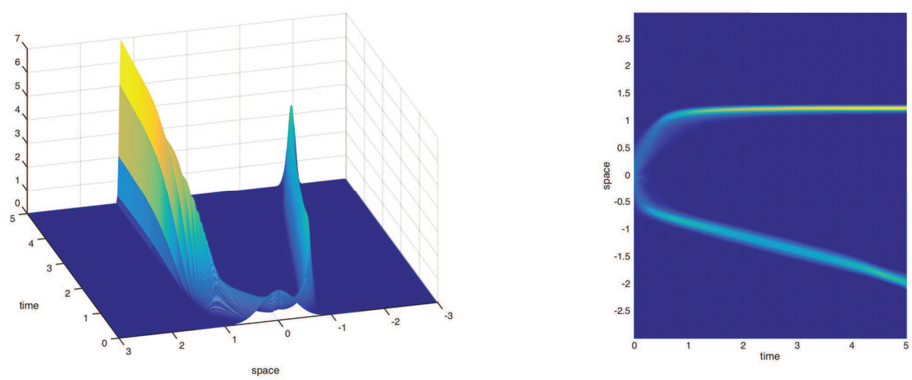

FIG. 3. *Test* 5.2.1*: 3D and 2D views in the (x, t) domain of the evolution of the density of agents.*

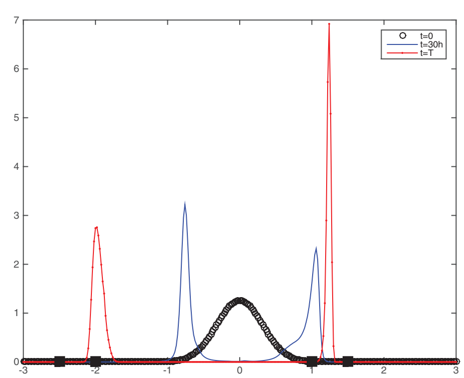

FIG. 4. *Test* 5.2.1*: Density of agents at times $t = 0$, 0.6, and 5 (black squares on the x axis represent the boundary of the "meeting areas").*

REFERENCES

[1] Y. ACHDOU, F. CAMILLI, AND L. CORRIAS, *On numerical approximation of the Hamilton-Jacobi-transport system arising in high frequency approximations*, Discrete Contin. Dyn. Syst. Ser. B, 19 (2014), pp. 629–650.

[2] Y. Achdou and I. Capuzzo-Dolcetta, *Mean field games: Numerical methods*, SIAM J. Numer. Anal., 48 (2010), pp. 1136–1162, https://doi.org/10.1137/090758477.

[3] L. Ambrosio, *Lecture notes on optimal transport problem*, in Mathematical Aspects of Evolving Interfaces, P. Colli and J. Rodrigues, eds., CIME Foundation Ser. 1812, Springer, Berlin, 2003, pp. 1–52.

[4] L. Ambrosio, *Transport equation and Cauchy problem for BV vector fields*, Invent. Math., 158 (2004), pp. 227–260.

[5] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed., Lectures in Math. ETH Zürich, Birkhäuser, Basel, 2008.

[6] M. Annunziato and A. Borzì, *A Fokker-Planck control framework for multidimensional stochastic processes*, J. Comput. Appl. Math., 237 (2013), pp. 487–507.

[7] M. Bardi and I. Capuzzo Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Birkäuser, Basel, 1996.

[8] G. Barles and P. E. Souganidis, *Convergence of approximation schemes for fully nonlinear second order equations*, Asymptot. Anal., 4 (1991), pp. 271–283.

[9] V. I. Bogachev and A. V. Kolesnikov, *The Monge-Kantorovich problem: Achievements, connections, and prospects*, Uspekhi Mat. Nauk, 67 (2012), pp. 3–110.

[10] V. I. Bogachev, N. V. Krylov, M. Röckner, and S. V. Shaposhnikov, *Fokker-Planck-Kolmogorov Equations*, Math. Surveys Monogr. 207, AMS, Providence, RI, 2015.

[11] V. I. Bogachev, M. Röckner, and S. V. Shaposhnikov, *Nonlinear evolution and transport equations for measures*, Dokl. Akad. Nauk, 429 (2009), pp. 7–11.

[12] 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), pp. 1262–1300.

[13] L. Bonaventura and R. Ferretti, *Semi-Lagrangian methods for parabolic problems in divergence form*, SIAM J. Sci. Comput., 36 (2014), pp. A2458–A2477, https://doi.org/10.1137/140969713.

[14] F. Camilli and M. Falcone, *An approximation scheme for the optimal control of diffusion processes*, RAIRO Modél. Math. Anal. Numér., 29 (1995), pp. 97–122.

[15] P. Cannarsa and C. Sinestrari, *Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control*, Progr. Nonlinear Differential Equations Appl. 58, Birkäuser Boston, Boston, MA, 2004.

[16] P. Cardaliaguet, *Notes on Mean Field Games: From P.-L. Lions' lectures at Collège de France*, Lecture Notes given at Tor Vergata, 2010.

[17] P. Cardaliaguet and S. Hadikhanloo, *Learning in mean field games: The fictitious play*, ESAIM Control Optim. Calc. Var., 23 (2017), pp. 569–591.

[18] E. Carlini, A. Festa, and F.-J. Silva, *The Hughes model for pedestrian dynamics and congestion modelling*, IFAC-PapersOnLine, 50 (2017), pp. 1655–1660.

[19] E. Carlini, A. Festa, F. J. Silva, and M.-T. Wolfram, *A semi-Lagrangian scheme for a modified version of the Hughes' model for pedestrian flow*, Dyn. Games Appl., 7 (2017), pp. 683–705.

[20] E. Carlini and F. J. Silva, *Semi-Lagrangian schemes for mean field game models*, in Proceedings of the 52nd Annual IEEE Conference on Decision and Control (CDC), IEEE, Washington, DC, 2013, pp. 3115–3120.

[21] E. Carlini and F. J. Silva, *A fully discrete semi-Lagrangian scheme for a first order mean field game problem*, SIAM J. Numer. Anal., 52 (2014), pp. 45–67, https://doi.org/10.1137/120902987.

[22] E. Carlini and F. J. Silva, *A semi-Lagrangian scheme for a degenerate second order mean field game system*, Discrete Contin. Dyn. Syst., 35 (2015), pp. 4269–4292.

[23] E. Carlini and F. J. Silva, *A semi-Lagrangian scheme for the Fokker-Planck equation*, IFAC-PapersOnLine, 49 (2016), pp. 272–277.

[24] J. S. Chang and G. Cooper, *A practical difference scheme for Fokker-Planck equations*, J. Comput. Phys., 6 (1970), pp. 1–16.

[25] L. Chen, E. R. Jakobsen, and A. Naess, *On Numerical Density Approximations of Solutions of SDEs with Unbounded Coefficients*, preprint, https://arxiv.org/abs/1506.05576, 2015.

[26] P. G. Ciarlet and J.-L. Lions, eds., *Handbook of Numerical Analysis. Vol.* II. *Finite Element Methods. Part* 1, Handbook of Numerical Analysis II, North–Holland, Amsterdam, 1991.

[27] K. Debrabant and E. R. Jakobsen, *Semi-Lagrangian schemes for linear and fully non-linear diffusion equations*, Math. Comp., 82 (2013), pp. 1433–1462.

[28] F. Delarue, F. Lagoutière, and N. Vauchelet, *Convergence Order of Upwind Type Schemes for Transport Equations with Discontinuous Coefficients*, preprint, https://arxiv.org/abs/1602.05746, 2016.

[29] M. Di Francesco, P. A. Markowich, J.-F. Pietschmann, and M.-T. Wolfram, *On the Hughes' model for pedestrian flow: The one-dimensional case*, J. Differential Equations, 250 (2011), pp. 1334–1362.

[30] R. J. DiPerna and P.-L. Lions, *Ordinary differential equations, transport theory and Sobolev spaces*, Invent. Math., 98 (1989), pp. 511–547.

[31] A. N. Drozdov and M. Morillo, *Solution of nonlinear Fokker-Planck equations*, Phys. Rev. E, 54 (1996), pp. 931–937.

[32] A. Figalli, *Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients*, J. Funct. Anal., 253 (2008), pp. 109–153.

[33] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*, 2nd ed., Stoch. Model. Appl. Prob. 25, Springer, New York, 2006.

[34] T. Funaki, *A certain class of diffusion processes associated with nonlinear parabolic equations*, Z. Wahrsch. Verw. Gebiete, 67 (1984), pp. 331–348.

[35] D. A. Gomes, J. Mohr, and R. Souza, *Discrete time, finite state space mean field games*, J. Math. Pures Appl., 93 (2010), pp. 308–328.

[36] M. Huang, R. P. Malhamé, and P. E. Caines, *Large population stochastic dynamic games: Closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle*, Commun. Inf. Syst., 6 (2006), pp. 221–251.

[37] R. L. Hughes, *The flow of large crowds of pedestrians*, Math. Comput. Simul., 53 (2000), pp. 367–370.

[38] A. Klar, P. Reuterswäd, and M. Seaïd, *A Semi-Lagrangian method for a Fokker-Planck equation describing fiber dynamics*, J. Sci. Comput., 38 (2009), pp. 349–367.

[39] H. J. Kushner, *Finite difference methods for the weak solutions of the Kolmogorov equations for the density of both diffusion and conditional diffusion processes*, J. Math. Anal. Appl., 53 (1976), pp. 251–265.

[40] H. J. Kushner, *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations*, Math. Sci. Engrg. 129, Academic Press, New York, London, 1977.

[41] H. J. Kushner and P. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, 2nd ed., Appl. Math. (New York) 24, Springer-Verlag, New York, 2001.

[42] F. Lagoutière and N. Vauchelet, *Analysis and simulation of nonlinear and nonlocal transport equations*, in Innovative Algorithms and Analysis, Springer INdAM Ser. 16, Springer, Cham, 2017, pp. 265–288.

[43] J.-M. Lasry and P.-L. Lions, *Jeux à champ moyen* I. *Le cas stationnaire*, C. R. Math. Acad. Sci. Paris, 343 (2006), pp. 619–625.

[44] J.-M. Lasry and P.-L. Lions, *Jeux à champ moyen* II. *Horizon fini et contrôle optimal*, C. R. Math. Acad. Sci. Paris, 343 (2006), pp. 679–684.

[45] J.-M. Lasry and P.-L. Lions, *Mean field games*, Jpn. J. Math., 2 (2007), pp. 229–260.

[46] C. Le Bris and P.-L. Lions, *Existence and uniqueness of solutions to Fokker-Planck type equations with irregular coefficients*, Comm. Partial Differential Equations, 33 (2008), pp. 1272–1317.

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

[48] O. A. Manita, M. S. Romanov, and S. V. Shaposhnikov, *Uniqueness of a probability solution of a nonlinear Fokker-Planck-Kolmogorov equation*, Dokl. Akad. Nauk, 461 (2015), pp. 18–22.

[49] O. A. Manita and S. V. Shaposhnikov, *Nonlinear parabolic equations for measures*, Dokl. Akad. Nauk, 447 (2012), pp. 610–614.

[50] O. A. Manita and S. V. Shaposhnikov, *Nonlinear parabolic equations for measures*, Algebra i Analiz, 25 (2013), pp. 64–93 (in Russian).

[51] S. Méléard, *Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models*, in Probabilistic Models for Nonlinear Partial Differential Equations (Montecatini Terme, 1995), Lecture Notes in Math. 1627, Springer, Berlin, 1996, pp. 42–95.

[52] G. N. Milstein, *The probability approach to numerical solution of nonlinear parabolic equations*, Numer. Methods Partial Differential Equations, 18 (2002), pp. 490–522.

[53] M. Mohammadi and A. Borzì, *Analysis of the Chang-Cooper discretization scheme for a class of Fokker-Planck equations*, J. Numer. Math., 23 (2015), pp. 271–288.

[54] A. Naess and J. M. Johnsen, *Response statistics of nonlinear, compliant offshore structures by the path integral solution method*, Probab. Engrg. Mech., 8 (1993), pp. 91–106.

[55] B. Piccoli and A. Tosin, *Time-evolving measures and macroscopic modeling of pedestrian flow*, Arch. Ration. Mech. Anal., 199 (2011), pp. 707–738.

[56] A. Quarteroni, *Numerical Models for Differential Problems*, Springer-Verlag, New York, 2016.

[57] A. Quarteroni, R. Sacco, and F. Saleri, *Numerical Mathematics*, 2nd ed., Springer, Berlin, 2007.

[58] A. Schlichting and C. Seis, *Convergence rates for upwind schemes with rough coefficients*, SIAM J. Numer. Anal., 55 (2017), pp. 812–840, https://doi.org/10.1137/16M1084882.

[59] B. F. Spencer, Jr., and L. A. Bergman, *On the numerical solution of the Fokker-Planck equation for nonlinear stochastic systems*, Nonlinear Dynam., 4 (1993), pp. 357–372.

[60] D. W. Stroock and S. R. S. Varadhan, *Multidimensional Diffusion Processes*, Grundlehren Math. Wiss. 233, Springer-Verlag, Berlin, New York, 1979.

[61] C. Villani, *Topics in Optimal Transportation*, Grad. Stud. Math. 58, AMS, Providence, RI, 2003.

[62] M. P. Zorzano, H. Mais, and L. Vazquez, *Numerical solution of two-dimensional Fokker-Planck equations*, Appl. Math. Comput., 98 (1999), pp. 109–117.