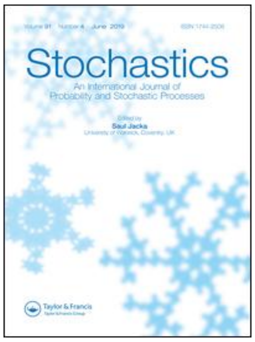

Stochastics

An International Journal of Probability and Stochastic Processes

ISSN: 1744-2508 (Print) 1744-2516 (Online) Journal homepage: https://www.tandfonline.com/loi/gssr20

The density evolution of the killed McKean–Vlasov process

Peter E. Caines, Daniel Ho & Qingshuo Song

To cite this article: Peter E. Caines, Daniel Ho & Qingshuo Song (2019): The density evolution of the killed McKean–Vlasov process, Stochastics, DOI: 10.1080/17442508.2019.1650041

To link to this article: https://doi.org/10.1080/17442508.2019.1650041

Published online: 06 Aug 2019.

STOCHASTICS
https://doi.org/10.1080/17442508.2019.1650041

The density evolution of the killed McKean–Vlasov process

Peter E. Caines[a], Daniel Ho[b] and Qingshuo Song [b]

[a]Department of Electrical and Computer Engineering, McGill University, Montreal, Canada; [b]Department of Mathematics, City University of Hong Kong, Kowloon, Hong Kong

ABSTRACT
The density evolution of McKean–Vlasov stochastic differential equations in the presence of an absorbing boundary is analysed where the solution to such equations corresponds to the dynamics of partially killed large populations. By using a fixed point theorem, we show that the density evolution is characterized as the solution of an integro-differential Fokker–Planck equation with Cauchy–Dirichlet data. This problem arises naturally within mean field game theory.

ARTICLE HISTORY
Received 28 December 2018
Accepted 26 July 2019

KEYWORDS
McKean–Vlasov stochastic differential equations; absorbing boundary; integro-differential Fokker–Planck equation; Cauchy–Dirichlet data

1. Introduction

Let $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}_{t\geq 0})$ be a filtered probability space satisfying the usual conditions on which there exists a $\mathbb{R}^d$-valued $\mathcal{F}_t$-Brownian motion. We consider the stochastic differential equation

$$\mathrm{d}X_t = \bar{b}(X_t, t)\,\mathrm{d}t + \mathrm{d}W_t; \quad X_0 \sim m_0,$$

where X_0 is the given initial state with its probability density m_0 on $\mathbb{R}^d$. A well-known result, see for instance Hörmander's Theorem in Section V.38 of [19], says that the density $m(x,t) = \mathbb{P}(X_t \in \mathrm{d}x)/\mathrm{d}x$ satisfies Fokker–Planck equation (FPK) with Cauchy data

$$\begin{aligned} \partial_t m &= -\mathrm{div}_x(\bar{b}\, m) + \tfrac{1}{2}\Delta m, \quad (0,\infty)\times\mathbb{R}^d, \\ m(x,0) &= m_0(x), \quad x\in\mathbb{R}^d. \end{aligned}$$

Recently, Mean Field Game theory attracted a great deal of attention in the control and other fields after it was initiated in a series of founding works by Huang, Caines, and Malhamé (e.g. [13–15]), and independently in that of Lasry and Lions (e.g. [17]). This led to extensive studies on the density evolution of McKean–Vlasov type stochastic differential equations (MV-SDE) of the following form:

$$\mathrm{d}X_t = b(X_t, \mathbb{E}[X_t^p])\,\mathrm{d}t + \mathrm{d}W_t; \quad X_0 \sim m_0 \tag{1}$$

CONTACT Qingshuo Song qsong@wpi.edu

for some positive integer p, see, for instance, [2,4,12] and the references therein. In this connection, it is well known that the density follows the integro-differential FPK

$$\begin{aligned} &\partial_t m = -\mathrm{div}_x \left(b\left(x, \int_{\mathbb{R}^d} x^p m(x,t)\, \mathrm{d}x\right) m\right) + \frac{1}{2}\Delta m, \quad (0,\infty)\times\mathbb{R}^d, \\ &m(x,0) = m_0(x), \quad x \in \mathbb{R}^d. \end{aligned} \tag{2}$$

In this paper, we study a similar integro-differential FPK associated to population density dynamic of the process (1) killed at the boundary of the unit ball. Indeed, the setup of killed population has been already applied to Mean Field Games in different contexts, see for instance, [3,8,9]. One of the closest references to the formulation in this paper is [3], where an FPK similar to our current study has been briefly sketched, see page 2217 of [3]. However, to the best of our knowledge, the corresponding study on the solvability and regularity properties of the FPK for the killed population is not available.

It is noted that, due to the loss of population at the boundary, the killed process is strictly submarkovian in B_1. As a result, its associated FPK is given with Dirichlet data along the boundary in addition to that of the counter-part FPK of the unkilled process (2). Our goal is to justify the following statement: *Under appropriate conditions on the drift function $b(\cdot,\cdot)$, the density of killed process is the classical solution of its associated integro-differential FPK with initial-boundary data.*

In this paper, in Section 2, we first present the precise problem formulation and its main result on the characterization of the density. In Section 3, we present the detailed proof, which is mainly based on the argument of the Leray–Schauder fixed point theorem. Section 4 gives a summary, and the last section is an Appendix presenting some useful facts from the existing literature.

2. Motivation, problem setup and main results

2.1. Motivation

An FPK equation and Hamilton–Jacobi–Bellman equation coupled pair arises naturally in Mean Field Game theory to describe the controlled underlying population dynamics. To better focus on the FPK equation in this paper, we present below a motivation from a class of Mean Field particle systems without control.

- Given a particle system of population size N, our interest is in the evolution of the mean field term given by

$$Z_t^N := \frac{1}{N}\sum_{i=1}^{N}(X_t^{i,N})^p, \tag{3}$$

where we only consider p as a positive integer. If $p=1$, then the mean field term Z_t^N corresponds to the population mean. Suppose the position $X_t^{i,N}$ of the ith particle follows the dynamics

$$\mathrm{d}X_t^{i,N} = b(X_t^{i,N}, Z_t^N)\,\mathrm{d}t + \mathrm{d}W_t^i; \quad X_0^{i,N} \sim m_0, \quad i = 1,2,\ldots,N \tag{4}$$

driven by an independent Brownian motion W^i with i.i.d. initial distribution $X_0^{i,N} \sim m_0$, then one could solve a system of N-equations (4) to track the mean field term Z_t^N. This bears a high computation cost if N is a fairly large number.

On the other hand, if N is a large number, a version of law of large number for exchangeable sequences implies that Z_t^N can be effectively approximated by a deterministic process $\int_{\mathbb{R}^d} x^p m(x,t)\,\mathrm{d}x$ for large N in the sense that

$$\lim_{N\to\infty} Z_t^N = \int_{\mathbb{R}^d} x^p m(x,t)\,\mathrm{d}x, \qquad \text{almost surely } \forall t > 0.$$

where the function $m(\cdot,t)$ denotes the density of X_t of (1) at its continuum limit or equivalently is the solution of FPK (2). Indeed, a rigorous treatment with the application of the Hewitt and Savage Theorem (see Theorem 5.14 of [4]) reveals that, the FPK (2) can be used as an approximation for a fairly broad class of symmetric functionals in a large system.

A similar argument may also be resorted to for a partially killed N-particle system.

- Suppose the process $X^{i,N}$ of the ith particle following MV-SDE (4) has an open unit ball B_1 as its state space. This means that $X^{i,N}$ explodes (i.e. absorbed) at the first exit time $\zeta^{i,N}$ from B_1. If we denote the size of the population at time t by

$$L_t := \sum_{i=1}^{N} I_{(t,\infty)}(\zeta^{i,N}),$$

then the L_t is monotonically decreasing from the initial size $L_0 = N$ to 0 as t goes to infinity. If we again consider the mean field term, then the law of large number implies that

$$Y_t^N := \frac{1}{N}\sum_{i=1}^{N}(X_t^{i,N})^p I_{(t,\infty)}(\zeta^{i,N}) \to \int_{B_1} x^p m(x,t)\,\mathrm{d}x, \qquad \text{almost surely } \forall t > 0. \tag{5}$$

In the above, $m(\cdot,t)$ is the density of the killed generic process X_t in B_1 with its absorbing boundary.

Therefore, a characterization of the density m in terms of its associated FPK is desirable for the killed process, however it's not available in the literature to the best of authors' knowledge.

2.2. Problem setup

We recall that W is a $\mathbb{R}^d$-valued Brownian motion on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}_{t\geq 0})$. Let B_1 be the open unit ball in $\mathbb{R}^d$ and p be a positive integer. We consider MV-SDE of the form (1) with the only difference being that the state space is B_1. To emphasize the role of the state space, we insert the symbol B_1 after the semicolon,

$$dX_t = (b(X_t, \mathbb{E}[X_t^p; B_1])\,\mathrm{d}t + \mathrm{d}W_t) I_{B_1}(X_t); \qquad X_0 \sim m_0. \tag{6}$$

In the above, the function $b : B_1 \times B_1 \ni (x,y) \mapsto b(x,y) \in \mathbb{R}^d$ is a given drift and $X_0 : \Omega \mapsto B_1$ is a given $\mathcal{F}_0$-measurable initial state with the density m_0 on B_1. Moreover, the mean field term in the drift function is understood to be given by

$$\mathbb{E}[X_t^p; B_1] := \mathbb{E}[X_t^p I_{B_1}(X_t)] = \int_{B_1} x^p m(x,t)\,\mathrm{d}x.$$

Note that, the boundary ∂B_1 is set to be the absorbing boundary, i.e. once X_t reaches the cemetery ∂B_1, the term $I_{B_1}(X_t)$ leads henceforth to $\mathrm{d}X_t = 0$ and the process X never returns to B_1. We adopt the convention

$$\zeta = \inf\{t > 0 : X_t \notin B_1\}$$

as the lifetime of X with $\inf \emptyset = +\infty$. We are interested in establishing

(1) The existence and uniqueness of the solution of an MV-SDE (6) up to the lifetime ζ,
(2) The density evolution $m(x,t) = \mathbb{P}(X_t \in \mathrm{d}x)/\mathrm{d}x$ for $(x,t) \in B_1 \times \mathbb{R}^+$, if it exists.

The above questions are well studied for MV-SDEs if the state X_t takes values in the whole space $\mathbb{R}^d$, see for instance [5]. However, if the state space is the bounded set B_1, one note that $\mathbb{P}(X_t \in B_1) < 1$ for $t > 0$, hence the process X_t has to be a submarkovian (see the definition of submarkovian in Page 9 of [6]). Indeed, due to the positive probability of the explosion on any time interval $(0,t)$, one can obtain,

$$\int_{B_1} m(x,t)\,\mathrm{d}x = \mathbb{P}(X_t \in B_1) = 1 - \mathbb{P}(X_t \in \partial B_1) < 1, \quad \forall t > 0,$$

as long as the solution X of (1) exists. To proceed, let us precisely define a solution of (6).

Definition 2.1: Given a complete filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}_{t\geq 0})$ in which is defined an $\mathbb{R}^d$-valued $\mathcal{F}_t$-adapted Brownian motion W and a B_1-valued $\mathcal{F}_0$-measurable random variable X_0 as the initial state, a process X is said to be the solution of (6) (up to the explosion time) with the state space B_1, if (X,β)
satisfies both

$$\mathrm{d}X_t = (b(X_t,\beta_t)\,\mathrm{d}t + \mathrm{d}W_t)\cdot I_{B_1}(X_t), \quad X_0 \sim m_0, \tag{7}$$

and

$$\beta_t = \mathbb{E}[X_t^p I_{B_1}(X_t)]. \tag{8}$$

For a two-variable function $m : \mathbb{R}^d \times \mathbb{R} \ni (x,t) \mapsto m(x,t) \in \mathbb{R}$, we often treat $m_t(x) = m(x,t)$ as a process $t \mapsto m_t(\cdot)$ with the state space taken to be a function space. If β_t were given by a known deterministic process in (7) a priori, one can directly write the density evolution by FPK (A3) with the substitution $\bar{b}(x,t) = b(x,\beta_t)$. Furthermore, one could use (8) to replace β_t in FPK (A3) by $\int_{B_1} x^p m_t \mathrm{d}x$, and obtain a new FPK in the following form for its density evolution at a heuristic level:

$$\partial_t m = \frac{1}{2}\Delta m - \mathrm{div}_x(m_t(x)\, b(x, \int_{B_1} x^p m_t(x)\,\mathrm{d}x)), \quad \text{on } B_1 \times (0,\infty);$$

$$
\begin{aligned}
&m(x,0) = m_0(x), \quad \text{on } \bar{B}_1;\\
&m(x,t) = 0, \quad \text{on } \partial B_1 \times (0,\infty),
\end{aligned} \tag{9}
$$

where in Equation (9) above, the divergence term shall read

$$
\mathrm{div}_x(m_t(x)b(x,y)) = \sum_{i=1}^{d} \partial_{x_i}(m_t(x)b^i(x,y)), \quad \text{where } y = \int_{B_1} x^p m_t(x)\,\mathrm{d}x.
$$

2.3. Main result

Before we present our main results, we briefly recall the notion of the elliptic Hölder space $C^{k+\gamma}(\mathcal{D},\mathcal{R})$ of [16]. For a non-negative integer k and a positive real $\gamma \in (0,1]$, we say a function $f : \mathcal{D} \mapsto \mathcal{R}$ belongs to $C^{k+\gamma}(\mathcal{D},\mathcal{R})$ if all of kth derivatives of f are γ-Hölder continuous and uniformly bounded, see its corresponding definition of Hölder norm given by (A1). For any positive real number $\alpha > 0$, there exists a unique decomposition in the form of $\alpha = k + \gamma$ for some non-negative integer k and positive real $\gamma \in (0,1]$. Therefore, we will denote $C^{2+0.5}$ simply by $C^{2.5}$ without any ambiguity.

A possible confusion arises from the meaning of the notations C^{k+1} and $C^{k+1.0}$, that is to say, when $\gamma = 1.0$. Recall that a function $f : \mathcal{D} \mapsto \mathcal{R}$ is said to be in C^k if all its derivatives up to the order k are continuous and uniformly bounded. For instance, $C^1(\mathbb{R},\mathbb{R})$ is the collection of all real functions with both f and f' being continuous and uniformly bounded. However, $C^{1.0}$ means $C^{k+\gamma}$ with $k=0$ and $\gamma = 1.0$, namely the sapce of Lipschitz continuous functions. An obvious inclusion is that $C^{k+1} \subset C^{k+1.0}$, for instance,

$$
f(x) = |x| \in C^{1.0}([0,1],\mathbb{R}) \setminus C^1([0,1],\mathbb{R}).
$$

Moreover, by $C_0^{k+\gamma}(\mathcal{D},\mathcal{R})$, we refer the space of functions $f : \mathcal{D} \mapsto \mathcal{R}$ in $C^{k+\gamma}(\mathcal{D},\mathcal{R})$, which are smoothly vanishing to zero outside of the domain. To reduce the notational complexity, we will drop the range from $C^{k+\gamma}(\mathcal{D},\mathbb{R})$ and write simply $C^{k+\gamma}(\mathcal{D})$ in the sequel if there is no risk of confusion. For more details about Hölder spaces, we refer to Appendix A1 at the end of this paper.

For our main result, we shall impose the following regularity assumption:

(A1) $b \in C^{1+\gamma}(B_1^2;\mathbb{R}^d)$ and $m_0 \in C_0^{2+\gamma}(B_1)$ for some $\gamma \in (0,1]$.

The main result is

Theorem 2.2: *If we assume* (A1), *then there exists a solution of MV-SDE (6), whose density satisfies FPK (9).*

2.4. An example

Next, we provide a special case of Theorem 2.2, whose statement could be easily verified from the symmetry of the initial density directly using Fourier series. The proof of the general case will be presented in Section 3.

We consider one dimensional MV-SDE (6) with the state space, the drift term, and initial density given by

$$B_1 = (-1, 1), \quad b(x, y) = y^2, \quad m_0(x) = \kappa e^{1/(x^2-1)} I_{(-1,1)}(x), \tag{10}$$

where κ is the normalization constant

$$\kappa = \left(\int_{-1}^{1} e^{1/(x^2-1)} \mathrm{d}x \right)^{-1}.$$

It is noted that both b and m_0 satisfy (A1) with $\gamma = 1$. We use the following $L^2(-1, 1)$-orthogonal basis: for all natural numbers n

$$\eta_n(x) = \sin(n\pi(x + 1)/2), \quad \forall x \in (-1, 1).$$

By Section 4.1 of [22], the equation

$$\begin{aligned} \partial_t m &= \tfrac{1}{2}\partial_{xx} m \quad \text{on } (-1, 1) \times (0, \infty); \\ m(x, 0) &= m_0(x), \quad \text{on } [-1, 1]; \\ m(\pm 1, t) &= 0, \quad \text{on } (0, \infty) \end{aligned} \tag{11}$$

has the unique solution

$$m(x, t) = \sum_{n \in \mathbb{N}} (m_0, \eta_n) e^{-n^2\pi^2 t/8} \eta_n.$$

By Page 322 of [1], the above function m given by Fourier series is also the density function of a Brownian motion with initial distribution m_0 absorbed at $\{\pm 1\}$, i.e.

$$dX_t = \mathrm{d}W_t \cdot I_{(-1,1)}(X_t); \quad X_0 \sim m_0. \tag{12}$$

If n is even, then $(m_0, \eta_n) = 0$, since m_0 is even and η_n is odd. Therefore, v is an even function of the form

$$m(x, t) = \sum_{n = odd} (m_0, \eta_n) e^{-n^2\pi^2 t/8} \eta_n. \tag{13}$$

Hence, if p is odd, we have

$$\beta_t = \mathbb{E}[X_t^p I_{(-1,1)}(X_t)] = 0, \forall t \geq 0.$$

Moreover, we observe that, due to the fact of $b(x, \beta_t) = b(x, 0) = 0$, (11) and (12) are equivalent to (9) and (6), respectively. Thus we conclude that

- With the setup (10), if p is an odd number, then the density function of (6) has an explicit form (13), and solves (9).

We also observe that, the density $m(x, t)$ goes to zero function as $t \to 0$, and the zero function is actually the stationary distribution of such a process. It is not hard to see this example has the following extensions: If there exists some $\gamma \in (0, 1]$ such that,

- $b \in C^{1+\gamma}(B_1^2; \mathbb{R}^d)$ with $b(x, 0) = 0$ for all x;
- $m_0 \in C_0^{2+\gamma}(B_1)$ is an even function,

then the density function of (6) solves (9). Furthermore, m has a representation via Fourier series, which goes to the zero function as $t \to 0$.

3. Proof of the main result

In this section, we prove Theorem 2.2. We outline the main proof of Theorem 2.2 in Section 3.1 based on some estimation results, whose proof will be provided in Section 3.2. Throughout the proof, K will be used for a generic constant, which will sometimes be written $K(\alpha, \beta)$ to indicate its dependence on α and β. We also use the elliptic and parabolic Hölder spaces, which are briefly recalled in Appendix A.1.

3.1. Definition of the mapping $\mathcal{T}$ and the proof of Theorem 2.2

We fix arbitrary $T > 0$ and $\gamma \in (0, 1)$, and define the Banach spaces $\mathcal{B}$ and $\mathcal{R}$ given by

$$\mathcal{B} = C^{1/2}((0, T); \mathbb{R}^d). \tag{14}$$

and

$$\mathcal{R} := C^{2+\gamma, 1+\frac{\gamma}{2}}(B_1 \times (0, T); \mathbb{R}). \tag{15}$$

To proceed, we introduce an operator $\mathcal{T} : \mathcal{B} \mapsto \mathcal{B}$ through the composition $\mathcal{T} = \mathcal{T}_1 \circ \mathcal{T}_2$, where $\mathcal{T}_1 : \mathcal{B} \mapsto \mathcal{R}$ and $\mathcal{T}_2 : \mathcal{R} \mapsto \mathcal{B}$ are defined as follows:

(1) Define $\mathcal{T}_1 : \beta \mapsto \mathcal{T}_1(\beta) := m$, where m solves the following equation with a given process β,

$$\begin{aligned} \partial_t m &= \tfrac{1}{2}\Delta m - \mathrm{div}_x(m\, b(x, \beta_t)), \quad \text{on } B_1 \times (0, T); \\ m(x, 0) &= m_0(x), \quad \text{on } \bar{B}_1; \\ m(x, t) &= 0, \quad \text{on } \partial B_1 \times (0, T). \end{aligned} \tag{16}$$

(2) Define $\mathcal{T}_2 : m \mapsto \mathcal{T}_2(m)$, where

$$\mathcal{T}_2(m) = \int_{B_1} x^p m_t(x)\, \mathrm{d}x. \tag{17}$$

Proof of Theorem 2.2: By Lemma 3.3, $\mathcal{T}$ is a mapping from the Banach space $\mathcal{B}$ to itself. Furthermore, the mapping $\mathcal{T} : \mathcal{B} \mapsto \mathcal{B}$ has the following properties:

(1) $\mathcal{T}$ is a continuous compact mapping by Lemma 3.4;
(2) $\{x \in \mathcal{B} : x = \lambda \mathcal{T} x,\ \lambda \in [0, 1]\}$ is bounded in $\mathcal{B}$ by Lemma 3.5.

By the Leray–Schauder's fixed point theorem (FPT) (see Theorem 11.2 of [11]), the mapping $\mathcal{T}$ has a fixed point in $\mathcal{B}$, that is to say, there exists $\hat{\beta} \in \mathcal{B}$ such that

$$\mathcal{T}\hat{\beta} = \hat{\beta}.$$

Set $\hat{m} = \mathcal{T}_1\hat{\beta}$, then the pair $(\hat{m}, \hat{\beta})$ solves FPK (9) due to the definition $\mathcal{T} = \mathcal{T}_1 \circ \mathcal{T}_2$ in (16) -(17). Also due to Proposition A.5, since $\hat{m}_t$ is the unique solution of (16), it is the density of the unique solution $\hat{X}_t$ of (7) given $\beta = \hat{\beta}$. Together with the definition of $\mathcal{T}_2$ given by (17) the process $\hat{\beta}_t$ satisfies (8). Therefore, the pair $(\hat{X}, \hat{\beta})$ solves (7)–(8), and this gives the solvability of (6) in the sense of Definition 2.1. ■

3.2. Estimates for the mapping $\mathcal{T}$

In the proof of Theorem 2.2, we have used Lemma 3.3, 3.4, 3.5 concerning the mapping $\mathcal{T}$, and we will present their proofs separately in this section. First, we shall verify that $\mathcal{T}$ is a well-defined mapping. This includes

(1) The unique solvability of FPK (16);
(2) Justification that the set $\mathcal{R}$ satisfies $\mathcal{T}_1(\mathcal{B}) \subset \mathcal{R} \subset \mathcal{T}_2^{-1}(\mathcal{B})$.

Lemma 3.1: *$\mathcal{T}_1 : \mathcal{B} \mapsto \mathcal{R}$ is a well-defined mapping with estimates*

$$|\mathcal{T}_1(\beta)|_{2+\gamma,1+\gamma/2} \le K(|\beta|_{1/2})|m_0|_{2+\gamma}.$$

Proof: Rewrite the FPK (16) into non-divergence form

$$\partial_t m = \tfrac{1}{2}\Delta m - b^\beta \circ \nabla m - m\, \mathrm{div}_x(b^\beta),$$

where

$$b^\beta(x,t) = b(x, \beta(t)).$$

Note that

- We have $b^\beta \in C^{1.0,\frac{1}{2}}(B_1 \times (0,T))$ by the application of Proposition A.4 with (A1) and (14); Moreover,

$$|b^\beta|_{1.0,1/2} = |b^\beta|_0 + [b^\beta]_{1.0,1/2} \le K|b|_{1.0}(|\beta|_{1/2} + 1).$$

- We also have $\mathrm{div}_x(b^\beta) \in C^{\gamma,\gamma/2}(B_1 \times (0,T))$. Indeed, one can write

$$\mathrm{div}_x(b^\beta(x,t)) = \sum_{i=1}^d \partial_{x_i} b^{(i)}(x, \beta_t)$$

and use Proposition A.4 once again and the fact that $\partial_{x_i} b^{(i)} \in C^\gamma(B_1^2)$, which yields

$$|\mathrm{div}_x(b^\beta)|_{\gamma,\gamma/2} \le \sum_{i=1}^d |\partial_{x_i} b^{(i)}(x,\beta_t)|_0 + \sum_{i=1}^d [\partial_{x_i} b^{(i)}(x,\beta_t)]_{\gamma,\gamma/2} \le K|b|_{1+\gamma}(|\beta|_{1/2} + 1).$$

Moreover, if we define $m_0^T(x,t) = m_0(x)$, then $m_0^T \in C^{1+\gamma/2, 2+\gamma}(B_1 \times (0,T))$. Therefore, by Theorem 10.3.3 of [2], there exists a unique solution $m \in C^{2+\gamma, 1+\frac{\gamma}{2}}(B_1 \times (0,T)) = \mathcal{R}$ for (16) and hence $\mathcal{T}_1$ is well defined.

If we set $\bar{m}(x,t) = e^{-\lambda t} m(x,t)$ with $\lambda = |b|_{1+\gamma}$, then $\mathrm{div}_x(b^\beta) + \lambda \le 0$ and $\bar{m}$ solves

$$
\begin{aligned}
\partial_t \bar{m} &= \tfrac{1}{2}\Delta \bar{m} - b^\beta \circ \nabla \bar{m} - \bar{m}\,(\mathrm{div}(b^\beta) + \lambda), \qquad \text{on } B_1 \times (0,T);\\
\bar{m}(x,0) &= m_0(x), \qquad \text{on } \bar{B}_1;\\
\bar{m}(x,t) &= 0, \qquad \text{on } \partial B_1 \times (0,T).
\end{aligned}
$$

Now we can invoke the estimation from Proposition A.6 to obtain

$$
|\bar{m}|_{2+\gamma, 1+\gamma/2} \le K(|b^\beta|_{\gamma,\gamma/2}, |\mathrm{div}_x(b^\beta) + \lambda|_{\gamma,\gamma/2})|m_0|_{2+\gamma}
$$

which is equivalent to

$$
|m|_{2+\gamma, 1+\gamma/2} \le K(|b|_{1+\gamma}, |\beta|_{1/2})|m_0|_{2+\gamma}.
$$

■

The estimation of $\mathcal{T}_2$ directly follows from its definition. We observe that the Hölder space $C^{1.0}((0,T);\mathbb{R}^d)$ used below is indeed the space of Lipschitz continuous functions, see the remark on $C^{1.0}$ and C^1 in Appendix A.1.

Lemma 3.2: *$\mathcal{T}_2 : \mathcal{R} \mapsto C^{1.0}((0,T);\mathbb{R}^d)$ is well defined with an estimate*

$$
|\mathcal{T}_2(m)|_{1.0} \le K|m|_{2+\gamma, 1+\gamma/2}.
$$

Proof: It is evident that $\mathcal{T}_2$ is well defined. To prove Lipschitz continuity of the process $\mathcal{T}_2(m)$, we shall show

$$
|\mathcal{T}_2(m)(t_1) - \mathcal{T}_2(m)(t_2)| \le K|t_1 - t_2|, \quad \forall 0 \le t_1 < t_2 \le T.
$$

This follows from the following estimate for the mapping $\mathcal{T}_2$ of $m \in \mathcal{R}$ and $0 < t_1 \le t_2 < T$:

$$
\begin{aligned}
|\mathcal{T}_2(m)(t_1) - \mathcal{T}_2(m)(t_2)| &\le \int_{B_1} |x|^p |m(x,t_1) - m(x,t_2)|\,\mathrm{d}x\\
&= \int_{B_1} |x|^p \cdot \int_{t_1}^{t_2} |\partial_t m|(x,s)\,\mathrm{d}s\,\mathrm{d}x\\
&\le \int_{B_1} |x|^p\,\mathrm{d}x\,|m|_{2+\gamma,1+\frac{\gamma}{2}}\,|t_2 - t_1|.
\end{aligned}
$$

■

Consequently, we obtain an estimation of $\mathcal{T}$ as a well-defined mapping as follows.

Lemma 3.3: *$\mathcal{T} : \mathcal{B} \mapsto C^{1.0}((0,T);\mathbb{R}^d) \subset \mathcal{B}$ is well defined with*

$$
|\mathcal{T}(\beta)|_{1.0} \le K(|\beta|_{\frac{1}{2}})|m_0|_{2+\gamma}.
$$

Proof: This is a consequence of Lemma 3.1 and 3.2. ■

So far we established that the operator $\mathcal{T}$ is well defined from its domain $\mathcal{B}$ to itself. Next, we prove key facts for the proof of the fixed point theorem: the continuity and the compactness of the operator $\mathcal{T}$. For this purpose, we briefly recall the following embedding properties for Hölder spaces. Consider two Hölder spaces C^γ and C^λ for $\gamma > \lambda > 0$. Then, $C^\gamma \subset C^\lambda$ holds and any bounded subset of C^γ is a compact subset of C^λ. Furthermore, if (1) $C^\gamma \ni \alpha_n \to \alpha$ pointwise; and (2) $|\alpha_n|_\gamma < K$ for any $n \in \mathbb{N}$, then $\alpha_n \to \alpha$ in C^λ, i.e. $|\alpha_n - \alpha|_\lambda \to 0$ as $n \to \infty$. However, $\alpha_n \to \alpha$ in C^γ (i.e. $|\alpha_n - \alpha|_\gamma \to 0$) may not be true.

Lemma 3.4: *$\mathcal{T}$ is continuous compact in $\mathcal{B}$.*

Proof: Lemma 3.3 implies that any sequence $\{\beta^n\}$ bounded in $\mathcal{B} = C^{1/2}((0,T);\mathbb{R}^d)$ maps to a sequence $\{\mathcal{T}(\beta^n)\}$ bounded in $C^{1.0}((0,T);\mathbb{R}^d)$, which is precompact in $C^{1/2}((0,T);\mathbb{R}^d)$. Thus, $\mathcal{T}$ is compact.

Next we establish the continuity of the mapping $\mathcal{T}$. If $\beta^n \to \beta^\infty$ in $\mathcal{B}$, we denote, for simplicity

$$m^n = \mathcal{T}_1(\beta^n),\ \alpha^n = \mathcal{T}_2(m^n), \quad \forall n \in \mathbb{N} \cup \{\infty\}.$$

Our objective is to show $\alpha^n \to \alpha^\infty$ in $\mathcal{B}$. Since $\{m^n : n \in \mathbb{N}\}$ is bounded in $C^{2+\gamma,1+\gamma/2}(B_1 \times (0,T);\mathbb{R})$ by Lemma 3.1, one has the pointwise convergence of m^n as $n \to \infty$ from the stability of the viscosity solution (see Section 6 of [7]), that is to say,

$$m^n(x,t) \to m^\infty(x,t), \quad \forall (x,t) \in B_1 \times (0,T).$$

In addition, $\{m^n : n \in \mathbb{N}\}$ is bounded in $C^{2+\gamma,1+\gamma/2}(B_1 \times (0,T);\mathbb{R})$ by Lemma 3.1. Hence, one can use dominated convergence theorem (see page 26 of [20]) to obtain,

$$\alpha^n(t) - \alpha^\infty(t) = \int_{B_1} x^p \cdot (m^n(x,t) - m^\infty(x,t))\,\mathrm{d}x \to 0, \quad \forall t \in (0,T).$$

Thus, α^n converges to α^∞ pointwisely. But we know $\{\alpha^n : n \in \mathbb{N}\}$ is bounded in $C^{1.0}((0,T);\mathbb{R}^d)$ by Lemma 3.3. Hence, pointwise convergence and boundedness in $C^{1.0}((0,T);\mathbb{R}^d)$ implies convergence in $C^{1/2}((0,T);\mathbb{R}^d)$, i.e. $\alpha^n \to \alpha^\infty$ in $\mathcal{B}$ (See remark before Lemma 3.4). ■

The following is the last piece required to complete the FPT.

Lemma 3.5: *The set $\{\beta : \beta = \lambda \mathcal{T}\beta, \lambda \in [0,1]\}$ is bounded in $\mathcal{B}$.*

Proof: We shall show that, there exists $M > 0$, s.t. if β solves (18) below for some $\lambda \in [0,1]$, then $|\beta|_{C^{1/2}((0,T);\mathbb{R}^d)} < M$.

$$\partial_t m = \tfrac{1}{2}\Delta m - \mathrm{div}_x(b(x,\beta_t)m), \quad B_1 \times (0,T)$$

$$
\begin{aligned}
&m(x,0) = m_0(x), \quad x \in \bar{B}_1 \\
&m(x,t) = 0, \quad \partial B_1 \times (0,T) \\
&\beta_t = \lambda \int_{B_1} x^p m(x,t)\,\mathrm{d}x \quad t \in (0,T).
\end{aligned} \tag{18}
$$

If $\lambda = 0$, then $\beta = 0$, then the conclusion trivially holds. If $\lambda \in (0,1]$, we first have

$$
|\beta|_0 \le \sup_{t\in(0,T)} \lambda \int_{B_1} |x|^p m(x,t)\,\mathrm{d}x \le \sup_{t\in(0,T)} \int_{B_1} m(x,t)\,\mathrm{d}x \le 1.
$$

So, it is enough to show the boundedness of $[\beta]_{1/2}$. To proceed, we write the stochastic representation by Proposition A.5 as follows.

$$
\begin{aligned}
\mathrm{d}X_t &= b(X_t,\beta_t)\,\mathrm{d}t + \mathrm{d}W_t, \\
\tau &= \inf\{t > 0 : X_t \notin B_1\} \\
\beta(t) &= \lambda \mathbb{E}^{m_0}\Big[|X_t|^p I_{[0,\tau)}(t)\Big].
\end{aligned}
$$

Without loss of generality, we set $0 < t_1 < t_2 < T$. Then, we have

$$
\begin{aligned}
|\beta(t_1) - \beta(t_2)|^2 &= \lambda^2 \Big|\mathbb{E}^{m_0}[X_{t_1} I_{[0,\tau)}(t_1)] - \mathbb{E}^{m_0}[X_{t_2} I_{[0,\tau)}(t_2)]\Big|^2 \\
&\le 2\lambda^2 \Big|\mathbb{E}^{m_0}[X_{t_1} I_{[t_1,t_2]}(\tau)]\Big|^2 + 2\lambda^2 \Big|\mathbb{E}^{m_0}[(X_{t_1} - X_{t_2}) I_{[t_2,T]}(\tau)]\Big|^2 \\
&\le 2\lambda^2 \Big|\mathbb{E}^{m_0}[I_{[t_1,t_2]}(\tau)]\Big|^2 + 2\lambda^2 \mathbb{E}^{m_0}[|X_{t_1} - X_{t_2}|^2] \\
&\le \lambda^2 K(|b|_0)\Big|t_1 - t_2\Big|, \quad \text{by (D9) of [11].}
\end{aligned}
$$

Therefore, we have

$$
[\beta]_{1/2} \le \lambda K^{1/2}(b_0).
$$

Thus, if we choose $M = 1 + \lambda K^{1/2}(|b_0|)$, we shall have

$$
|\beta|_{1/2} = |\beta|_0 + [\beta]_{1/2} \le 1 + \lambda K^{1/2}(|b_0|) = M.
$$

■

4. Summary

In this note, we showed that under assumption (A1) on the drift function and the initial density, the killed McKean–Vlasov process (6) solves FPK (9). We observe that uniqueness of the solution to the FPK (9) has not been established in this paper and this will be the subject of future work. If the uniqueness were true, as an application, one can approximate the population pth mean(5) of large system in a bounded region by its associated FPK. It may also be interesting to consider the estimates of the population pth moment normalized by the number of survivals, which is closely related to Y^N given by (5). Further applications can be found in the connections of non-linear filtering theory with MFG theory, see for instance [21].

Note

1. $\bar{b}\nabla f = \sum_{i=1}^{d} \bar{b}_i \partial_i f$.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was partially supported by Research Grants Council of the Hong Kong Special Administrative Region, China (CityU 11201518).

ORCID

Qingshuo Song http://orcid.org/0000-0002-9289-7462

References

[1] R.F. Bass, *Stochastic Processes*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2011.
[2] A. Bensoussan, J. Frehse, and P. Yam, *Mean Field Games and Mean Field Type Control Theory*, Springer, New York, 2013.
[3] L. Campi and M. Fischer, *n-player games and mean-field games with absorption*, Annal. Appl. Probab. 28(4) (2018), pp. 2188–2242.
[4] P. Cardaliaguet, Notes on mean field games, 2013, Available at https://www.ceremade.dauphine.fr/cardaliaguet/MFG20130420.pdf.
[5] R. Carmona and F. Delarue, *Probabilistic theory of mean field games with applications. I*, Vol. 83, Probability Theory and Stochastic Modelling, Mean field FBSDEs, control, and games, Springer, Cham, 2018.
[6] K.L. Chung and J.B. Walsh, *Markov Processes, Brownian Motion, and Time Symmetry*, Vol. 249, Springer, New York, 2006.
[7] M.G. Crandall, H. Ishii, and P.L. Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc. (N.S.) 27(1) (1992), pp. 1–67.
[8] J. Cvitanić, J. Ma, and J. Zhang, *The law of large numbers for self-exciting correlated defaults*, Stoch. Process. Their Appl. 122(8) (2012), pp. 2781–2810.
[9] D. Firoozi, A. Pakniyat, and P.E. Caines, *A mean field game – hybrid systems approach to optimal execution problems in finance with stopping times*. Proceedings of the 56th IEEE CDC, Melbourne, 2017.
[10] W.H. Fleming, and H.M. Soner, *Controlled Markov processes and viscosity solutions*, Vol. 25, 2nd ed., Stochastic Modelling and Applied Probability. Springer, New York, 2006.
[11] D. Gilbarg and N.S. Trudinger, *Elliptic partial differential equations of second order*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.
[12] D.A. Gomes, E.A. Pimentel and V. Voskanyan, *Regularity Theory for Mean-Field Game Systems*. SpringerBriefs in Mathematics. Cham, Switzerland, 2016.
[13] M. Huang, P.E. Caines and R.P. Malhame, *Individual and mass behavior in large population stochastic wireless power control problems: Centralized and nash equilibrium solutions*, in *Proceedings of the 42nd IEEE CDC, Maui, HI, USA*, 2003, pp. 98–103.
[14] M. Huang, P.E. Caines and R.P. Malhame, *Large population stochastic dynamic games: closed-loop mckean-vlasov systems and the nash certainty equivalence principle*, Commun. Inf. Syst. 6(3) (2006), pp. 221–251.
[15] M. Huang, P.E. Caines and R.P. Malhame, *Large-population cost-coupled lqg problems with non-uniform agents: Individual-mass behavior and decentralized epsilon-nash equilibria*, IEEE Trans. Automatic. Control 52 (2007), pp. 1560–1571.

[16] N.V. Krylov, *Lectures on elliptic and parabolic equations in Hölder spaces*, Vol. 12, Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 1996.
[17] J-M Lasry and P-L Lions, *Mean field games*, Jpn. J. Math 2 (2007), pp. 229–260.
[18] G.A. Pavliotis, *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*. Texts in Applied Mathematics. Springer, New York, 2014.
[19] L.C.G. Rogers and D. Williams, *Diffusions, Markov Processes and Martingales: Volume 2, Itô Calculus*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000.
[20] W. Rudin, *Real and Complex Analysis*, McGraw-Hill Book Co., New York-Düsseldorf-Johannesburg, 1970.
[21] N. Şen and P.E. Caines, *Nonlinear filtering theory for McKean-Vlasov type stochastic differential equations*, SIAM J. Control Optim. 54(1) (2016), pp. 153–174.
[22] W.A. Strauss, *Partial Differential Equations: An Introduction*, John Wiley & Sons, Ltd., Chichester, 2007.

Appendix

A.1 Hölder space

For convenience, we introduce the notion of elliptic Hölder space $C^{k+\gamma}(\mathcal{D};\mathcal{R})$ and parabolic Hölder space $C^{2k+\gamma,k+\frac{\gamma}{2}}(\mathcal{D}\times(0,T);\mathcal{R})$ from [16]. In this paper, the domain $\mathcal{D}$ may be B_1, B_1^2 or $(0,T)$; and the range $\mathcal{R}$ may be $\mathbb{R}$ or $\mathbb{R}^d$. If the range $\mathcal{R}$ is $\mathbb{R}$, then $\mathcal{R}$ may be abbreviated, for instance, $C^\gamma(\mathcal{D})$ means $C^\gamma(\mathcal{D};\mathbb{R})$. For each $\gamma>0$, we also treat $C^\gamma(\mathcal{D})$ as the same as $C^\gamma(\bar{\mathcal{D}})$ by natural bijective isometry. Indeed, if a function f belongs to $C^\gamma(\mathcal{D})$, then f is uniformly continuous and there exists a unique function g on $\mathcal{D}$ such that

$$g=f \text{ on } \mathcal{D},\ g(x)=\lim_{\mathcal{D}\ni y\to x} f(y), \quad \text{and } |f|_\gamma=|g|_\gamma.$$

A.1.1 Elliptic Hölder space

Let $\mathcal{D}$ be a domain in $\mathbb{R}^d$ and $\mathcal{R}$ be a range in $\mathbb{R}^{d_1}$. For $u:\mathcal{D}\mapsto\mathcal{R}$, we define a uniform norm by $|u|_0=\sup_{\mathcal{D}}|u|$, and we denote by $\partial_{x_i}^{\alpha_i}u$ the α_ith order partial derivative in the variable x_i, if it exists. For multi-index $\alpha=(\alpha_i: i=1,\dots d)$, we use $D^\alpha u=\partial_{x_1}^{\alpha_1}\cdots\partial_{x_d}^{\alpha_d}u$.

For $k\in\mathbb{N}\cup\{0\}$, we denote by $C^k_{loc}(\mathcal{D},\mathcal{R})$ the set of all functions $u:\mathcal{D}\mapsto\mathcal{R}$ whose derivatives $D^\alpha u$ for $|\alpha|\le k$ are continuous in $\mathcal{D}$. One can define a norm in $C^k_{loc}(\mathcal{D},\mathcal{R})$ by

$$|u|_k=\sum_{i=0}^k \max_{|\alpha|=i}|D^\alpha u|_0.$$

Then the functions u having finite norm consists of Banach space, and we refer it to $C^k(\mathcal{D},\mathcal{R})$. For instance, $u=e^x:\mathbb{R}\mapsto\mathbb{R}$ belongs to $C^k_{loc}(\mathbb{R},\mathbb{R})$ but not $C^k(\mathbb{R},\mathbb{R})$.

For $\gamma\in(0,1]$, we can also define a Hölder seminorm for a function $u\in C(\mathcal{D},\mathbb{R})$ by,

$$[u]_\gamma=\sup_{x,y\in\mathcal{D},x\neq y}\frac{|u(x)-u(y)|}{|x-y|^\gamma}.$$

Definition A.1: For a decimal number $\gamma\in(0,1]$ and an integer $k\in\mathbb{N}\cup\{0\}$, Hölder space $C^{k+\gamma}(\mathcal{D},\mathcal{R})$ is the Banach space of all functions $u\in C^k(\mathcal{D},\mathcal{R})$ for which the norm

$$|u|_{k+\gamma}=|u|_k+\max_{|\alpha|=k}[D^\alpha u]_\gamma \tag{A1}$$

is finite.

In the above, we emphasize that γ is a decimal number (writing with decimal point) and k is an integer to avoid the following ambiguity. Note that $C^{1.0}(\mathcal{D},\mathcal{R})$ is 1-Hölder space (or Lipschitz continuous space) with a finite norm w.r.t.

$$|u|_{1.0} = |u|_0 + \sup_{x,y\in\mathcal{D},x\neq y} \frac{|u(x)-u(y)|}{|x-y|^\gamma},$$

while $C^1(\mathcal{D},\mathcal{R})$ is a continuous differentiable function space with a finite norm w.r.t.

$$|u|_1 = |u|_0 + \max_{i=1}^{d} |\partial_{x_i} u|_0.$$

For instance, $f(x) = |x|$ is in $C^{1.0}([-1,1]) \setminus C^1([-1,1])$ with its norm

$$|f|_{1.0} = |f|_0 + [f]_{1.0} = 2.$$

Another example is that $g(x) = x^2 \mathrm{sgn}(x)$ is in $C^{2.0}([-1,1] \setminus C^2([-1,1])$ with

$$|g|_{2.0} = |g|_0 + |g'|_0 + [g']_{1.0} = 5.$$

In general, $C^{k+1}(\mathcal{D},\mathcal{R})$ is a proper subset of $C^{k+1.0}(\mathcal{D},\mathcal{R})$.

Next, we use the extension of $u : \mathcal{D} \mapsto \mathcal{R}$ with $\tilde{u}(x) = u(x)I_{\mathcal{D}}(x) : \mathbb{R}^d \mapsto \mathbb{R}^{d_1}$ by taking values the same as u in $\mathcal{D}$ otherwise zero.

Definition A.2: Let $\mathcal{D}$ be a bounded set in $\mathbb{R}^d$. The space $C_0^{k+\gamma}(\mathcal{D},\mathcal{R})$ is defined by

$$C_0^{k+\gamma}(\mathcal{D},\mathcal{R}) = \{u \in C^{k+\gamma}(\mathcal{D},\mathcal{R}) : u(x)I_{\mathcal{D}}(x) \in C^{k+\gamma}(\mathbb{R}^d,\mathcal{R})\}.$$

A.1.2 Parabolic Hölder space

Let $\mathcal{D}$ be a domain in $\mathbb{R}^d$, $\mathcal{Q} = \mathcal{D} \times (0,T)$ be the parabolic domain in $\mathbb{R}^{d+1}$ for some $T > 0$, and $\mathcal{R}$ be a range in $\mathbb{R}^{d_1}$. We are going to define norms for $u : \mathcal{Q} \mapsto \mathcal{R}$ in the following.

First, we define parabolic metric on $\mathbb{R}^{d+1}$: for any $z_1 = (x_1,t_1), z_2 = (x_2,t_2) \in \mathbb{R}^{d+1}$

$$\rho(z_1,z_2) = |x_1 - x_2| + |t_1 - t_2|^{1/2}.$$

Then, we set the parabolic Hölder seminorm for $u \in C(\mathcal{Q})$ by, $\gamma \in (0,1)$

$$[u]_{\gamma,\gamma/2} = \sup_{z_1,z_2\in\mathcal{Q},z_1\neq z_2} \frac{|u(z_1)-u(z_2)|}{\rho^\gamma(z_1,z_2)}.$$

Definition A.3: For $\gamma \in (0,1]$ and $k \in \mathbb{N} \cup \{0\}$, the parabolic Hölder space $C^{2k+\gamma,k+\gamma/2}(\mathcal{Q},\mathcal{R})$ is the Banach space of all functions $u \in C(\mathcal{Q},\mathcal{R})$ for which the norm

$$|u|_{2k+\gamma,k+\gamma/2} = |u|_0 + \sum_{i=1}^{k} |D_t^i u|_0 + \sum_{i=1}^{2k} \max_{|\alpha|=i} |D_x^\alpha u|_0 + \max_{|\alpha|=2k} [D_t^k D_x^\alpha u]_{\gamma,\gamma/2}.$$

is finite.

In this text, we only use $C^{2k+\gamma,k+\gamma/2}(\mathcal{Q},\mathcal{R})$ for $k=0,1$. We also need the following elementary fact between elliptic Hölder and parabolic Hölder spaces.

Proposition A.4: *If $f \in C^\delta((0,T);B_1)$ and $g \in C^\gamma(B_1^2)$ be two functions for some constants $\delta \in (0,1/2]$ and $\gamma \in (0,1]$, then $h(x,t) = g(x,f(t))$ belongs to $C^{2\delta\gamma,\delta\gamma}(B_1 \times (0,T))$ with*

$$[h]_{2\delta\gamma,\delta\gamma} \leq K[g]_\gamma([f]_\delta + 1).$$

Proof: The result follows from the following inequalities:

$$|h(x_1,t_1) - h(x_2,t_2)| = |g(x_1,f(t_1)) - g(x_2,f(t_2))| \leq [g]_\gamma(|f(t_1) - f(t_2)|^2 + |x_1 - x_2|^2)$$

Since $|f(t_1) - f(t_2)| \leq [f]_\delta |t_1 - t_2|^\delta$, $|t_1 - t_2| + |x_1 - x_2| \leq 3$, and $2\delta \leq 1$, we conclude that

$$|h(x_1,t_1) - h(x_2,t_2)| \leq K[g]_\gamma([f]_\delta + 1)(|t_1 - t_2|^{1/2} + |x_1 - x_2|)^{2\delta\gamma}.$$

■

A.2 Relations between PDEs and SDE without mean field term

We consider the density of a killed process on the unit ball B_1 given by

$$dX_t = (\bar{b}(X_t,t)\,dt + dW_t)I_{B_1}(X_t),\; X_0 \sim m_0 \tag{A2}$$

for a function $\bar{b} \in C^{2+\delta,1+\frac{\delta}{2}}(B_1 \times (0,T))$. One can define a semigroup $\{P_{s,t} : 0 \le s \le t\}$ on a Banach space $C_0 = C_0(B_1)$ by

$$P_{s,t}f(x) = \mathbb{E}^{x,s}[f(X_t)].$$

It can be checked that

- $P_{t,t} = I$;
- $P_{s,t}P_{t,r} = P_{s,r}$ for $0 \le s \le t \le r$;

Recall that the generator

$$L_t f(x) = \lim_h \frac{P_{t,t+h} - I}{h} f(x).$$

In this case, the generator can be written explicitly as [1]

$$L_t f(x) = \bar{b}(x,t)\cdot \nabla f(x) + \tfrac{1}{2}\Delta f(x)$$

and the domain $\mathcal{D}(L_t)$ of the generator includes the smooth test function set $C_0^\infty(B_1)$. Moreover, the adjoint operator of L_t is given by

$$L_t^* f(x) = -\mathrm{div}_x(\bar{b}f) + \tfrac{1}{2}\Delta f.$$

Formally, if we denote the density of X_t on B_1 by $m(t,\cdot)$, i.e.

$$P_{0,t}f(x) = (m_t, f), \quad \forall f \in C_0.$$

It is also noted that, X_t is a submarkovian on B_1, since $\int_{B_1} m_t(x)\,dx = 1 - \mathbb{P}(X_t \in \partial B_1) \le 1$. One can carry out

$$\frac{d}{dt}P_{s,t}f(x) = \lim_h \frac{P_{s,t+h} - P_{s,t}}{h} f(x) = P_{s,t}L_t f(x), \quad \forall f \in C_0^\infty.$$

Taking $s = 0$, it becomes

$$\frac{d}{dt}(m_t, f) = (m_t, L_t f), \quad \forall f \in C_0^\infty$$

which implies the Kolmogorov forward equation with appropriate initial and boundary conditions:

$$\begin{aligned} \partial_t m &= L_t^* m, \quad B_1 \times (0,T), \\ m(x,0) &= m_0(x), \quad x \in \bar{B}_1, \\ m(x,t) &= 0, \quad \partial B_1 \times (0,T). \end{aligned} \tag{A3}$$

Next, if we denote $u(x,t) = P_{t,T}g(x)$ for some fixed $g \in C_0$, and $u \in C^{1,2}$ for some fixed T, then we can write

$$\frac{d}{dt}P_{t,r}g(x) = \lim_h \frac{P_{t+h,r} - P_{t,r}}{h} g(x) = \lim_h \frac{I - P_{t,t+h}}{h} P_{t+h,r}g(x) = -L_t P_{t,r}g(x).$$

This implies, by taking $r = T$, the Kolmogorov backward equation with some terminal and boundary conditions:

$$\begin{aligned} \partial_t u + L_t u &= 0, \quad B_1 \times (0,T), \\ u(x,T) &= g(x), \quad x \in \bar{B}_1, \\ u(x,t) &= 0, \quad \partial B_1 \times (0,T). \end{aligned} \tag{A4}$$

Proposition A.5: *Assume $\bar{b} \in C^{2+\delta,1+\delta/2}((0,\infty) \times B_1)$, $m_0, g \in C_0^{\delta}(B_1)$. Then,*

(1) *The density $m(x,t)$ of X of* (A2) *on the open set B_1 is the unique solution of* (A3) *in $C^{2+\delta,1+\delta/2}$;*
(2) *$u(x,t) = P_{t,T}g(x)$ is the unique solution of* (A4) *in $C^{2+\delta,1+\delta/2}$.*

Proof: One can first write down non-divergence form of (A3). Then, the uniqueness of the solution and its regularity result of forward Equation (A3) and backward Equation (A4) directly follow from Theorem 10.3.3 of [2]. The relation of (A3) and transition density of the submarkovian process X is referred to Section 4.1 of [18]. The stochastic representation of (A4) to the function $P_{t,T}g(x)$ is referred to Section 40.2 of [1]. ■

We need the following estimate in this paper. Consider

$$\begin{aligned} \partial_t u &= L_t u + cu, \quad B_1 \times (0,T), \\ u(x,0) &= g(x), \quad x \in \bar{B}_1, \\ u(x,t) &= 0, \quad \partial B_1 \times (0,T). \end{aligned} \tag{A5}$$

Proposition A.6: *If $b, c \in C^{\delta,\delta/2}$, $g \in C_0^{2+\delta}$, and $c \le 0$, then* (A5) *is uniquely solvable satisfying*

$$|u|_{2+\delta,1+\delta/2} \le K(|b|_{\delta,\delta/2}, |c|_{\delta,\delta/2})|g|_{2+\delta}.$$

Proof: Unique solvability is implied by Theorem 10.3.3 of [16] and the estimate is given by Theorem 10.2.2 of [16]. ■