Existence of solutions to path-dependent kinetic equations and related forward - backward systems

Vassili Kolokoltsov[1] [2], **Wei Yang**[3]

Abstract

This paper is devoted to path-dependent kinetics equations arising, in particular, from the analysis of the coupled backward – forward systems of equations of mean field games. We present local well-posedness, global existence and some regularity results for these equations.

Key words: kinetic equation, global existence, path dependence, nonlinear Markov process, stable processes, coupled backward–forward systems.

1   Introduction

For a Banach space $\mathbf{B}$ we denote by $\mathbf{B}^*$ the dual Banach space of $\mathbf{B}$. The pairing between $f \in \mathbf{B}$ and $\mu \in \mathbf{B}^*$ is denoted by (f, μ). The norm in $\mathbf{B}^*$ is defined by $\|\mu\|_{\mathbf{B}^*} = \sup_{\|f\|_{\mathbf{B}} \leq 1} |(f, \mu)|$. For a $T > 0$, we denotes by $C([0, T], \mathbf{B})$ the Banach space of continuous curves $\eta_. : [0, T] \to B$ equipped with the norm $\sup_{t \in [0, T]} \|\eta_t\|_{\mathbf{B}}$.

A deterministic dynamic in $\mathbf{B}^*$ can be naturally specified by a vector-valued ordinary differential equation

$$\dot{\mu}_t = \Psi(t, \mu_t) \tag{1.1}$$

with a given initial value $\mu \in \mathbf{B}^*$, where the mapping $(t, \eta) \mapsto \Psi(t, \eta)$ is from $\mathbf{R}^+ \times \mathbf{B}^*$ to $\mathbf{B}^*$. More generally, one often meets the situations when $\dot{\mu}$ does not belong to $\mathbf{B}^*$, but to some its extension. Namely, let $\mathbf{D}$ be a dense subset of $\mathbf{B}$, which is itself a Banach space with the norm $\| \|_{\mathbf{D}} \geq \| \|_{\mathbf{B}}$. A deterministic dynamic in $\mathbf{B}^*$ can be specified by equation (1.1), where the

---

mapping $(t, \eta) \mapsto \Psi(t, \eta)$ is from $\mathbf{R}^+ \times \mathbf{B}^*$ to $\mathbf{D}^*$. Written in weak form, equation (1.1) means that, for all $f \in \mathbf{D}$,

$$(f, \dot{\mu}_t) = (f, \Psi(t, \mu_t)). \tag{1.2}$$

In many applications, equation (1.2) appears in the form

$$\frac{d}{dt}(f, \mu_t) = (A[t, \mu_t]f, \mu_t), \quad \mu_0 = \mu, \tag{1.3}$$

where the mapping $(t, \eta) \mapsto A[t, \eta]$ is from $\mathbf{R}^+ \times \mathbf{B}^*$ to bounded linear operators $A[t, \eta] : \mathbf{D} \mapsto \mathbf{B}$ such that, for each pair $(t, \eta) \in \mathbf{R}^+ \times \mathbf{B}^*$, $A[t, \eta]$ generates a strongly continuous semigroup in $\mathbf{B}$. Of major interest is the case when $\mathbf{B}^*$ is the space of measures on a locally compact space. It turns out that, in this case and under mild technical assumptions, an evolution (1.2) preserving positivity has to be of form (1.3) with the operators $A[t, \eta]$ generating Feller processes, see Theorems 6.8.1 and 11.5.1 from [7].

Equation (1.3) will be referred to as the *general kinetic equation*. It contains most of the basic equations from non-equilibrium statistical mechanics and evolutionary biology, see monograph [7] for an extensive discussion.

In this paper we are mostly interested in yet more general equation. Namely, let $\mathcal{M}$ be a closed convex subset of $\mathbf{B}^*$, which is also closed in $\mathbf{D}^*$. For a $T > 0$, let $C([0, T], \mathcal{M}(\mathbf{D}^*))$ denote a closed convex subset of $C([0, T], \mathbf{D}^*)$ consisting of curves with values in $\mathcal{M}$, and $C_\mu([0, T], \mathcal{M}(\mathbf{D}^*))$ a closed convex subset of $C([0, T], \mathcal{M}(\mathbf{D}^*))$, consisting of curves $\{\eta_.\}$ with initial data $\eta_0 = \mu \in \mathcal{M}$.

The main object of this paper is a "path-dependent" version of equation (1.3), that is

$$\frac{d}{dt}(f, \mu_t) = (A[t, \{\mu_s\}_{0 \le s \le T}]f, \mu_t), \quad \mu_0 = \mu, \tag{1.4}$$

where $(t, \{\eta_s\}_{0 \le s \le T}) \mapsto A[t, \{\eta_s\}_{0 \le s \le T}]$ maps $\mathbf{R}^+ \times C_\mu([0, T], \mathcal{M}(\mathbf{D}^*))$ to bounded linear operators $\mathbf{D} \mapsto \mathbf{B}$. We refer to equation (1.4) as the general *path-dependent kinetic equation*. It should hold for all test functions $f \in \mathbf{D}$. Compared to equation (1.4), equation (1.3) is often referred to as a *path-independent* case.

When the operators A only depend on the *history* of the trajectory of $\{\mu_.\} \in C_\mu([0, T], \mathcal{M}(\mathbf{D}^*))$, that is

$$\frac{d}{dt}(f, \mu_t) = (A[t, \{\mu_{\le t}\}]f, \mu_t), \quad \mu_0 = \mu, \tag{1.5}$$

2

we call (1.5) an *adapted kinetic equation*, where $\{\mu_{\leq t}\}$ is a shorthand for $\{\mu_s\}_{0 \leq s \leq t}$. Adapted kinetic equations can be seen as analytic analogs of stochastic differential equations with adapted coefficients, and their well-posedness can be obtained by similar methods. When the generators A only depend on the *future* of the trajectory of $\{\mu.\} \in C_\mu([0, T], \mathcal{M}(\mathbf{D}^*))$, that is

$$\frac{d}{dt}(f, \mu_t) = (A[t, \{\mu_{\geq t}\}]f, \mu_t), \quad \mu_0 = \mu, \tag{1.6}$$

we call (1.6) an *anticipating kinetic equation*, where $\{\mu_{\geq t}\}$ is a shorthand for $\{\mu_s\}_{t \leq s \leq T}$.

Remark 1.1. *The terminology of adaptiveness and anticipation here should not be associated with any randomness, as in more standard usage of these words.*

Equation (1.4) has many applications. Let us briefly explain the crucial role played by this equation in the mean field game (MFG) methodology, which is based on the analysis of coupled systems of forward – backward evolutions and which constitutes a quickly developing area of research in modern theory of optimization, see detail e.g. in [2, 3, 4, 5, 10].

Assume that the objective of an agent described by a controlled stochastic process $X(s)$ (passing through x at time t), given an evolution $\hat{\mu}.$ of the empirical distributions of a large number of other players, is to maximize (over a suitable class of controls $\{u.\}$) the payoff

$$V(t, x, \hat{\mu}_{\geq t}, u_{\geq t}) = \mathbb{E}\left[\int_t^T J(s, X(s), \hat{\mu}_s, u_s)\, ds + V^T(X(T))\right],$$

By dynamic programming the optimal payoff of such an agent

$$V(t, x, \hat{\mu}_{\geq t}) = \sup_{u.} V(t, x, \hat{\mu}_{\geq t}, u_{\geq t})$$

should satisfy certain HJB equation (backward evolution). On the other hand, when all optimal controls $\{u_t = u_t(\hat{\mu}_{\geq t})\}$ are found, the empirical measure $\mu.$ of the resulting process satisfies the controlled kinetic equation of type (1.3) (forward equation), that is

$$\frac{d}{dt}(f, \mu_t) = (A[t, \mu_t, u_t(\hat{\mu}_{\geq t})]f, \mu_t), \quad \mu_0 = \mu. \tag{1.7}$$

The main consistency condition of MFG is in the requirement that the initial $\hat{\mu}$ coincides with the resulting μ. Equalizing $\hat{\mu}. = \mu.$ in (1.7) clearly leads to anticipating kinetic equation of type (1.6).

3

Our main results concern the well-posedness of adaptive kinetic equations (1.5), the local well-posedness and global existence of anticipating and general path dependent kinetic equations and finally some regularity result for path-independent equations arising from their probabilistic interpretations.

The rest of the paper is organised as follows. In Section 2 our main results are formulated and in Section 3 they are proved. Section 4 yields some regularity results for the solutions of kinetic equations leading also to simple verifiable conditions for compactness assumption (2.12) of our main global existence result. Section 5 show some examples.

2 Main results

Let us recall the notion of propagators needed for the proper formulation of our results.

For a set S, a family of mappings $U^{t,r}$ from S to itself, parametrized by the pairs of numbers $r \leq t$ (resp. $t \leq r$) from a given finite or infinite interval is called a *(forward) propagator* (resp. a *backward propagator*) in S, if $U^{t,t}$ is the identity operator in S for all t and the following *chain rule*, or *propagator equation*, holds for $r \leq s \leq t$ (resp. for $t \leq s \leq r$):

$$U^{t,s}U^{s,r} = U^{t,r}.$$

A backward propagator $U^{t,r}$ of bounded linear operators on a Banach space $\mathbf{B}$ is called *strongly continuous* if the operators $U^{t,r}$ depend strongly continuously on t and r.

Suppose $U^{t,r}$ is a strongly continuous backward propagator of bounded linear operators on a Banach space with a common invariant domain $\mathbf{D}$. Let A_t, $t \geq 0$, be a family of bounded linear operators $\mathbf{D} \mapsto \mathbf{B}$ that are strongly continuous in t outside a set S of zero-measure in $\mathbf{R}$. Let us say that the family A_t *generates* $U^{t,r}$ on $\mathbf{D}$ if, for any $f \in \mathbf{D}$, the equations

$$\frac{d}{ds}U^{t,s}f = U^{t,s}A_s f, \quad \frac{d}{ds}U^{s,r}f = -A_s U^{s,r}f, \quad 0 \leq t \leq s \leq r, \qquad (2.1)$$

hold for all s outside S with the derivatives taken in the topology of B. In particular, if the operators A_t depend strongly continuously on t, equations (2.1) hold for all s and $f \in \mathbf{D}$, where for $s = t$ (resp. $s = r$) it is assumed to be only a right (resp. left) derivative. In the case of propagators in the space of measures, the second equation in (2.1) is called *the backward Kolomogorov equation*.

We can now formulate our main results.

Theorem 2.1 (local well-posedness for general "path-dependent" case).
Let $\mathcal{M}$ be a bounded convex subset of $\mathbf{B}^*$ with $\sup_{\mu \in \mathcal{M}} \|\mu\|_{\mathbf{B}^*} \leq K$, which is closed in the norm topologies of both $\mathbf{B}^*$ and $\mathbf{D}^*$. Suppose that

(i) the linear operators $A[t, \{\xi.\}] : \mathbf{D} \mapsto \mathbf{B}$ are uniformly bounded and Lipschitz in $\{\xi.\}$, i.e. for any $\{\xi.\}, \{\eta.\} \in C_\mu([0, T], \mathcal{M}(\mathbf{D}^*))$

$$\sup_{t \in [0,T]} \|A[t, \{\xi.\}] - A[t, \{\eta.\}]\|_{\mathbf{D} \mapsto \mathbf{B}} \leq c_1 \sup_{t \in [0,T]} \|\xi_t - \eta_t\|_{\mathbf{D}^*}, \qquad (2.2)$$

$$\sup_{t \in [0,T]} \|A[t, \{\xi.\}]\|_{\mathbf{D} \mapsto \mathbf{B}} \leq c_1 \qquad (2.3)$$

for a positive constant c_1;

(ii) for any $\{\xi.\} \in C_\mu([0, T], \mathcal{M}(\mathbf{D}^*))$, let the operator curve $A[t, \{\xi.\}] : \mathbf{D} \mapsto \mathbf{B}$ generate a strongly continuous backward propagator of bounded linear operators $U^{t,s}[\{\xi.\}]$ in $\mathbf{B}$, $0 \leq t \leq s$, on the common invariant domain $\mathbf{D}$, such that

$$\|U^{t,s}[\{\xi.\}]\|_{\mathbf{D} \mapsto \mathbf{D}} \leq c_2 \text{ and } \|U^{t,s}[\{\xi.\}]\|_{\mathbf{B} \mapsto \mathbf{B}} \leq c_3, \quad t \leq s, \qquad (2.4)$$

for some positive constants c_2, c_3, and with their dual propagators $\tilde{U}^{s,t}[\{\xi.\}]$ preserving the set $\mathcal{M}$.

Then, if

$$c_1 c_2 c_3 K T < 1, \qquad (2.5)$$

the Cauchy problem

$$\frac{d}{dt}(f, \mu_t) = (A[t, \{\mu.\}]f, \mu_t), \quad \mu_0 = \mu, \ t \in [0, T], \qquad (2.6)$$

is well posed, that is for any $\mu \in \mathcal{M}$, it has a unique solution $\Phi^t(\mu) \in \mathcal{M}$ (that is (2.6) holds for all $f \in \mathbf{D}$) that depends Lipschitz continuously on time t and the initial data in the norm of $\mathbf{D}^*$, i.e.

$$\|\Phi^s(\mu) - \Phi^t(\mu)\|_{\mathbf{D}^*} \leq c_1 c_2 (s - t), \quad 0 \leq t \leq s \leq T, \qquad (2.7)$$

and for $\mu, \eta \in \mathcal{M}$

$$\|\{\Phi^\cdot(\mu)\} - \{\Phi^\cdot(\eta)\}\|_{C([0,T], \mathcal{M}(\mathbf{D}^*))} \leq \frac{c_2}{1 - c_1 c_2 c_3 K T} \|\mu - \eta\|_{\mathbf{D}^*}. \qquad (2.8)$$

Theorem 2.2. *(global well-posedness for general "path-independent" case) Under the assumptions in Theorem 2.1, but without the locality constraint (2.5), the Cauchy problem for kinetic equation*

$$\frac{d}{dt}(f, \mu_t) = (A[t, \mu_t]f, \mu_t), \quad \mu_s = \mu \qquad (2.9)$$

5

is well-posed, i.e. for any $\mu \in \mathcal{M}$, $s \in [0,T]$, it has a unique solution $\tilde{U}^{t,s}(\mu) \in \mathcal{M}$, $t \in [s,T]$, and the transformations $\tilde{U}^{t,s}$ of $\mathcal{M}$ form a propagator depending Lipschitz continuously on time t and the initial data in the norm of $\mathbf{D}^*$, i.e.

$$\|\tilde{U}^{t,s}(\mu) - \tilde{U}^{t,s}(\eta)\|_{\mathbf{D}^*} \le c(T,K)\|\mu - \eta\|_{\mathbf{D}^*}, \qquad (2.10)$$

with a constant $c(T,K)$ depending on T and K.

Theorem 2.3. *(global wellposedness for an "adapted" case) Under the assumptions in Theorem 2.1, but without the locality constraint (2.5), the Cauchy problem*

$$\frac{d}{dt}(f, \mu_t) = (A[t, \{\mu_{\le t}\}]f, \mu_t), \quad \mu_0 = \mu, \qquad (2.11)$$

is well posed in $\mathcal{M}$ and its unique solution depends Lipschitz continuously on initial data in the norm of $\mathbf{D}^*$.

Theorem 2.4 (global existence of the solution for general "path dependent" case). *Under the assumptions in Theorem 2.1, but without the locality constraint (2.5), assume additionally that for any t from a dense subset of $[0,T]$, the set*

$$\{\tilde{U}^{t,0}[\{\xi_\cdot\}]\mu : \ \{\xi_\cdot\} \in C_\mu([0,T], \mathcal{M}(\mathbf{D}^*))\} \qquad (2.12)$$

is relatively compact in $\mathcal{M}$. Then a solution to the Cauchy problem

$$\frac{d}{dt}(f, \mu_t) = (A[t, \{\mu_s\}_{s\in[0,T]}]f, \mu_t), \quad \mu_0 = \mu, t \ge 0, \qquad (2.13)$$

exists in $\mathcal{M}$.

In Proposition 4.3 in Section 4, we give the conditions under which the compactness assumption (2.12) holds.

3 Proofs of the main results

Proof of Theorem 2.1
By duality, for any $\{\xi_\cdot^1\}, \{\xi_\cdot^2\} \in C_\mu([0,T], \mathcal{M}(\mathbf{D}^*))$

$$(f, (\tilde{U}^{t,0}[\{\xi_\cdot^1\}] - \tilde{U}^{t,0}[\{\xi_\cdot^2\}])\mu) = ((U^{0,t}[\{\xi_\cdot^1\}] - U^{0,t}[\{\xi_\cdot^2\}])f, \mu).$$

Next, we need to estimate the difference of the two propagators. Define an operator-valued function $Y(r) := U^{0,r}[\{\xi_\cdot^2\}]U^{r,t}[\{\xi_\cdot^1\}]$. Since $U^{t,t}[\{\xi_\cdot^i\}], i =$

1, 2, are identity operators, $Y(t) = U^{0,t}[\{\xi_\cdot^2\}]$ and $Y(0) = U^{0,t}[\{\xi_\cdot^1\}]$. By (2.1), we get

$$
U^{0,t}[\{\xi_\cdot^2\}] - U^{0,t}[\{\xi_\cdot^1\}] = U^{0,r}[\{\xi_\cdot^2\}]U^{r,t}[\{\xi_\cdot^1\}]\big|_{r=0}^{t}
$$
$$
= \int_0^t \frac{d}{dr}\left(U^{0,r}[\{\xi_\cdot^2\}]U^{r,t}[\{\xi_\cdot^1\}]\right)dr
$$
$$
= \int_0^t \left(U^{0,r}[\{\xi_\cdot^2\}]A[r,\{\xi_\cdot^2\}]U^{r,t}[\{\xi_\cdot^1\}] - U^{0,r}[\{\xi_\cdot^2\}]A[r,\{\xi_\cdot^1\}]U^{r,t}[\{\xi_\cdot^1\}]\right)ds
$$
$$
= \int_0^t U^{0,r}[\{\xi_\cdot^2\}](A[r,\{\xi_\cdot^2\}] - A[r,\{\xi_\cdot^1\}])U^{r,t}[\{\xi_\cdot^1\}]ds.
$$

Then, together with assumptions (2.2) and (2.4),

$$
\|(\tilde{U}^{t,0}[\{\xi_\cdot^1\}] - \tilde{U}^{t,0}[\{\xi_\cdot^2\}])\mu)\|_{\mathbf{D}^*}
$$
$$
\leq \|U^{0,t}[\{\xi_\cdot^1\}] - U^{0,t}[\{\xi_\cdot^2\}]\|_{\mathbf{D}\mapsto\mathbf{B}}\|\mu\|_{\mathbf{B}^*} \tag{3.1}
$$
$$
\leq c_1 c_2 c_3 t K \|\{\xi_\cdot^1\} - \{\xi_\cdot^2\}\|_{C([0,T],\mathcal{M}(\mathbf{D}^*))}.
$$

Consequently, if (2.5) holds, the mapping $\{\xi_\cdot\} \mapsto \{\tilde{U}^{t,0}[\{\xi_\cdot\}]\}_{t\in[0,T]}$ is a contraction in $C_\mu([0,T],\mathcal{M}(\mathbf{D}^*))$. Hence by the contraction principle there exists a unique fixed point for this mapping and hence a unique solution to equation (2.6).

Inequality (2.7) follows directly from (2.6). Finally, if $\Phi^t(\mu) = \mu_t$ and $\Phi^t(\eta) = \eta_t$, then

$$
\mu_t - \eta_t = \tilde{U}^{t,0}[\{\mu_\cdot\}]\mu - \tilde{U}^{t,0}[\{\eta_\cdot\}]\eta
$$
$$
= (\tilde{U}^{t,0}[\{\mu_\cdot\}] - \tilde{U}^{t,0}[\{\eta_\cdot\}])\mu + \tilde{U}^{t,0}[\{\eta_\cdot\}](\mu - \eta).
$$

From (2.4) and (3.1),

$$
\|\{\mu_\cdot\} - \{\eta_\cdot\}\|_{C([0,T],\mathcal{M}(\mathbf{D}^*))} \leq c_1 c_2 c_3 T K \|\{\mu_\cdot\} - \{\eta_\cdot\}\|_{C([0,T],\mathcal{M}(\mathbf{D}^*))} + c_2 \|\mu - \eta\|_{\mathbf{D}^*} \tag{3.2}
$$

implying (2.8). $\qquad\square$

Proof of Theorem 2.2 The global unique solution of (2.9) is constructed by extending local unique solutions of (2.6) via iterations, as is routinely performed in the theory of ordinary differential equations (ODE). $\quad\square$

Proof of Theorem 2.3

For a $\mu \in \mathcal{M}$, let us construct an approximating sequence $\{\xi_\cdot^n\} \in C([0,T], \mathcal{M}(\mathbf{D}^*))$, $n = 0, 1, \cdots$, by defining $\xi_t^0 = \mu$ for $t \in [0,T]$ and then recursively

$$\xi_t^n = \tilde{U}^{t,0}[\xi_{\leq t}^{n-1}]\mu, \quad \forall t \in [0,T].$$

By non-anticipation, arguing as in the proof of (3.1) above, we first get the estimate

$$\sup_{0 \leq r \leq t} ||\xi_r^1 - \xi_r^0||_{\mathbf{D}^*} \leq c_1 c_2 c_3 K t,$$

and then recursively

$$\sup_{0 \leq r \leq t} ||\xi_r^n - \xi_r^{n-1}||_{\mathbf{D}^*} \leq c_1 c_2 c_3 K \int_0^t \sup_{0 \leq r \leq s} ||\xi_r^{n-1} - \xi_r^{n-2}||_{\mathbf{D}^*} \, ds$$

that implies (by straightforward induction) that, for all $t \in [0,T]$,

$$||\{\xi_\cdot^n\} - \{\xi_\cdot^{n-1}\}||_{C([0,t], \mathcal{M}(\mathbf{D}^*))} = \sup_{0 \leq r \leq t} ||\xi_r^n - \xi_r^{n-1}||_{\mathbf{D}^*} \leq \frac{1}{n!}(c_1 c_2 c_3 K t)^n.$$

Hence, the partial sums on the r.h.s. of the obvious equation

$$\xi_{\leq t}^n = (\xi_{\leq t}^n - \xi_{\leq t}^{n-1}) + \cdots + (\xi_{\leq t}^1 - \xi_{\leq t}^0) + \xi_{\leq t}^0$$

converge, and thus the sequence $\xi_\cdot^n$ converges in $C([0,T], \mathbf{D}^*)$. The limit is clearly a solution to (2.11).

To prove uniqueness and continuous dependence on the initial condition, let us assume that μ_t and η_t are some solutions with the initial conditions μ and η respectively. Instead of (3.2), we now get

$$||\{\mu_\cdot\} - \{\eta_\cdot\}||_{C([0,t], \mathcal{M}(\mathbf{D}^*))} \leq c_1 c_2 c_3 K \int_0^t ||\{\mu_\cdot\} - \{\eta_\cdot\}||_{C([0,s], \mathcal{M}(\mathbf{D}^*))} \, ds + c_2 ||\mu - \eta||_{\mathbf{D}^*}.$$

By Gronwall's lemma, this implies

$$||\{\mu_\cdot\} - \{\eta_\cdot\}||_{C([0,t], \mathcal{M}(\mathbf{D}^*))} \leq c_2 ||\mu - \eta||_{\mathbf{D}^*} e^{c_1 c_2 c_3 K t}, \quad t \in [0,T]$$

yielding uniqueness and Lipchitz continuity of solutions with respect to initial data. $\square$

Proof of Theorem 2.4

Since $\mathcal{M}$ is convex, the space $C_\mu([0,T], \mathcal{M}(\mathbf{D}^*))$ is also convex. Since the dual operators $\tilde{U}^{t,0}[\{\xi_\cdot\}]$ preserve the set $\mathcal{M}$, for any $\{\xi_\cdot\} \in C_\mu([0,T], \mathcal{M}(\mathbf{D}^*))$, the curve $\tilde{U}^{t,0}[\{\xi_\cdot\}]\mu$ belongs to $C_\mu([0,T], \mathcal{M}(\mathbf{D}^*))$ as a function of t. Hence,

the mapping $\{\xi.\} \to \{\tilde{U}^{t,0}[\{\xi.\}]\mu, t \in [0,T]\}$ is from $C_\mu([0,T], \mathcal{M}(\mathbf{D}^*))$ to itself. Moreover, by (3.1), this mapping is Lipschitz continuous.

Denote $\hat{C} = \{\{\tilde{U}^{\cdot,0}[\{\xi.\}]\mu\} : \{\xi.\} \in C_\mu([0,T], \mathcal{M}(\mathbf{D}^*))\}$. Together with (2.7), the assumption that set (2.12) is compact in $\mathcal{M}$ for any t from a dense subset of $[0,T]$ implies that the set $\hat{C}$ is relatively compact in $C_\mu([0,T], \mathcal{M}(\mathbf{D}^*))\}$ (Arzela-Ascoli Theorem, see A.21 in [6]).

Finally, by Schauder fixed point theorem, there exists a fixed point in $\hat{C} \subset C_\mu([0,T], \mathcal{M}(\mathbf{D}^*))$, which gives the existence of a solution to (2.13). $\square$

4 Nonlinear Markov evolutions and its regularity

This section is designed to provide a probabilistic interpretation and, as a consequence, certain regularity properties for nonlinear Markov evolution μ_t solving kinetic equation (2.9) in the case when $\mathbf{B} = C_\infty(\mathbf{R}^d)$ and $\mathcal{M} = \mathcal{P}(\mathbf{R}^d)$ is the set of probability measures on $\mathbf{R}^d$, so that $\mathbf{B}^*$ is the space of signed Borel measures on $\mathbf{R}^d$ and $K = \sup_{\mu \in \mathcal{P}(\mathbf{R}^d)} \|\mu\|_{\mathbf{B}^*} = 1$. As a consequence, w shall present a simple criterion for the main compactness assumption of Theorem 2.4.

We will use the following notations.

$C_{Lip}(\mathbf{R}^d)$ is the Banach space of bounded Lipschitz continuous functions f on $\mathbf{R}^d$ with the norm $\|f\|_{Lip} = \sup_x |f(x)| + \sup_{x \neq y} \frac{|f(x)-f(y)|}{|x-y|}$.

$C_\infty(\mathbf{R}^d)$ is the Banach space of bounded continuous functions f on $\mathbf{R}^d$ with $\lim_{x \to \infty} f(x) = 0$, equipped with sup-norm $\|f\|_{C_\infty(\mathbf{R}^d)} := \sup_x |f(x)|$.

$C_\infty^1(\mathbf{R}^d)$ is the Banach space of continuously differentiable and bounded functions f on $\mathbf{R}^d$ such that the derivative f' belongs to $C_\infty(\mathbf{R}^d)$, equipped with the norm $\|f\|_{C_\infty^1(\mathbf{R}^d)} := \sup_x |f'(x)|$.

$C_\infty^2(\mathbf{R}^d)$ is the Banach space of twice continuously differentiable and bounded functions f on $\mathbf{R}^d$ such that the first derivative f' and the second derivative f'' belong to $C_\infty(\mathbf{R}^d)$, equipped with the norm $\|f\|_{C_\infty^2(\mathbf{R}^d)} := \sup_x(|f'(x)| + |f''(x)|)$.

Let $\{A[t,\mu] : t \geq 0, \mu \in \mathcal{P}(\mathbf{R}^d)\}$ be a family of operators in $C_\infty(\mathbf{R}^d)$ of the Lévy-Khintchin type, that is

$$
\begin{aligned}
A[t,\mu]f(z) = \frac{1}{2}(G(t,z,\mu)\nabla, \nabla)f(z) + (b(t,z,\mu), \nabla f(z)) \\
+ \int (f(z+y) - f(z) - (\nabla f(z), y)\mathbf{1}_{B_1}(y))\nu(t,z,\mu,dy),
\end{aligned}
\tag{4.1}
$$

where ∇ denotes the gradient operator; for $(t,z,\mu) \in [0,T] \times \mathbf{R}^d \times \mathcal{P}(\mathbf{R}^d)$, $G(t,z,\mu)$ is a symmetric non-negative matrix, $b(t,z,\mu)$ is a vector, $\nu(t,z,\mu,\cdot)$

is a Lévy measure on $\mathbf{R}^d$, i.e.

$$\int_{\mathbf{R}^d} \min(1, |y|^2)\nu(t, z, \mu, dy) < \infty, \quad \nu(t, z, \mu, \{0\}) = 0, \qquad (4.2)$$

depending measurably on t, z, μ, and $\mathbf{1}_{B_1}$ denotes, as usual, the indicator function of the unit ball in $\mathbf{R}^d$. Assume that each operator (4.1) generates a Feller process with one and the same domain $\mathbf{D}$ such that $C_\infty^2(\mathbf{R}^d) \subseteq \mathbf{D} \subseteq C_\infty^1(\mathbf{R}^d)$.

Proposition 4.1. *Suppose the assumptions of Theorem 2.2 are fulfilled with generators $A[t, \mu]$ of type (4.1) and a probability measure μ is given. Then there exists a family of processes $\{X_{s,t}^\mu : \mu \in \mathcal{P}(\mathbf{R}^d)\}$ defined on a certain filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathcal{P}(\mathbf{R}^d))$ such that $\mu_t = \mathcal{L}(X_{s,t}^\mu)$ solves the Cauchy problem for equation (2.9) with initial condition μ and $\{X_{s,t}^\mu\}$ solves the nonlinear martingale problem, specified by the family $\{A[t, \mu]\}$, that is, for any $f \in \mathbf{D}$,*

$$f(X_{s,t}^\mu) - \int_s^t A[\tau, \mathcal{L}(X_{s,\tau}^\mu)]f(X_{s,\tau}^\mu)d\tau, \quad s \le t \qquad (4.3)$$

is a martingale.

Proof. By the assumptions of Theorem 2.2, a solution $\mu_t \in \mathcal{P}(\mathbf{R}^d)$ of equation (2.9) with initial condition $\mu_s = \mu$ specifies a propagator $\tilde{U}^{t,r}[\mu.]$, $s \le r \le t$, of linear transformations in $\mathbf{B}^*$, solving the Cauchy problems for equation

$$\frac{d}{dt}(f, \nu_t) = (A[t, \mu_t]f, \nu_t). \qquad (4.4)$$

In its turn, for any $\nu \in \mathcal{P}(\mathbf{R}^d)$, equation (4.4) specifies marginal distributions of a usual (linear) Markov process $\{X_{s,t}^\mu(\nu)\}$ in $\mathbf{R}^d$ with the initial measure ν. Clearly, the process $\{X_{s,t}^\mu(\mu)\}$ is a solution to our martingale problem. $\square$

We shall refer to the family of processes constructed in Proposition 4.1 as to *nonlinear Markov process* generated by the family $A[t, \mu]$.

Using martingales allows us to prove the following useful regularity property for the solution of kinetic equations.

Proposition 4.2. *Suppose the assumptions of Theorem 2.2 are fulfilled for a kinetic equation of "path-independent" type (2.9) with generators $A[t, \mu]$ of type (4.1). Let $\{X_{s,t}^\mu\}$ denote a nonlinear Markov process constructed from the family of generators $A[t, \mu]$ by Proposition 4.1. Assume, for $p \in (0, 2]$*

10

and $P > 0$, the following boundedness condition holds:

$$\sup_{x \in R^d, \, t \geq 0, \, \mu \in \mathcal{M}} \max \left\{ |G(t, x, \mu)|, |b(t, x, \mu)|, \int \min(|y|^2, |y|^p) \nu(t, x, \mu, dy) \right\} \leq P,$$
(4.5)

and the initial measure $\mu_s = \mu$ has a finite pth order moment, i.e.

$$\int |x|^p \mu(dx) = p_\mu < \infty.$$

Then the distributions $\mathcal{L}(X_{s,t}^\mu) = \Phi^{t,s}(\mu)$, solving the Cauchy problem for equation (2.9) with initial condition μ_s have uniformly bounded pth moments, i.e.

$$\int |x|^p \Phi^{t,s}(\mu)(dx) \leq c(T, P)[1 + p_\mu],$$
(4.6)

and are $\frac{1}{2}$-Hölder continuous with respect to t in the space $(C_{Lip}(\mathbf{R}^d))^*$, i.e.

$$\|\Phi^{t_1,s}(\mu) - \Phi^{t_2,s}(\mu)\|_{(C_{Lip}(\mathbf{R}^d))^*} \leq c(T, P)\sqrt{|t_1 - t_2|}, \quad \forall t_1, t_2 \geq s \geq 0, \ (4.7)$$

with a positive constant c.

Proof. For a fixed trajectory $\{\mu_t\}_{t \geq 0}$ with initial value μ, one can consider $\{X_{s,t}^\mu\}$ as a usual Markov process. Using the estimates for the moments of such processes from formula (5.61) of [8] (more precisely, its straightforward extension to time non-homogeneous case), one obtains from (4.5) that

$$\mathbb{E} \left[\min \left(|X_{s,t}^\mu - \hat{x}|^2, |X_{s,t}^\mu - \hat{x}|^p \right) | X_{s,s}^\mu = \hat{x} \right) \right] \leq e^{C(T,P)(t-s)} - 1.$$
(4.8)

This implies (4.6). Moreover, (4.8) implies that

$$\mathbb{E} \left[|X_{s,t}^\mu - \hat{x}| \mathbf{1}_{|X_{s,t}^\mu - \hat{x}| \leq 1} | X_{s,s}^\mu = \hat{x} \right] \leq (e^{C(T,P)(t-s)} - 1)^{1/2} \leq C(T, P)\sqrt{t - s}$$
(4.9)

$$\mathbb{E} \left[|X_{s,t}^\mu - \hat{x}| \mathbf{1}_{|X_{s,t}^\mu - \hat{x}| \geq 1} | X_{s,s}^\mu = \hat{x} \right] \leq (e^{C(T,P)(t-s)} - 1)^{1/p} \leq C(T, P)(t - s)^{1/p},$$
(4.10)

and consequently

$$\mathbb{E} \left(|X_{s,t}^\mu - \hat{x}| \, | X_{s,s}^\mu = \hat{x} \right) \leq C(T, P)\sqrt{t - s},$$
(4.11)

where constants $C(T, P)$ can have different values in various formulas above.

Since $\Phi^{t,s}(\mu)$ is the distribution law of the process $\{X^\mu_{s,t}\}$,

$$
\begin{aligned}
||\Phi^{s,t_1}(\mu) - \Phi^{s,t_2}(\mu)||_{(C_{Lip}(\mathbf{R}^d))^*} &= \sup_{||\psi||_{C_{Lip}(\mathbf{R}^d)} \leq 1} \left| \mathbb{E}\psi(X^\mu_{s,t_1}) - \mathbb{E}\psi(X^\mu_{s,t_2}) \right| \\
&\leq \sup_{||\psi||_{C_{Lip}(\mathbf{R}^d)} \leq 1} \mathbb{E}\left| \psi(X^\mu_{s,t_1}) - \psi(X^\mu_{s,t_2}) \right| \\
&\leq \mathbb{E}\left| X^\mu_{s,t_1} - X^\mu_{s,t_2} \right|.
\end{aligned}
$$
(4.12)

From (4.11), (4.12) and Markov property, we get (4.7) as required. □

Remark 4.1. *For the case of diffusions with $G = 1$, Hölder continuity (4.7) was proved in [4].*

Our main purpose for presenting Proposition 4.2 lies in the following corollary.

Proposition 4.3. *Under the assumptions of Theorem 2.1 for generators $A[t, \{\mu.\}]$ of Lévy-Khintchin type (4.1), but without locality condition (2.5), suppose the boundedness condition (4.5) holds for some $p \in (0, 2]$ and $P > 0$. Then the compactness condition from Theorem 2.4 (stating that set (2.12) is compact in $\mathcal{P}(\mathbf{R}^d)$) holds for any initial measure μ with a finite moment of pth order.*

Proof. It follows from (4.6) and an observation that a set of probability laws on $\mathbf{R}^d$ with a bounded pth moment, $p > 0$, is tight and hence relatively compact. □

5   Basic examples of operators $A[t, \mu]$

In this section, we present some basic examples of generators that fit to assumptions of our main Theorems and are relevant to the study of mean field games.

Notice that the most nontrivial condition of Theorem 2.1 is (ii), as it concerns the difficult question from the theory of usual Markov process, on when a given pre-generator of Lévy-Khintchine type does really generate a Markov process. Even more difficult is the situation with time-dependent generators, as the standard semigroup methods (resolvents and Hille-Phillips-Iosida theorem) are not applicable.

Example 5.1. *Nonlinear Lévy processes are specified by a families of generators of type (4.1) such that all coefficients do not depend on z, i.e.*

$$A[t,\mu]f(x) = \frac{1}{2}(G(t,\mu)\nabla,\nabla)f(x) + (b(t,\mu),\nabla f)(x)$$
$$+ \int [f(x+y) - f(x) - (y,\nabla f(x))\mathbf{1}_{B_1}(y)]\nu(t,\mu,dy).$$

The following statement is a consequence of Proposition 7.1 from [7].

Proposition 5.1. *Supposed that the coefficients G, b, ν are continuous in t and Lipschitz continuous in μ in the norm of Banach space $(C^2_\infty(\mathbf{R}^d))^*$, i.e.*

$$\|G(t,\mu) - G(t,\eta)\| + \|b(t,\mu) - b(t,\eta)\| + \int \min(1,|y|^2)|\nu(t,\mu,dy) - \nu(t,\eta,dy)|$$
$$\leq c\|\mu - \eta\|_{(C^2_\infty(\mathbf{R}^d))^*}, \quad \forall t \geq 0,$$

with a positive constant c, then condition (ii) of Theorem 2.1 holds with $\mathbf{D} = C^2_\infty(\mathbf{R}^d)$.

Notice that the most natural examples of a functional F on measures that are Lipschitz continuous (or even smooth) in space $(C^2_\infty(\mathbf{R}^d))^*$ are supplied by smooth functions of monomials $\int g(x_1, \cdots, x_n)\mu(dx_1)\cdots\mu(dx_n)$ with sufficient smooth functions g.

Example 5.2. *McKean-Vlasov diffusion are specified by the following stochastic differential equation*

$$dX_t = b(t, X_t, \mu_t)dt + \sigma(t, X_t, \mu_t)dW_t,$$

where drift coefficient $b : \mathbf{R}^+ \times \mathbf{R}^d \times \mathcal{P}(\mathbf{R}^d) \to \mathbf{R}^d$, diffusion coefficient $\sigma : \mathbf{R}^+ \times \mathbf{R}^d \times \mathcal{P}(\mathbf{R}^d) \to \mathbf{R}^d$ and W_t is a standard Brownian motion. The corresponding generator is given by

$$A[t,\mu]f(x) = (b(t,x,\mu),\nabla f(x)) + \frac{1}{2}(G(t,x,\mu),\nabla^2 f(x)), \quad f \in C^2_\infty(\mathbf{R}^d),$$

where $G(t,x,\mu) = tr\{\sigma(t,x,\mu)\sigma^T(t,x,\mu)\}$. It is well known (and follows from Ito's calculus) that if the coefficients of a diffusion are Lipshitz continuous, the corresponding SDE is well posed, implying the following.

Proposition 5.2. *If G, b are continuous in t, Lipshitz continuous in x and Lipshitz continuous in μ in the topology of $(C^2_\infty(\mathbf{R}^d))^*$, then the condition (ii) of Theorem 2.1 is satisfied.*

Example 5.3. *Nonlinear stable-like processes (including tempered ones) are specified by the families*

$$A[t,\mu]f(x) = (b(t,x,\mu), \nabla f(x)) + \int (f(x+y) - f(x))\nu(t,x,\mu,dy)$$

$$+ \int_0^K d|y| \int_{S^{d-1}} a(t,x,s) \frac{f(x+y) - f(x) - (y, \nabla f(x))}{|y|^{\alpha_t(x,s)+1}} \omega_t(ds).$$

$$(5.1)$$

Here $s = y/|y|$, $K > 0$, ω_t are certain finite Borel measures on S^{d-1} and $\nu(t,x,\mu,dy)$ are finite measures, a, α are positive bounded functions with $\alpha \in (0,2)$.

The following result is a corollary of (a straightforward time-nonhomogeneous extension of) Proposition 4.6.2 of [8].

Proposition 5.3. *If all coefficients are continuous in t, a, α are C^1-functions in x, s, b and ν are Lipshitz continuous in x and μ (with μ taken in the topology $(C_\infty^2(\mathbf{R}^d))^*$, then all conditions of Theorem 2.1 are satisfied with $\mathbf{D} = C_\infty^2(\mathbf{R}^d)$.*

Example 5.4. *Processes of order at most one are specified by the families*

$$A[t,\mu]f(x) = (b(t,x,\mu), \nabla f(x)) + \int_{\mathbf{R}^d} (f(x+y) - f(x))\nu(t,x,\mu,dy),$$

with the Lévy measures ν having finite first moment $\int |y|\nu(t,x,\mu,dy)$. The next result is established in Theorem 4.17 of [7].

Proposition 5.4. *If b, ν are continuous in t and Lipschitz continuous in μ, i.e.*

$$||b(t,x,\mu) - b(t,x,\eta)|| + \int |y|\nu(t,x,\mu,dy) - \nu(t,x,\eta,dy)|$$

$$\leq c||\mu - \eta||_{(C_\infty^1(\mathbf{R}^d))^*}, \qquad \forall t \geq 0, x \in \mathbf{R}^d$$

and Lipshitz continuous in x, then condition (ii) of Theorem 2.1 is satisfied with $\mathbf{D} = C_\infty^1(\mathbf{R}^d)$.

The generators of order at most one describe a variety of models including spatially homogeneous and mollified Boltzmann equation and interacting α-stable laws with $\alpha < 1$.

Example 5.5. *Pure jump processes are specified by integral generators of the form*

$$A[t, \mu]f(x) = \int_{\mathbf{R}^d} (f(y) - f(x))\nu(t, x, \mu, dy). \qquad (5.2)$$

If measures $\nu(t, x, \mu_t, u, dy)$ are uniformly bounded, the conditions of Theorem 2.1 are satisfied with $\mathbf{D} = C_\infty(\mathbf{R}^d)$. For unbounded rates we refer to [7] (and references therein) for a detailed discussion.

Let us note finally that not all interesting evolution of type (1.3) satisfy the Lipschitz continuity assumption used in our main results. For instance, a different type of continuity should be applied for coefficients depending on measures via their quantiles, e.g. value at risk (VAR). This type of evolution is analyzed in [9] inspired by preprint [1].

References

[1] D. Crisan, Th. Kurtz and Y. Lee. Conditional distributions, exchangeable particle systems, and stochastic partial differential equations. Preprint 2012.

[2] O. Guéant, J.-M. Lasry, P.-L. Lions. Mean Field Games and Applications. Paris-Princeton Lectures on Mathematical Finance 2010, Springer, p. 205-266.

[3] M. Huang, R.P. Malhamé and P.E. Caines. Nash equilibria for large-population linear stochastic systens with weakly coupled agents. In: E.K. Boukas, R. P. Malhamé (Eds). Analysis, Control and Optimization of Complex Dynamic Systems. Springer 2005, 215-252.

[4] 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, Communications in information and systems, **6** (2006), 221-252.

[5] M. Huang, P. E. Caines and R. P. Malhamé. The NCE (mean field) principle with locality dependent cost interactions. IEEE Trans. Automat. Control, **55:12** (2010), 2799-2805.

[6] O. Kallenberg. Foundations of Modern Probability, Springer-Verlag, New York.

[7] V. N. Kolokoltsov. Nonlinear Markov processes and kinetic equations. Cambridge Tracks in Mathematics 182, Cambridge Univ. Press, 2010.

[8] V. N. Kolokoltsov. Markov processes, semigroups and generators, De Gryuter, 2011

[9] V. N. Kolokoltsov 2012. Nonlinear diffusions and stable-like processes with coefficients depending on the median or VaR http://arxiv.org/abs/1207.5925. To appear in 'Applied Mathematics and Optimization'.

[10] V. N. Kolokoltsov, Jiajie Li, Wei Yang. Mean field games and nonlinear Markov processes, arXiv:1112.3744v2, 2011.