Markov processes and martingale problems*

Markus Fischer, University of Padua

May 4, 2012

1 Introduction

In the late 1960s, D.W. Stroock and S.R.S. Varadhan introduced a way of characterizing Markov processes, the martingale problem approach, which is based on a mixture of probabilistic and analytic techniques. A process, or rather its law, is determined by the martingale property of a certain class of test processes. Those test processes are defined in terms of an operator that, in the case of a Markov process, is a restriction of the infinitesimal generator of the corresponding Markov semigroup. The martingale problem approach is particularly useful when dealing with questions of approximation and convergence in law for stochastic processes. The approach also allows to establish the Markov property for a given process.

The purpose of these notes is to present basic ideas and results of the Stroock-Varadhan theory, first in connection with Markov processes taking values in a locally compact separable metric space, then in connection with a class of continuous processes, namely Itô diffusions or solutions of Itô stochastic differential equations. In Section 2, we recall the definition of martingale and related notions and state the key observation of the theory, introducing the kind of processes that will serve as test processes. In Section 3, the definition of solution of the martingale problem for an operator is given together with results on existence and uniqueness of solutions; the connection with the Markov property is discussed as well. In Section 4, the martingale problem formulation for solutions of Itô stochastic differential equations is presented;

*Notes for the third and last part of a mini course on Markov processes, Graduate School in Mathematics, University of Padua, February / March 2012. Parts one and two of the course were given by Prof.s Paolo Dai Pra and Paolo Guiotto.

the relationship between the respective concepts of existence and uniqueness of solutions is explained. The martingale formulation allows, in particular, to obtain existence of solutions of stochastic differential equations under mild assumptions on the coefficients, and to establish the Markov property. Two applications to specific problems are sketched in Subsections 5.1 and 5.2, respectively. Results on weak convergence of probability measures are collected in the appendix.

The main source for Sections 2 and 3 is Chapter 4 in Ethier and Kurtz [1986], while Section 4 is based on Chapter 5, especially Section 5.4, in Karatzas and Shreve [1991]; also see the work by Stroock and Varadhan [1979] and the references therein. The book by Klenke [2008] is a useful reference on background material. Subsection 5.1 is based on Example 26.29 in Klenke [2008, pp. 584-586]. The martingale approach in connection with convergence of empirical measures sketched in Subsection 5.2 goes back at least to Oelschläger [1984].

Table 1: Notation

$\mathcal{X}$	a topological space
$\mathcal{B}(\mathcal{X})$	σ-algebra of Borel sets over $\mathcal{X}$, i.e., smallest σ-algebra containing all open (closed) subsets of $\mathcal{X}$
$\mathcal{P}(\mathcal{X})$	space of probability measures on $\mathcal{B}(\mathcal{X})$, endowed with topology of weak convergence of probability measures
$\boldsymbol{C}(\mathcal{X})$	space of all continuous functions $\mathcal{X} \to \mathbb{R}$
$\boldsymbol{C}_b(\mathcal{X})$	space of all bounded continuous functions $\mathcal{X} \to \mathbb{R}$
$\boldsymbol{C}_c(\mathcal{X})$	space of all continuous functions $\mathcal{X} \to \mathbb{R}$ with compact support
E or (E, d)	locally compact separable metric space
$\boldsymbol{B}(E)$	Banach space (under sup norm) of all bounded Borel measurable functions $E \to \mathbb{R}$
$\boldsymbol{C}_0(E)$	Banach space (under sup norm) of all bounded continuous functions $E \to \mathbb{R}$ vanishing at infinity
$\boldsymbol{C}^k(\mathbb{R}^d)$	space of all continuous functions $\mathbb{R}^d \to \mathbb{R}$ with continuous partial derivatives up to order k
$\boldsymbol{C}_c^k(\mathbb{R}^d)$	space of all continuous functions $\mathbb{R}^d \to \mathbb{R}$ with compact support and continuous partial derivatives up to order k
$\mathbb{T}$	a subset of $\mathbb{R}$, usually $[0, T]$ or $[0, \infty)$

$\boldsymbol{C}(\mathbb{T},\mathcal{X})$	space of all continuous functions $\mathbb{T}\to\mathcal{X}$
$\boldsymbol{D}(\mathbb{T},\mathcal{X})$	space of all càdlàg functions $\mathbb{T}\to\mathcal{X}$ (i.e., functions continuous from the right with limits from the left)
$\wedge$	minimum (as binary operator)

2 Martingales and a key observation

Let us first recall some basics on martingales. A *stochastic basis* is a pair $((\Omega,\mathcal{F},\mathbf{P}),(\mathcal{F}_t)_{t\in\mathbb{T}})$ where $(\Omega,\mathcal{F},\mathbf{P})$ is a probability space and $(\mathcal{F}_t)_{t\in\mathbb{T}}$ a filtration in $\mathcal{F}$.

Definition 2.1. A real-valued process $(X(t))_{t\in\mathbb{T}}$ defined on a stochastic basis $((\Omega,\mathcal{F},\mathbf{P}),(\mathcal{F}_t)_{t\in\mathbb{T}})$ is called a *martingale* with respect to $(\mathcal{F}_t)$ if

(i) $X(t)\in L^1(\Omega,\mathcal{F}_t,\mathbf{P})$ for all $t\in\mathbb{T}$;

(ii) $\mathbf{E}\left[X(t)|\mathcal{F}_s\right]=X(s)$ $\mathbf{P}$-almost surely for all $s,t\in\mathbb{T}$ with $s\leq t$.

A real-valued process X is called a martingale if X is a martingale with respect to $(\mathcal{F}_t^X)$, the filtration generated by X.

Examples of martingales are the simple one-dimensional random walk (with $\mathbb{T}=\mathbb{N}_0$), the compensated Poisson process $(N(t)-\lambda t)_{t\geq 0}$, where $N(.)$ is a standard Poisson process with intensity $\lambda>0$, and Brownian motion (or Wiener process). If W is a one-dimensional Wiener process, then $W(.)$ and $(W(t)^2-t)_{t\geq 0}$ are both martingales.

Let $\mathbb{T}=[0,T]$ for some $T>0$, let $((\Omega,\mathcal{F},\mathbf{P}),(\mathcal{F}_t)_{t\in\mathbb{T}})$ be a stochastic basis, and let $Y\in L^1(\Omega,\mathcal{F}_T,\mathbf{P})$. Set $Y(t)\doteq\mathbf{E}\left[Y|\mathcal{F}_t\right]$, $t\in[0,T]$. Then $(Y(t))_{t\in[0,T]}$ is a martingale with respect to $(\mathcal{F}_t)$.

Definition 2.2. A real-valued process $(X(t))_{t\in\mathbb{T}}$ defined on a stochastic basis $((\Omega,\mathcal{F},\mathbf{P}),(\mathcal{F}_t)_{t\in\mathbb{T}})$ is called a *submartingale* with respect to $(\mathcal{F}_t)$ if

(i) $X(t)\in L^1(\Omega,\mathcal{F}_t,\mathbf{P})$ for all $t\in\mathbb{T}$;

(ii) $\mathbf{E}\left[X(t)|\mathcal{F}_s\right]\geq X(s)$ $\mathbf{P}$-almost surely for all $s,t\in\mathbb{T}$ with $s\leq t$.

A real-valued process X is called a *supermartingale* if the process $-X$ is a submartingale. The following two facts explain the terminology.

A submartingale stays below the martingale: Let $(X(t))_{t\in[0,T]}$ be a submartingale with respect to $(\mathcal{F}_t)$. Set $Y(t)\doteq\mathbf{E}\left[X(T)|\mathcal{F}_t\right]$. Then $(Y(t))_{t\in[0,T]}$ is a martingale and $X(t)\leq Y(t)$ for all $t\in[0,T]$.

Let $f \in \boldsymbol{C}(\mathbb{R}^d)$ be *subharmonic* (i.e., $\Delta f \geq 0$) such that ∇f, Δf have subexponential growth. Let W be standard d-dimensional Brownian motion (with respect to a filtration $(\mathcal{F}_t)$). Then $(f(W(t)))_{t\geq 0}$ is a submartingale (with respect to $(\mathcal{F}_t)$).

A useful tool in stochastic calculus is localization by stopping times. There is a corresponding notion for martingales.

Definition 2.3. A real-valued process $(X(t))_{t\in\mathbb{T}}$ defined on a stochastic basis $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t)_{t\in\mathbb{T}})$ is called a *local martingale* with respect to $(\mathcal{F}_t)$ if there exists a non-decreasing sequence $(\tau_n)_{n\in\mathbb{N}}$ of $(\mathcal{F}_t)$-stopping times such that $\tau_n \nearrow \infty$ $\mathbf{P}$-almost surely and, for every $n \in \mathbb{N}$, $(X(\tau_n \wedge t))_{t\in\mathbb{T}}$ is an $(\mathcal{F}_t)$-martingale.

From now on $\mathbb{T} = [0, \infty)$. Let $X = ((\Omega, \mathcal{F}), (\mathcal{F}_t)_{t\geq 0}, (\mathbf{P}_x)_{x\in E}, (X(t))_{t\geq 0})$ be an E-valued *Markov family*, that is,

(i) $(\Omega, \mathcal{F})$ is a measurable space and $(\mathcal{F}_t)$ a filtration in $\mathcal{F}$,

(ii) for every $x \in E$, $\mathbf{P}_x$ is a probability measure on $\mathcal{F}$,

(iii) for every $B \in \mathcal{B}(E)$, the mapping $E \ni x \mapsto \mathbf{P}_x(B) \in [0, 1]$ is measurable,

(iv) for all $t \geq s \geq 0$, every $x \in E$, every $B \in \mathcal{B}(E)$, $\mathbf{P}_x(X(0) = x) = 1$ and

$$\mathbf{P}_x\left(X(t) \in B | \mathcal{F}_s\right) = \mathbf{P}_{X(s)}\left(X(t-s) \in B\right) \quad \mathbf{P}_x\text{-almost surely.}$$

Remark 2.1. A Markov family is often realized on a canonical space. Three standard choices, depending on regularity of the trajectories, are as follows:

a) $\Omega \doteq E^{[0,\infty)}$, $\mathcal{F} \doteq \otimes^{[0,\infty)}\mathcal{B}(E)$ for "wild" processes (no path regularity);

b) $\Omega \doteq \boldsymbol{D}([0, \infty), E)$ the space of càdlàg functions $[0, \infty) \to E$, $\mathcal{F} \doteq \mathcal{B}(\boldsymbol{D}([0, \infty), E))$, where $\boldsymbol{D}([0, \infty), E)$ is equipped with the Skorohod topology (càdlàg processes);

c) $\Omega \doteq \boldsymbol{C}([0, \infty), E)$ the space of continuous functions $[0, \infty) \to E$, $\mathcal{F} \doteq \mathcal{B}(\boldsymbol{C}([0, \infty), E))$, where $\boldsymbol{C}([0, \infty), E)$ is equipped with the topology of uniform convergence on compacts (continuous processes).

In all three cases, Ω is a space of functions on $[0,\infty)$, and the σ-algebra $\mathcal{F}$ is generated by the one-dimensional projections. The process $X(.)$ can be chosen as the canonical (or coordinate) process, that is, $X(t,\omega) \doteq \omega(t)$, $\omega \in \Omega$, $t \geq 0$. The filtration $(\mathcal{F}_t)$ becomes the canonical filtration (filtration generated by the coordinate process). In order to determine a Markov family it then remains to choose the family $(\mathbf{P}_x)$ of probability measures on $\mathcal{F}$.

Let $P(.,.,.)$ be the Markov kernel (transition probabilities) and $(S(t))_{t\geq 0}$ the Markov (or transition) semigroup associated with the Markov family X. Thus, for $t \geq 0$, $x \in E$, $B \in \mathcal{B}(E)$, $f \in \boldsymbol{B}(E)$,

$$P(t,x,B) = \mathbf{P}_x\left(X(t) \in B\right), \qquad S(t)(f)(x) = \mathbf{E}_x\left[f\big(X(t)\big)\right].$$

Suppose that X is such that $(S(t))_{t\geq 0}$ is strongly continuous as a semigroup on $\boldsymbol{C}_0(E)$. Let A be its infinitesimal generator, that is, A is the linear operator $\mathrm{dom}(A) \subset \boldsymbol{C}_0(E) \rightarrow \boldsymbol{C}_0(E)$ given by

$$\begin{aligned} \mathrm{dom}(A) &\doteq \left\{ f \in \boldsymbol{C}_0(E) : \lim_{h\to 0+} \frac{1}{h}\left(S(f)(h) - f\right) \text{ exists in } \boldsymbol{C}_0(E) \right\} \\ A(f) &\doteq \lim_{h\to 0+} \frac{1}{h}\left(S(f)(h) - f\right), \quad f \in \mathrm{dom}(A). \end{aligned}$$

For $f \in \mathrm{dom}(A)$ define a process M_f by

$$M_f(t) \doteq f(X(t)) - f(X(0)) - \int_0^t A(f)(X(s))ds, \quad t \geq 0.$$

Clearly, M_f is a real-valued $(\mathcal{F}_t)$-adapted process with $M_f(0) = 0$ and trajectories that are uniformly bounded on compact time intervals. Moreover, if X has continuous trajectories, then M_f has continuous trajectories.

The following observation is the key to the characterization of Markov processes in terms of martingale problems.

Proposition 2.1. *Let $f \in \mathrm{dom}(A)$, $x \in E$. Then M_f is an $(\mathcal{F}_t)$-martingale under $\mathbf{P}_x$.*

Proof. Since $(S(t))$ is strongly continuous, the mapping $[0,\infty)t \mapsto S(t)(f) \in \boldsymbol{C}_0(E)$ is differentiable whenever $f \in \mathrm{dom}(A)$, and

$$\frac{d}{dt}S(t)(f) = S(t)A(f) = A\,S(t)(f).$$

Let $t \geq s \geq 0$. It is enough to show that $\mathbf{E}\left[M_f(t) - M_f(s)|\mathcal{F}_s\right] = 0$, where expectation is taken with respect to $\mathbf{P}_x$. We have, using Fubini's theorem and the time derivative of $S(.)$,

$$
\begin{aligned}
&\mathbf{E}\left[M_f(t) - M_f(s)|\mathcal{F}_s\right] \\
&= \mathbf{E}\left[f(X(t))|\mathcal{F}_s\right] - f(X(s)) - \mathbf{E}\left[\int_s^t A(f)(X(r))dr|\mathcal{F}_s\right] \\
&= \int_E f(y)P(t-s, X(s), dy) - f(X(s)) - \int_s^t \int_E A(f)(y)P(r-s, X(s), dy)dr \\
&= S(t-s)(f)(X(s)) - f(X(s)) - \int_s^t S(r-s)A(f)(X(s))dr \\
&= S(t-s)(f)(X(s)) - f(X(s)) - \int_s^t \frac{d}{dr}S(r-s)(f)(X(s))dr \\
&= S(t-s)(f)(X(s)) - f(X(s)) - (S(t-s)(f)(X(s)) - S(0)(f)(X(s))) \\
&= 0.
\end{aligned}
$$

$\square$

3 Martingale problems

Let $A\colon \mathrm{dom}(A) \subset \boldsymbol{B}(E) \to \boldsymbol{B}(E)$ be a linear operator.

Definition 3.1. A triple $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t)_{t\geq 0}, (X(t))_{t\geq 0})$ with $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t))$ a stochastic basis and X an E-valued $(\mathcal{F}_t)$-adapted stochastic process is a *solution of the martingale problem* for A if, for all $f \in \mathrm{dom}(A)$,

$$
M_f(t) \doteq f(X(t)) - f(X(0)) - \int_0^t A(f)(X(s))ds, \quad t \geq 0,
$$

is a martingale with respect to $(\mathcal{F}_t)$.

If the filtration is not specified, it is understood to be the natural filtration of the process X. Let us write $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.)) \in \mathbf{MP}(A)$ if $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.))$ is a solution of the martingale problem for A. Let us write $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.)) \in \mathbf{MP}(A, \nu)$ if ν is a probability measure on $\mathcal{B}(E)$, $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.)) \in \mathbf{MP}(A)$ and $\mathbf{P} \circ (X(0))^{-1} = \nu$, that is, $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.))$ is a solution of the martingale problem for A with initial distribution ν. We will sometimes write $X \in \mathbf{MP}(A)$ or $X \in \mathbf{MP}(A, \nu)$ with the obvious interpretation.

The following criterion for a process of the form of the test processes M_f to be a martingale is sometimes useful.

Lemma 3.1. *Let $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t))$ be a stochastic basis and $(X(t))_{t\geq 0}$ be an E-valued $(\mathcal{F}_t)$-adapted stochastic process. Let $f, g \in \mathcal{B}(E)$, $\lambda > 0$. Then*

$$f(X(t)) - \int_0^t g(X(s))ds, \quad t \geq 0,$$

is a martingale with respect to $(\mathcal{F}_t)$ if and only if

$$e^{-\lambda t} f(X(t)) + \int_0^t e^{-\lambda s} \left(\lambda f(X(s)) - g(X(s))\right) ds, \quad t \geq 0.$$

is a martingale with respect to $(\mathcal{F}_t)$.

For a proof of Lemma 3.1, see §4.3.2 in Ethier and Kurtz [1986, pp. 174-175].

If A is the infinitesimal generator of a Feller Markov process, then it is dissipative (i.e., $\|\lambda f - A(f)\| \geq \lambda \|f\|$ for all $\lambda > 0$, $f \in \mathrm{dom}(A)$). A converse holds for operators A that admit solutions of the corresponding martingale problem.

Proposition 3.1. *Suppose $A\colon \mathrm{dom}(A) \subset \boldsymbol{B}(E) \to \boldsymbol{B}(E)$ is a linear operator such that $\mathbf{MP}(A, \delta_x)$ is non empty for every $x \in E$. Then A is dissipative.*

Proof. Fix $x \in E$ and let $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.)) \in \mathbf{MP}(A, \delta_x)$. Let $\lambda > 0$, $f \in \mathrm{dom}(A)$. By Lemma 3.1,

$$e^{-\lambda t} f(X(t)) - \int_0^t e^{-\lambda s} \left(\lambda f(X(s)) - A(f)(X(s))\right) ds, \quad t \geq 0,$$

is a martingale. Thus for all $t \geq 0$,

$$\mathbf{E}\left[e^{-\lambda t} f(X(t)) + \int_0^t e^{-\lambda s} \left(\lambda f(X(s)) - A(f)(X(s))\right) ds\right] = \mathbf{E}\left[f(X(0))\right].$$

Since $\mathbf{E}\left[f(X(0))\right] = f(x)$, letting t tend to infinity,

$$f(x) = \mathbf{E}\left[\int_0^\infty e^{-\lambda s} \left(\lambda f(X(s)) - A(f)(X(s))\right) ds\right]$$

Therefore,

$$|f(x)| \leq \int_0^\infty e^{-\lambda s} \|\lambda f - A(f)\| ds = \frac{1}{\lambda} \|\lambda f - A(f)\|,$$

hence $\lambda \|f\| \leq \|\lambda f - A(f)\|$. □

A martingale problem can only determine the law of its solution processes. The law of a stochastic process is determined by its finite-dimensional distributions.

Definition 3.2. *Uniqueness* is said to hold for the martingale problem for (A, ν) if, whenever $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.))$, $((\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{P}}), (\tilde{\mathcal{F}}_t), \tilde{X}(.))$ are solutions of the martingale problem for (A, ν), it follows that $\mathbf{P} \circ X^{-1} = \tilde{\mathbf{P}} \circ \tilde{X}^{-1}$. *Uniqueness* is said to hold for the martingale problem for A if, for every $\nu \in \mathcal{P}(E)$, uniqueness holds for the martingale problem for (A, ν).

Definition 3.3. The martingale problem for (A, ν) is said to be *well-posed* if $\mathbf{MP}(A, \nu)$ is non-empty and uniqueness holds for the martingale problem for (A, ν). The martingale problem for A is said to be *well-posed* if the martingale problem for (A, ν) is well-posed for every $\nu \in \mathcal{P}(E)$.

Remark 3.1. The solution processes are often required to have continuous or càdlàg paths. In this case, a solution of the martingale problem corresponds to a probability measure on the Borel sets of the canonical space $\boldsymbol{C}([0, \infty), E)$ or $\boldsymbol{D}([0, \infty), E)$, the solution process being the corresponding coordinate process. Martingale problems and the notions of uniqueness and well-posedness can thus be restricted to $\boldsymbol{C}([0, \infty), E)$ or $\boldsymbol{D}([0, \infty), E)$.

Remark 3.2. To ensure that the martingale problem for A is well-posed it is often enough to consider only initial distributions of the form δ_x for some $x \in E$. In particular, if E is compact, then well-posedness of the martingale problem for (A, δ_x) for all $x \in E$ implies that the martingale problem for A is well-posed.

Theorem 3.1. *Let* $A : \mathrm{dom}(A) \subset \boldsymbol{B}(E) \to \boldsymbol{B}(E)$ *be linear and dissipative, and let* $\nu \in \mathcal{P}(E)$. *Set* $L \doteq \mathrm{cl}(\mathrm{dom}(A))$, *that is,* L *is the closure of* $\mathrm{dom}(A)$ *in* $\boldsymbol{B}(E)$. *Suppose that* L *is measure-determining*[1] *and such that* $\mathrm{cl}(\mathrm{range}(\lambda_0 \,\mathrm{Id} - A)) = L$ *for some* $\lambda_0 > 0$. *If* $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.)) \in \mathbf{MP}(A, \nu)$, *then* $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.))$ *is a Markov process corresponding to the semigroup on* L *generated by* $\bar{A}$, *the closure of* A, *and uniqueness holds for the martingale problem for* (A, ν).

Theorem 3.1 is a version of Theorem 4.4.1 in Ethier and Kurtz [1986, pp. 182-184].

[1] A subset $L \subset \boldsymbol{B}(E)$ is called *measure-determining* if $\int f d\mu = \int f d\nu$ for all $f \in L$ implies $\mu = \nu$, where $\mu, \nu \in \mathcal{P}(E)$.

Remark 3.3. Suppose that A is the infinitesimal generator of a strongly continuous contraction semigroup on $\boldsymbol{C}_0(E)$. Then

(i) $\mathrm{dom}(A)$ is dense in $\boldsymbol{C}_0(E)$, hence $L = \boldsymbol{C}_0(E)$, and L is measure-determining;

(ii) the resolvent set $\rho(A)$ contains $\{\lambda \in \mathbb{C} : \Re(\lambda) > 0\}$, hence $\mathrm{range}(\lambda \,\mathrm{Id} - A) = L$ for every $\lambda > 0$;

(iii) A is a closed operator, hence $A = \bar{A}$.

In particular, if $A : \mathrm{dom}(A) \subset \boldsymbol{C}_0(E) \to \boldsymbol{C}_0(E)$ is a Markov generator associated with a Markov family $((\Omega, \mathcal{F}), (\mathcal{F}_t), (\mathbf{P}_x)_{x\in E}, X(.))$ that has the Feller property, then the assumptions of Theorem 3.1 are satisfied. Thus, for every $x \in E$, $((\Omega, \mathcal{F}, \mathbf{P}_x), (\mathcal{F}_t), X(.))$ is a solution of the martingale problem for (A, δ_x), and uniqueness holds. Theorem 3.1 is formulated in order to be applicable also to Markov pre-generators.

By Remark 3.3, uniqueness holds for the martingale problems associated with a Markov (pre-)generator. Conversely, uniqueness of solutions for the martingale problems associated with a linear operator A implies that the solutions have the Markov property.

Theorem 3.2. *Let* $A\colon \mathrm{dom}(A) \subset \boldsymbol{B}(E) \to \boldsymbol{B}(E)$ *be linear. Assume that uniqueness of solutions holds for* $\mathbf{MP}(A)$*. Then the following hold:*

a) If $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.)) \in \mathbf{MP}(A)$*, then* $X(.)$ *is a Markov process with respect to* $(\mathcal{F}_t)$*.*

b) Assume in addition that $\mathrm{dom}(A) \subset \boldsymbol{C}_b(E)$*. If* $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.)) \in \mathbf{MP}(A)$ *such that* $X(.)$ *has càdlàg trajectories, then* $X(.)$ *is a strong Markov process with respect to* $(\mathcal{F}_t)$*, that is, for all* $g \in \boldsymbol{B}(E)$*, all* $\mathbf{P}$*-almost surely finite* $(\mathcal{F}_t)$*-stopping times* τ*, all* $t \geq 0$*,*

$$\mathbf{E}\left[g(X(\tau + t))|\mathcal{F}_\tau\right] = \mathbf{E}\left[g(X(\tau + t))|X(\tau)\right].$$

Theorem 3.2 and Theorem 3.3 below are versions of Theorem 4.4.2 in Ethier and Kurtz [1986, pp. 184-186].

Proof of Theorem 3.2 (sketch). a) Let $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.)) \in \mathbf{MP}(A)$. Let $t > s \geq 0$, $g \in \boldsymbol{B}(E)$ be arbitrary. We have to show that $\mathbf{E}\left[g(X(t))|\mathcal{F}_s\right] = \mathbf{E}\left[g(X(t))|X(s)\right]$. It is enough to show that

$$\mathbf{E}\left[\mathbf{1}_{\Gamma_0}\, \mathbf{E}\left[g(X(t))|\mathcal{F}_s\right]\right] = \mathbf{E}\left[\mathbf{1}_{\Gamma_0}\, \mathbf{E}\left[g(X(t))|X(s)\right]\right] \tag{3.1}$$

for all $\Gamma_0 \in \mathcal{F}_s$ such that $\mathbf{P}(\Gamma_0) > 0$. Fix such an event $\Gamma_0 \in \mathcal{F}_s$. Define probability measures $\mathbf{P}_1$, $\mathbf{P}_2$ on $\mathcal{F}$ by

$$\mathbf{P}_1(\Gamma) \doteq \frac{1}{\mathbf{P}(\Gamma_0)} \mathbf{E}\left[\mathbf{1}_{\Gamma_0} \mathbf{E}\left[\mathbf{1}_{\Gamma}|\mathcal{F}_s\right]\right], \quad \mathbf{P}_2(\Gamma) \doteq \frac{1}{\mathbf{P}(\Gamma_0)} \mathbf{E}\left[\mathbf{1}_{\Gamma_0} \mathbf{E}\left[\mathbf{1}_{\Gamma}|X(s)\right]\right].$$

Set $Y(t) \doteq X(t+s)$ and $\tilde{\mathcal{F}}_t \doteq \mathcal{F}_{s+t}$, $t \geq 0$. Let ν be the conditional distribution of $X(s)$ given Γ_0, that is,

$$\nu(B) \doteq \mathbf{P}\left(X(s) \in B|\Gamma_0\right), \quad B \in \mathcal{B}(E).$$

We check that $((\Omega, \mathcal{F}, \mathbf{P}_1), (\tilde{\mathcal{F}}_t), Y(.))$, $((\Omega, \mathcal{F}, \mathbf{P}_2), (\tilde{\mathcal{F}}_t), Y(.))$ are both solutions of the martingale problem for (A, ν). As to the initial distribution, since $Y(0)$ is $\mathcal{F}_s$-measurable, we have for all $B \in \mathcal{B}(E)$,

$$\begin{aligned}
\mathbf{P}_1\left(Y(0) \in B\right) &= \frac{1}{\mathbf{P}(\Gamma_0)} \mathbf{E}\left[\mathbf{1}_{\Gamma_0} \mathbf{E}\left[\mathbf{1}_{\{Y(0)\in B\}}|\mathcal{F}_s\right]\right] \\
&= \frac{1}{\mathbf{P}(\Gamma_0)} \mathbf{E}\left[\mathbf{1}_{\Gamma_0 \cap \{Y(0)\in B\}}\right] \\
&= \mathbf{P}\left(X(s) \in B|\Gamma_0\right),
\end{aligned}$$

hence $\mathbf{P}_1 \circ (Y(0))^{-1} = \nu$. Similarly, since $Y(0)$ is also $\sigma(X(s))$-measurable, $\mathbf{P}_2 \circ (Y(0))^{-1} = \nu$. As to the martingale property, let $n \in \mathbb{N}$, $0 \leq t_0 < \ldots < t_n < t_{n+1}$, $h_0, \ldots, h_n \in \boldsymbol{B}(E)$, $f \in \mathrm{dom}(A)$ be arbitrary elements; set

$$\Psi(Y) \doteq \left(f(Y(t_{n+1})) - f(Y(t_n)) - \int_{t_n}^{t_{n+1}} A(f)(Y(r))dr\right) \prod_{i=0}^{n} h_i(Y(t_i)).$$

Since $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.)) \in \mathbf{MP}(A)$, we have

$$\mathbf{E}\left[\Psi(Y)|\mathcal{F}_s\right] = \mathbf{E}\left[\Psi(X(s+.))|\mathcal{F}_s\right] = 0.$$

This implies

$$\mathbf{E}_{\mathbf{P}_1}\left[\Psi(Y)\right] = 0 = \mathbf{E}_{\mathbf{P}_2}\left[\Psi(Y)\right],$$

which establishes the martingale property. By hypothesis, uniqueness holds for $\mathbf{MP}(A, \nu)$. This means, in particular, that any two solutions of $\mathbf{MP}(A, \nu)$ have the same one-dimensional distributions. Since $((\Omega, \mathcal{F}, \mathbf{P}_1), (\tilde{\mathcal{F}}_t), Y(.))$, $((\Omega, \mathcal{F}, \mathbf{P}_2), (\tilde{\mathcal{F}}_t), Y(.))$ are both solutions of $\mathbf{MP}(A, \nu)$, it follows that, for all $t \geq 0$, all $g \in \boldsymbol{B}(E)$,

$$\mathbf{E}_{\mathbf{P}_1}\left[g(Y(t))\right] = \mathbf{E}_{\mathbf{P}_1}\left[g(Y(t))\right],$$

which implies (3.1).

b) The argument is similar to a). Instead of $\mathbf{E}\left[\Psi(X(s+.))|\mathcal{F}_s\right] = 0$ one uses the fact that $\mathbf{E}\left[\Psi(X(\tau+.))|\mathcal{F}_\tau\right] = 0$ for any $\mathbf{P}$-almost surely finite stopping time τ; this is a consequence of the stopping (optional sampling) theorem and the boundedness of $\Psi(X(\tau+.))$. The hypothesis that $\mathrm{dom}(A) \subset \boldsymbol{C}_b(E)$ guarantees that $\Psi(X(\tau+.))$ has càdlàg trajectories whenever $X(.)$ has càdlàg trajectories. □

Uniqueness of solutions for a martingale problem (A,ν) means that any two solutions have the same finite-dimensional distributions. It is actually enough to check that all one-dimensional marginal distributions coincide. Notice that, in the proof of Theorem 3.2, only uniqueness of one-dimensional distributions was used.

Theorem 3.3. *Let $A\colon \mathrm{dom}(A) \subset \boldsymbol{B}(E) \to \boldsymbol{B}(E)$ be linear. Suppose that, whenever $\nu \in \mathcal{P}(E)$ and $((\Omega,\mathcal{F},\mathbf{P}),(\mathcal{F}_t),X(.))$, $((\bar{\Omega},\bar{\mathcal{F}},\bar{\mathbf{P}}),(\bar{\mathcal{F}}_t),\bar{X}(.))$ are in $\mathbf{MP}(A,\nu)$, it holds that*

$$\mathbf{P}\left(X(t) \in B\right) = \bar{\mathbf{P}}\left(\bar{X}(t) \in B\right) \quad \textit{for all } t \geq 0, \ \textit{all } B \in \mathcal{B}(E).$$

Then uniqueness holds for $\mathbf{MP}(A)$.

Proof (sketch). Let $((\Omega,\mathcal{F},\mathbf{P}),(\mathcal{F}_t),X(.))$, $((\tilde{\Omega},\tilde{\mathcal{F}},\tilde{\mathbf{P}}),(\tilde{\mathcal{F}}_t),\tilde{X}(.))$ be solutions of the martingale problem for (A,ν). In order to check that $\mathbf{P}\circ X^{-1} = \tilde{\mathbf{P}}\circ\tilde{X}^{-1}$, it is enough to show that for every $n \in \mathbb{N}$,

$$\mathbf{E}_{\mathbf{P}}\left[\prod_{i=1}^{n} g_i(X(t_i))\right] = \mathbf{E}_{\tilde{\mathbf{P}}}\left[\prod_{i=1}^{n} g_i(\tilde{X}(t_i))\right] \tag{3.2}$$

for all $0 \leq t_1 < \ldots < t_n$, all $g_1,\ldots,g_n \in \boldsymbol{B}(E)$. It suffices to consider strictly positive test functions $g_1,\ldots,g_n$. Show (3.2) by induction over n. By hypothesis, (3.2) holds if $n = 1$.

Induction step $n \to n+1$. Let $0 \leq t_1 < \ldots < t_n$, and let $g_1,\ldots,g_n \in \boldsymbol{B}(E)$ be strictly positive. Define probability measures Q, $\tilde{Q}$ on $\mathcal{F}$ and $\tilde{\mathcal{F}}$, respectively, by

$$\begin{aligned}
Q(\Gamma) &\doteq \frac{1}{\mathbf{E}_{\mathbf{P}}\left[\prod_{i=1}^{n} g_i(X(t_i))\right]}\,\mathbf{E}_{\mathbf{P}}\left[\mathbf{1}_{\Gamma}\cdot\prod_{i=1}^{n} g_i(X(t_i))\right], \qquad \Gamma \in \mathcal{F},\\
\tilde{Q}(\tilde{\Gamma}) &\doteq \frac{1}{\mathbf{E}_{\tilde{\mathbf{P}}}\left[\prod_{i=1}^{n} g_i(\tilde{X}(t_i))\right]}\,\mathbf{E}_{\tilde{\mathbf{P}}}\left[\mathbf{1}_{\tilde{\Gamma}}\cdot\prod_{i=1}^{n} g_i(\tilde{X}(t_i))\right], \qquad \tilde{\Gamma} \in \tilde{\mathcal{F}}.
\end{aligned}$$

Set $Y(t) \doteq X(t_n+t)$, $\tilde{Y}(t) \doteq \tilde{X}(t_n+t)$, $t \geq 0$. As in the proof of Theorem 3.2, one shows that $((\Omega, \mathcal{F}, Q), (\mathcal{F}_{t_n+t}), Y(.))$, $((\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{Q}), (\tilde{\mathcal{F}}_{t_n+t}), \tilde{X}(.))$ are both in $\mathbf{MP}(A)$. By induction hypothesis, it follows that

$$\mathbf{E}_Q\left[g(Y(0))\right] = \mathbf{E}\left[g(\tilde{Y}(0))\right]$$

for all $g \in \boldsymbol{B}(E)$ with $g > 0$. This implies $Q \circ (Y(0))^{-1} = \tilde{Q} \circ (\tilde{Y}(0))^{-1}$, that is, the two initial distributions coincide. Using the hypothesis on uniqueness of one-dimensional distributions, one sees that, for $g \in \boldsymbol{B}(E)$,

$$\mathbf{E}_Q\left[g(Y(t))\right] = \mathbf{E}\left[g(\tilde{Y}(t))\right] \quad \text{for all } t \geq 0.$$

Consequently, again using the induction hypothesis,

$$\mathbf{E}_{\mathbf{P}}\left[g(X(t)) \prod_{i=1}^{n} g_i(X(t_i))\right] = \mathbf{E}_{\tilde{\mathbf{P}}}\left[g(\tilde{X}(t)) \prod_{i=1}^{n} g_i(\tilde{X}(t_i))\right],$$

which establishes (3.2) for $n+1$. $\square$

Theorem 3.3 sometimes allows to link uniqueness for the martingale problems associated with one operator to existence of solutions of the martingale problems for a second operator.

Definition 3.4. Let $A\colon \mathrm{dom}(A) \subset \boldsymbol{B}(E) \to \boldsymbol{B}(E)$, $\tilde{A}\colon \mathrm{dom}(\tilde{A}) \subset \boldsymbol{B}(\tilde{E}) \to \boldsymbol{B}(\tilde{E})$ be linear operators, where $(\tilde{E}, \tilde{d})$ is a second locally compact separable metric space. Let $\nu \in \mathcal{P}(E)$, $\tilde{\nu} \in \mathcal{P}(\tilde{E})$, and let $H : E \times \tilde{E} \to \mathbb{C}$ be measurable. The martingale problems for (A, ν) and $(\tilde{A}, \tilde{\nu})$ are said to be *dual* with respect to the *duality function* H if, whenever $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.)) \in \mathbf{MP}(A, \nu)$, $((\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{P}}), (\tilde{\mathcal{F}}_t), Y(.)) \in \mathbf{MP}(\tilde{A}, \tilde{\nu})$, it holds that

$$\int_{\tilde{E}} \mathbf{E}_{\mathbf{P}}\left[H(X(t), y)\right] \tilde{\nu}(dy) = \int_{E} \mathbf{E}_{\tilde{\mathbf{P}}}\left[H(x, Y(t))\right] \nu(dx).$$

The duality relation introduced in Definition 3.4 can sometimes be used to prove that uniqueness holds for a martingale problem, which in general is more difficult to establish than existence of solutions.

Theorem 3.4. *Suppose that (E, d) is complete (in addition to being separable and locally compact). Let $A\colon \mathrm{dom}(A) \subset \boldsymbol{B}(E) \to \boldsymbol{B}(E)$, $\tilde{A}\colon \mathrm{dom}(\tilde{A}) \subset \boldsymbol{B}(\tilde{E}) \to \boldsymbol{B}(\tilde{E})$ be linear operators. Let $H\colon E \times \tilde{E} \to \mathbb{C}$ be measurable such that the set $\{H(., y) : y \in \tilde{E}\}$ is measure-determining for $\mathcal{P}(E)$. Suppose that, for every $y \in \tilde{E}$, every $\nu \in \mathcal{P}(E)$ with compact support, the martingale* *problems for (A, ν) and $(\tilde{A}, \delta_y)$ are dual with respect to H. If $\mathbf{MP}(\tilde{A}, \delta_y)$ is non-empty for all $y \in \tilde{E}$, then, given any $\nu \in \mathcal{P}(E)$, uniqueness holds for $\mathbf{MP}(A, \nu)$.*

Theorem 3.4 is a simplified version of Proposition 4.4.7 in Ethier and Kurtz [1986, pp. 189-190].

Proof of Theorem 3.4. For $y \in \tilde{E}$, let $(\ldots, Y_y(.))$ be a solution of the martingale problem for $(\tilde{A}, \delta_y)$; by hypothesis, $\mathbf{MP}(\tilde{A}, \delta_y)$ is non-empty. Let $\nu \in \mathcal{P}(E)$ with compact support. If $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.))$, $((\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbf{P}}), (\bar{\mathcal{F}}_t), \bar{X}(.))$ are two solutions of the martingale problem for (A, ν), then for all $t \geq 0$, all $y \in \tilde{E}$,

$$\mathbf{E}_{\mathbf{P}}\left[H(X(t), y)\right] = \int_E \mathbf{E}_y\left[H(x, Y_y(t))\right] \nu(dx) = \mathbf{E}_{\bar{\mathbf{P}}}\left[H(\bar{X}(t), y)\right].$$

Since $\{H(., y) : y \in \tilde{E}\}$ is measure-determining by hypothesis, it follows that the one-dimensional distributions of any two solutions of $\mathbf{MP}(A, \nu)$ coincide. By Theorem 3.3, this implies that uniqueness holds for $\mathbf{MP}(A, \nu)$.

Uniqueness for $\mathbf{MP}(A, \nu)$ when ν does not have compact support is established by an approximation (conditioning) argument using the fact that any probability measure on a complete and separable metric space is tight. □

4 Stochastic differential equations

Let $b\colon \mathbb{R}^d \to \mathbb{R}^d$, $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d \times d_1}$ be measurable functions. The stochastic differential equation of Itô type associated with b, σ can be written formally as

$$dX(t) = b(X(t))dt + \sigma(X(t))dW(t), \tag{4.1}$$

where W is a standard Wiener process. Definition 4.1 makes rigorous the idea of solutions of Equation (4.1).

Definition 4.1. A quadruple $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), W(.), X(.))$ is a *solution* of the stochastic differential equation associated with (b, σ) if

(i) $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space,

(ii) $(\mathcal{F}_t)$ is a filtration in $\mathcal{F}$ (satisfying the usual conditions),

(iii) W is a d_1-dimensional Wiener process with respect to $(\mathcal{F}_t)$,

(iv) X is an $(\mathcal{F}_t)$-adapted $\mathbb{R}^d$-valued continuous process such that the processes $b(X(.))$, $\sigma(X(.))$ are integrable and, $\mathbf{P}$-almost surely,

$$X(t) = X(0) + \int_0^t b(X(s))ds + \int_0^t \sigma(X(s))dW(s), \quad t \geq 0.$$

The integrability condition in Definition 4.1 means that for all $T > 0$, all $i \in \{1, \ldots, d\}$, all $j \in \{1, \ldots, d_1\}$,

$$\mathbf{P}\left(\int_0^T |b_i(X(s))|ds < \infty\right) = 1, \quad \mathbf{P}\left(\int_0^T |\sigma_{ij}(X(s))|^2 ds < \infty\right) = 1.$$

That condition ensures that the integrals (Lebesgue and Itô, respectively) appearing in the equation

$$X(t) = X(0) + \int_0^t b(X(s))ds + \int_0^t \sigma(X(s))dW(s) \tag{4.2}$$

are well-defined for all $t \geq 0$. The integral equation (4.2) gives a rigorous meaning to the differential equation (4.1).

Let us write $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), W(.), X(.)) \in \mathbf{SDE}(b, \sigma)$ to indicate that $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), W(.), X(.))$ is a *solution* of the stochastic differential equation associated with (b, σ). Given a distribution $\nu \in \mathcal{P}(\mathbb{R}^d)$, let us write $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), W(.), X(.)) \in \mathbf{SDE}(b, \sigma; \nu)$ to indicate that $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), W(.), X(.))$ is a *solution* of the stochastic differential equation associated with (b, σ) such that $\mathbf{P} \circ (X(0))^{-1} = \nu$; that is, the solution has *initial distribution* ν.

The notion of solution introduced in Definition 4.1 is that of a weak solution. A solution $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), W(.), X(.)) \in \mathbf{SDE}(b, \sigma)$ is called a *strong solution* of the stochastic differential equation associated with (b, σ) if $X(.)$ is adapted to the (augmentation of) the filtration generated by the Wiener process W and the initial condition $X(0)$, that is, the filtration

$$\mathcal{G}_t \doteq \sigma(X(0), W(s), \text{events of measure zero} : s \leq t), \quad t \geq 0.$$

Two notions of uniqueness are relevant for stochastic differential equations.

Definition 4.2. *Pathwise uniqueness* is said to hold for the stochastic differential equation associated with (b, σ) if, whenever $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), W(.), X(.))$, $((\Omega, \mathcal{F}, \mathbf{P}), (\tilde{\mathcal{F}}_t), W(.), \tilde{X}(.))$ are two solutions of $\mathbf{SDE}(b, \sigma)$ such that $X(0) = \tilde{X}(0)$ $\mathbf{P}$-almost surely, it follows that

$$\mathbf{P}\left(X(t) = \tilde{X}(t) \text{ for all } t \geq 0\right) = 1.$$

Definition 4.3. *Uniqueness in law* (or *weak uniqueness*) is said to hold for the stochastic differential equation associated with (b, σ) if, whenever $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), W(.), X(.))$, $((\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{P}}), (\tilde{\mathcal{F}}_t), \tilde{W}(.), \tilde{X}(.))$ are two solutions of $\mathbf{SDE}(b, \sigma)$ such that $\mathbf{P} \circ (X(0))^{-1} = \tilde{\mathbf{P}} \circ (\tilde{X}(0))^{-1}$, it follows that $\mathbf{P} \circ X^{-1} = \tilde{\mathbf{P}} \circ \tilde{X}^{-1}$.

For pathwise uniqueness, only solutions of the SDE associated with (b, σ) that differ in the solution process (and possibly the filtration) are compared. In particular, the driving Wiener process is the same. Pathwise uniqueness holds if almost sure equality of the initial condition implies almost sure equality of the solution trajectories. For uniqueness in law, on the other hand, any two solutions of the SDE associated with (b, σ) are compared. Uniqueness in law holds if having the same initial distribution implies that the solution processes have the same law or, equivalently, the same finite-dimensional distributions.

Remark 4.1. If the coefficients b, σ are globally Lipschitz continuous, then pathwise uniqueness holds for the stochastic differential equation associated with b, σ and, given any $\nu \in \mathcal{P}(\mathbb{R}^d)$, $\mathbf{SDE}(b, \sigma; \nu)$ contains a strong solution. This is the basic existence and uniqueness result for Itô stochastic differential equations; see, for instance, Section 5.2.B in Karatzas and Shreve [1991, pp. 286-291]. The global Lipschitz condition can be relaxed in that b, σ need only be locally Lipschitz continuous provided they satisfy a sublinear growth condition.

Remark 4.2. Tanaka's example: Consider the equation

$$X(t) = X(0) + \int_0^t \operatorname{sgn}(X(s)) dW(s), \quad t \geq 0, \tag{4.3}$$

where $\operatorname{sgn}(x) \doteq \mathbf{1}_{(0,\infty)}(x) - \mathbf{1}_{(-\infty,0]}(x)$, $x \in \mathbb{R}$, is the sign function and W is a standard one-dimensional Wiener process. Equation (4.3) corresponds to a stochastic differential equation with $d = d_1 = 1$ and coefficients $b \equiv 0$, $\sigma(x) = \operatorname{sgn}(x)$. Using Lèvy's characterization of Brownian motion [e.g. Karatzas and Shreve, 1991, p. 157], one sees that weak uniqueness holds for the stochastic differential equation associated with $(0, \operatorname{sgn})$. Fixing an initial distribution, for simplicity δ_0, it can be shown [e.g. Karatzas and Shreve, 1991, pp. 301-302] that Equation (4.3) admits a (weak) solution, but no strong solution; in particular, $\mathbf{SDE}(0, \operatorname{sgn}; \delta_0)$ is non-empty but contains no strong solution.

Pathwise uniqueness implies uniqueness in law. Pathwise uniqueness plus existence of (weak) solutions imply existence of strong solutions. In this situation, any solution of $\mathbf{SDE}(b,\sigma)$ is actually a strong solution, as Theorem 4.1 shows.

Theorem 4.1 (Yamada & Watanabe, Kallenberg)**.** *Suppose that pathwise uniqueness holds for the stochastic differential equation associated with* (b,σ) *and that* $\mathbf{SDE}(b,\sigma;\delta_x)$ *is non-empty given any* $x\in\mathbb{R}^d$*. Then there exists a Borel measurable map* $h\colon \mathbb{R}^d\times \boldsymbol{C}([0,\infty),\mathbb{R}^{d_1})\to\boldsymbol{C}([0,\infty),\mathbb{R}^d)$ *which is progressive such that, given any* $((\Omega,\mathcal{F},\mathbf{P}),(\mathcal{F}_t),W(.),X(.))\in\mathbf{SDE}(b,\sigma)$*, it holds for* $\mathbf{P}$*-almost all* $\omega\in\Omega$*,*

$$X(t,\omega)=h\left(X(0,\omega),W(.,\omega)\right)(t)\quad \textit{for all } t\geq 0.$$

Moreover, if $((\Omega,\mathcal{F},\mathbf{P}),(\mathcal{F}_t))$ *is a stochastic basis carrying a* d_1*-dimensional* $(\mathcal{F}_t)$*-Wiener process* $W(.)$ *and an* $\mathbb{R}^d$*-valued* $\mathcal{F}_0$*-measurable random variable* ξ*, then*

$$\tilde{X}(t,\omega)\doteq h\left(\xi(\omega),W(.,\omega)\right)(t),\quad (t,\omega)\in[0,\infty)\times\Omega,$$

defines a strong solution of $\mathbf{SDE}(b,\sigma)$ *with initial condition* $\tilde{X}(0)=\xi$ $\mathbf{P}$*-almost surely.*

For a proof of Theorem 4.1 see Kallenberg [1996], where the measurable dependence on the initial condition is established. Also see the original work by Yamada and Watanabe [1971] and Section 5.3.D in Karatzas and Shreve [1991, pp. 308-311].

Suppose that $\mathbf{SDE}(b,\sigma)$ is non-empty, and let $((\Omega,\mathcal{F},\mathbf{P}),(\mathcal{F}_t),W(.),X(.))\in\mathbf{SDE}(b,\sigma)$. Let $f\in\boldsymbol{C}^2(\mathbb{R}^d)$. By Itô's formula, for $t\geq 0$,

$$\begin{aligned}
&f(X(t))-f(X(0))\\
&=\sum_{i=1}^d\int_0^t\frac{\partial f}{\partial x_i}(X(s))dX_i(s)+\frac{1}{2}\sum_{i,j=1}^d\int_0^t\frac{\partial^2 f}{\partial x_i\partial x_j}(X(s))d\langle X_i(s),X_j(s)\rangle\\
&=\sum_{i=1}^d\int_0^t\frac{\partial f}{\partial x_i}(X(s))b_i(X(s))ds+\frac{1}{2}\sum_{i,j=1}^d\int_0^t\frac{\partial^2 f}{\partial x_i\partial x_j}(X(s))(\sigma\sigma^{\mathsf{T}})_{ij}(X(s))ds\\
&\quad+\sum_{i=1}^d\sum_{k=1}^{d_1}\int_0^t\frac{\partial f}{\partial x_i}(X(s))\sigma_{ik}(X(s))dW_k(s)\\
&=\int_0^t\mathcal{A}(f)(X(s))ds+\int_0^t\nabla f(X(s))\cdot\sigma(X(s))dW(s),
\end{aligned}$$

where $\mathcal{A} = \mathcal{A}_{(b,\sigma)}$ is given by

$$\mathcal{A}(f)(x) \doteq \underbrace{\sum_{i=1}^{d} b_i(x) \frac{\partial f}{\partial x_i}(x)}_{=\langle b(x), \nabla f(x)\rangle} + \frac{1}{2} \underbrace{\sum_{i,j=1}^{d} (\sigma\sigma^{\mathsf{T}})_{ij}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x)}_{=\mathrm{trace}(D^2 f(x)(\sigma\sigma^{\mathsf{T}})(x))}. \tag{4.4}$$

Observe that $\mathcal{A}$ is a second-order linear partial differential operator.

The stochastic integral process $\int_0^{\cdot} \nabla f(X(s)) \cdot \sigma(X(s)) dW(s)$ is a continuous local martingale with respect to $(\mathcal{F}_t)$. The process

$$M_f(t) \doteq f(X(t)) - f(X(0)) - \int_0^t \mathcal{A}(f)(X(s)) ds, \quad t \geq 0, \tag{4.5}$$

is therefore a continuous local martingale (starting in zero). If f is differentiable with compact support and b, σ are bounded on compact sets, then $\int_0^{\cdot} \nabla f(X(s)) \cdot \sigma(X(s)) dW(s)$ is a true martingale, hence M_f is a martingale. Consequently, $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.))$ is a solution of the local martingale problem for $(\mathcal{A}, \nu)$ with $\mathrm{dom}(\mathcal{A}) = \boldsymbol{C}^2(\mathbb{R}^d)$, where $\nu \doteq \mathbf{P} \circ (X(0))^{-1}$. If b, σ are bounded on compact sets, then $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.))$ is a solution of the martingale problem for $(\mathcal{A}, \nu)$ with $\mathrm{dom}(\mathcal{A}) = \boldsymbol{C}^2_c(\mathbb{R}^d)$. Actually, there is equivalence between weak solutions of a stochastic differential equation and solutions of the martingale problem for the associated operator provided that only continuous solution processes are considered.

Theorem 4.2. *Let $\mathcal{A}$ be the differential operator associated with b, σ, and let $\nu \in \mathcal{P}(\mathbb{R}^d)$.*

a) There exists a solution of $\mathbf{SDE}(b, \sigma; \nu)$ if and only if there exists a solution in $\boldsymbol{C}([0,\infty), \mathbb{R}^d)$ of the local martingale problem for $(\mathcal{A}, \nu)$ with $\mathrm{dom}(\mathcal{A}) = \boldsymbol{C}^2(\mathbb{R}^d)$.

b) If there is a solution in $\boldsymbol{C}([0,\infty), \mathbb{R}^d)$ of $\mathbf{MP}(\mathcal{A}, \nu)$ with $\mathrm{dom}(\mathcal{A}) = \boldsymbol{C}^2_c(\mathbb{R}^d)$, then there exists a solution of $\mathbf{SDE}(b, \sigma; \nu)$. If b, σ are bounded on compacts, then equivalence holds as in a).

c) Uniqueness holds for the (local) martingale problem for $\mathcal{A}$ if and only if uniqueness in law holds for $\mathbf{SDE}(b, \sigma)$.

Proof (sketch). By Itô's formula, a solution of $\mathbf{SDE}(b, \sigma; \nu)$ gives rise to a solution in $\boldsymbol{C}([0,\infty), \mathbb{R}^d)$ of the (local) martingale problem for $(\mathcal{A}, \nu)$. The difficulty in showing the converse implication lies in constructing a driving Wiener process for the stochastic differential equation.

In order to show "local $\mathbf{MP} \to \mathbf{SDE}$," suppose that $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.))$ is a solution of the local martingale problem for $(\mathcal{A}, \nu)$ with $\mathrm{dom}(\mathcal{A}) = \boldsymbol{C}^2(\mathbb{R}^d)$. Then M_f as given by (4.5) is a local martingale for any $f \in \boldsymbol{C}^2(\mathbb{R}^d)$, in particular, for functions f of the form $f(x) = x_i$ or $f(x) = x_i \cdot x_j$, $i, j \in \{1, \ldots, d\}$. Consequently,

$$Y(t) \doteq X(t) - X(0) - \int_0^t b(X(s))ds, \quad t \geq 0,$$

is a d-dimensional vector of continuous local martingales. If $d = d_1$ and $\sigma(x)$ is invertible for every $x \in \mathbb{R}^d$, then

$$W(t) \doteq \int_0^t \sigma^{-1}(X(s))dY(s), \quad t \geq 0,$$

is well-defined. Being the stochastic integral with respect to a continuous local martingale, W is a continuous local martingale. Using the martingale property of M_f for the test functions $f(x) = x_i \cdot x_j$, one computes the quadratic covariations of W and concludes, using Lèvy's characterization of Brownian motion, that W is a d_1-dimensional Wiener process. The proof in the general case relies on introducing an auxiliary independent Wiener process and the use of a martingale representation theorem [cf. Karatzas and Shreve, 1991, pp. 315-317].

Part c) is a consequence of a) and b), respectively, since the solution process is the same for the (local) martingale problem as for the stochastic differential equation. □

Remark 4.3. The domain of the differential operator $\mathcal{A}$ can be restricted from $\boldsymbol{C}^2(\mathbb{R}^d)$ to $\boldsymbol{C}^\infty(\mathbb{R}^d)$ and from $\boldsymbol{C}^2_c(\mathbb{R}^d)$ to $\boldsymbol{C}^\infty_c(\mathbb{R}^d)$, respectively.

Remark 4.4 (Markov property). By Theorem 3.2, uniqueness for the martingale problem implies the Markov property for its solutions. In view of Theorem 4.2, if b, σ are bounded on compacts and such that uniqueness in law holds for $\mathbf{SDE}(b, \sigma)$, then any solution of $\mathbf{SDE}(b, \sigma)$ possesses the strong Markov property.

Uniqueness for the martingale problem for the differential operator $\mathcal{A}$ is related to existence of a solutions for a Cauchy problem associated with $\mathcal{A}$; cf. Section 5.4.E in Karatzas and Shreve [1991, pp. 325-327]. Theorem 4.3 below should also be compared to the general duality result given in Theorem 3.4.

Theorem 4.3 (Stroock & Varadhan)**.** *Let $\mathcal{A}$ be the differential operator associated with b, σ. Suppose b, σ are such that, for every $f \in \boldsymbol{C}_c^\infty(\mathbb{R}^d)$, the Cauchy problem*

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) = \mathcal{A}(u)(t,x) & \text{if } (t,x) \in (0,\infty) \times \mathbb{R}^d, \\ u(0,x) = f(x) & \text{if } (t,x) \in \{0\} \times \mathbb{R}^d, \end{cases}$$

has a solution $u_f \in \boldsymbol{C}^{1,2}((0,\infty) \times \mathbb{R}^d) \cap \boldsymbol{C}([0,\infty) \times \mathbb{R}^d)$ that is bounded on any strip of the form $[0,T] \times \mathbb{R}^d$. Then, for every $x \in \mathbb{R}^d$, uniqueness holds in $\boldsymbol{C}([0,\infty), \mathbb{R}^d)$ for $\mathbf{MP}(\mathcal{A}, \delta_x)$ with $\mathrm{dom}(\mathcal{A}) = \boldsymbol{C}_c^\infty(\mathbb{R}^d)$.

Proof. Let $x \in \mathbb{R}^d$. Suppose $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t), X(.))$, $((\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{P}}), (\tilde{\mathcal{F}}_t), \tilde{X}(.))$ are two solutions of $\mathbf{MP}(\mathcal{A}, \delta_x)$ such that X, $\tilde{X}$ are continuous. Fix $T > 0$, let $f \in \boldsymbol{C}_c^\infty(\mathbb{R}^d)$, and set $g(t,x) \doteq u_f(T-t,x)$, $(t,x) \in [0,T] \times \mathbb{R}^d$. Then $g \in \boldsymbol{C}^{1,2}((0,T) \times \mathbb{R}^d) \cap \boldsymbol{C}_b([0,\infty) \times \mathbb{R}^d)$ with

$$\frac{\partial g}{\partial t} = -\mathcal{A}(g) \text{ in } (0,T) \times \mathbb{R}^d, \qquad g(T,.) = f(.).$$

By (the proof of) Theorem 4.2 (solution of **MP** yields solution of **SDE**), there exists a d_1-dimensional Wiener process on (an extension of) $(\Omega, \mathcal{F}, \mathbf{P})$ such that, $\mathbf{P}$-almost surely,

$$X(t) = X(0) + \int_0^t b(X(s))ds + \int_0^t \sigma(X(s))dW(s), \quad t \geq 0.$$

An analogous representation holds for $\tilde{X}(.)$. By Itô's formula, the processes $(g(t,X(t)))_{t \in [0,T]}$ and $(g(t,\tilde{X}(t)))_{t \in [0,T]}$ are local martingales under $\mathbf{P}$ and $\tilde{\mathbf{P}}$, respectively; g being bounded, they are actually martingales. In particular, since $X(0) = x$ $\mathbf{P}$-almost surely and $\tilde{X}(0) = x$ $\tilde{\mathbf{P}}$-almost surely,

$$\begin{aligned} \mathbf{E}_P\left[g(T,X(T))\right] &= \mathbf{E}_P\left[g(0,X(0))\right] = g(0,x), \\ \mathbf{E}_{\tilde{P}}\left[g(T,\tilde{X}(T))\right] &= \mathbf{E}_{\tilde{P}}\left[g(0,\tilde{X}(0))\right] = g(0,x). \end{aligned}$$

Since $g(T,.) = f(.)$, it follows that

$$\mathbf{E}_P\left[f(X(T))\right] = \mathbf{E}_{\tilde{P}}\left[f(\tilde{X}(T))\right].$$

This equality is valid for all $T \geq 0$, all $f \in \boldsymbol{C}_c^\infty(\mathbb{R}^d)$. Since $\boldsymbol{C}_c^\infty(\mathbb{R}^d)$ is measure-determining, it follows that the one-dimensional distributions of the two solution processes coincide. Thus Theorem 3.3 applies, showing that uniqueness holds for $\mathbf{MP}(A, \delta_x)$. $\square$

Remark 4.5. A sufficient condition for the existence of (smooth) solutions of the Cauchy problem as required in the assumptions of Theorem 4.3 is as follows. The coefficients b, σ are bounded and Hölder continuous, and the diffusion matrix $\sigma\sigma^{\mathsf{T}}$ is uniformly positive definite, that is,

$$\exists c > 0 : \forall x, \xi \in \mathbb{R}^d : \ \xi^{\mathsf{T}} \sigma\sigma^{\mathsf{T}}(x)\xi \geq c|\xi|^2.$$

A proof can be found, for instance, in Chapter 1 of Friedman [1964, pp. 1-32]; see in particular Theorems 10 and 12 there.

Existence of solutions of the martingale problem associated with the coefficients b, σ can be established using a discretization procedure and weak convergence of processes; cf. Theorem 6.1.6 in Stroock and Varadhan [1979, pp. 143-145] and Section 5.4.D in Karatzas and Shreve [1991, pp. 323-325].

Theorem 4.4 (Stroock & Varadhan)**.** *Let $\mathcal{A}$ be the differential operator associated with b, σ. Assume that b, σ are bounded and continuous. Let $\nu \in \mathcal{P}(\mathbb{R}^d)$. Then there exists a solution of $\mathbf{MP}(\mathcal{A}, \nu)$ with $\mathrm{dom}(\mathcal{A}) \subset \boldsymbol{C}_c^2(\mathbb{R}^d)$.*

Proof. Let $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_t))$ be a stochastic basis carrying a d_1-dimensional $(\mathcal{F}_t)$-Wiener process $W(.)$ and an $\mathbb{R}^d$-valued $\mathcal{F}_0$-measurable random variable ξ such that $\mathbf{P} \circ \xi^{-1} = \nu$. For $n \in \mathbb{N}$, define a process $X^{(n)}$ recursively by $X^{(n)}(0) \doteq \xi$ and, if $t \in (t_j^n, t_j^n]$ for some $j \in \mathbb{N}_0$ where $t_j^n \doteq j \cdot 2^{-n}$,

$$X^{(n)}(t) \doteq X^{(n)}(t_j^n) + b(X^{(n)}(t_j^n)) \left(t - t_j^n\right) + \sigma(X^{(n)}(t_j^n)) \left(W(t) - W(t_j^n)\right).$$

Then $X^{(n)}$ is $(\mathcal{F}_t)$-adapted, continuous, and it solves the integral equation

$$X^{(n)}(t) = \xi + \int_0^t b^{(n)}(s, X^{(n)}(.))ds + \int_0^t \sigma^{(n)}(s, X^{(n)}(.))dW(s),$$

where $b^{(n)} : [0,\infty) \times \boldsymbol{C}([0,\infty), \mathbb{R}^d) \to \mathbb{R}^d$, $\sigma^{(n)} : [0,\infty) \times \boldsymbol{C}([0,\infty), \mathbb{R}^d) \to \mathbb{R}^{d \times d_1}$ are progressive functionals given by

$$b^{(n)}(t, \phi) \doteq b(\phi(\lfloor t \rfloor_n)), \qquad \sigma^{(n)}(t, \phi) \doteq \sigma(\phi(\lfloor t \rfloor_n));$$

here $\lfloor . \rfloor_n$ is the projection on the dyadic mesh of order n, that is, $\lfloor t \rfloor_n \doteq t_{j(t)}^n$ with $j(t) \in \mathbb{N}_0$ such that $t \in [t_{j(t)}^n, t_{j(t)+1}^n)$.

Set $\theta_n \doteq \mathbf{P} \circ (X^{(n)})^{-1}$. Thus $\theta_n \in \mathcal{P}(\boldsymbol{C}([0,\infty), \mathbb{R}^d))$ is the law of $X^{(n)}$ under $\mathbf{P}$. Use Kolmogorov's condition (Theorem A.5)[2] to check that the family

[2] To establish relative compactness, one could directly use the characterization of Theorem A.4 together with an estimate on the moments of the modulus of continuity of the underlying Itô diffusions; see, for instance, Fischer & Nappo, "On the moments of the modulus of continuity of Itô processes," *Stoch. Anal. Appl.*, 28(1), 103–122, 2010.

$(\theta_n)_{n\in\mathbb{N}}$ is relatively compact in $\mathcal{P}(\boldsymbol{C}([0,\infty),\mathbb{R}^d))$. Consequently, $(\theta_n)_{n\in\mathbb{N}}$ has limit points with respect to the topology of weak convergence of probability measures. Let $\theta \in \mathcal{P}(\boldsymbol{C}([0,\infty),\mathbb{R}^d))$ be a limit point; we may assume that $\theta_n \stackrel{w}{\to} \theta$ as $n \to \infty$, otherwise choose a convergent subsequence and relabel the indices.

Define operators $\mathcal{A}^{(n)}\colon [0,\infty) \times \boldsymbol{C}_c^2(\mathbb{R}^d) \to \boldsymbol{C}_b(\boldsymbol{C}([0,\infty),\mathbb{R}^d))$ by

$$\mathcal{A}_t^{(n)}(f)(\phi) \doteq \sum_{i=1}^{d} b_i^{(n)}(t,\phi)\frac{\partial f}{\partial x_i}(\phi(t)) + \frac{1}{2}\sum_{i,j=1}^{d} (\sigma^{(n)}\sigma^{(n)\mathsf{T}})_{ij}(t,\phi)\frac{\partial^2 f}{\partial x_i \partial x_j}(\phi(t)).$$

For $f \in \boldsymbol{C}_c^2(\mathbb{R}^d)$, define processes $M_f(.)$, $M_f^{(n)}(.)$ on the canonical space $(\boldsymbol{C}([0,\infty),\mathbb{R}^d), \mathcal{B}(\boldsymbol{C}([0,\infty),\mathbb{R}^d))$ by

$$M_f(t,\phi) \doteq f(\phi(t)) - f(\phi(0)) - \int_0^t \mathcal{A}(f)(\phi(s))ds,$$

$$M_f^{(n)}(t,\phi) \doteq f(\phi(t)) - f(\phi(0)) - \int_0^t \mathcal{A}_s^{(n)}(f)(\phi)ds.$$

Denote by $\Phi\colon [0,\infty) \times \boldsymbol{C}([0,\infty),\mathbb{R}^d) \to \mathbb{R}^d$ the canonical process given by $\Phi(t,\phi) \doteq \phi(t)$, and let $(\mathcal{B}_t)$ denote the canonical filtration in $\mathcal{B}(\boldsymbol{C}([0,\infty),\mathbb{R}^d))$, that is, $\mathcal{B}_t \doteq \sigma(\Phi(s) : s \leq t)$, $t \geq 0$. By Itô's formula and the boundedness of the coefficients,

$$f(X^{(n)}(t)) - f(X^{(n)}(0)) - \int_0^t \mathcal{A}_s(f)(X^{(n)}(.))ds, \quad t \geq 0,$$

is a martingale with respect to $(\mathcal{F}_t)$ under $\mathbf{P}$. Since $\mathbf{P} \circ (X^{(n)})^{-1} = \theta_n$, this implies that $M_f^{(n)}$ is a martingale with respect to (the θ_n-augmentation of) $(\mathcal{B}_t)$ under θ_n.

We want to show that, for every $f \in \boldsymbol{C}_c^2(\mathbb{R}^d)$, M_f is a $(\mathcal{B}_t)$-martingale under θ. Let $t > s \geq 0$, and let $G \in \boldsymbol{C}_b(\boldsymbol{C}([0,\infty),\mathbb{R}^d))$ be $\mathcal{B}_s$-measurable; it is then enough to show that

$$\mathbf{E}_\theta\left[(M_f(t) - M_f(s)) \cdot G\right] = 0. \tag{4.6}$$

Since $M_f(r)$, $M_f^{(n)}(r)$ are in $\boldsymbol{C}_b(\boldsymbol{C}([0,\infty),\mathbb{R}^d))$ for all $r \geq 0$, the mapping theorem for weak convergence implies that

$$\mathbf{E}_{\theta_n}\left[(M_f(t) - M_f(s)) \cdot G\right] \stackrel{n\to\infty}{\longrightarrow} \mathbf{E}_\theta\left[(M_f(t) - M_f(s)) \cdot G\right].$$

It is not difficult to check that $M_f^{(n)}(r) \to M_f(r)$ uniformly on compacts in $\boldsymbol{C}([0,\infty),\mathbb{R}^d)$, given any $r \geq 0$. Since $(\theta_n)_{n\in\mathbb{N}}$ is tight, this implies that

$$\lim_{n\to\infty} \mathbf{E}_{\theta_n}\left[\left(M_f(t) - M_f(s) - M_f^{(n)}(t) + M_f^{(n)}(s)\right) \cdot G\right] = 0.$$

By the martingale property of $M_f^{(n)}$ under θ_n,

$$\mathbf{E}_{\theta_n}\left[\left(M_f^{(n)}(t) - M_f^{(n)}(s)\right)\cdot G\right] = 0,$$

hence

$$\begin{aligned}&\mathbf{E}_{\theta_n}\left[(M_f(t) - M_f(s))\cdot G\right]\\ &= \mathbf{E}_{\theta_n}\left[\left(M_f(t) - M_f(s) - M_f^{(n)}(t) + M_f^{(n)}(s)\right)\cdot G\right].\end{aligned}$$

Consequently,

$$\mathbf{E}_{\theta}\left[(M_f(t) - M_f(s))\cdot G\right] = \lim_{n\to\infty}\mathbf{E}_{\theta_n}\left[(M_f(t) - M_f(s))\cdot G\right] = 0,$$

which establishes (4.6). □

5 Two examples

5.1 Uniqueness in law for the Wright-Fisher diffusion

This example is taken from Klenke [2008, pp. 584-586]. Consider the scalar equation (in differential form)

$$dX(t) = \mathbf{1}_{[0,1]}(X(t))\cdot\sqrt{\gamma(1-X(t))X(t)}dW(t), \tag{5.1}$$

where W is a one-dimensional Wiener process and $\gamma > 0$ a parameter. The coefficients of the stochastic differential equation (5.1) are thus $b \equiv 0$ and $\sigma(x) \doteq \mathbf{1}_{[0,1]}(x)\cdot\sqrt{\gamma(1-x)x}$, $x\in\mathbb{R}$. Notice that σ is bounded and continuous. The associated differential operator $\mathcal{A}$ is given by

$$\mathcal{A}(f)(x) \doteq \frac{\gamma}{2}\mathbf{1}_{[0,1]}(x)\cdot(1-x)xf''(x),\quad x\in\mathbb{R},$$

with $\mathrm{dom}(\mathcal{A}) = \boldsymbol{C}_c^2(\mathbb{R})$. By Theorem 4.4 and part b) of Theorem 4.2, Equation (5.1) possesses a (weak) solution for any given initial distribution. It is intuitively clear and not hard to show that, if $((\Omega,\mathcal{F},\mathbf{P}),(\mathcal{F}_t),W(.),X(.))$ is a solution of (5.1) such that $X(0)\in[0,1]$ $\mathbf{P}$-almost surely, then $X(t)\in[0,1]$ for all $t\geq 0$ with probability one.

We want to show that uniqueness in law holds for (5.1) given any deterministic initial condition in $[0,1]$. In view of Theorem 4.2, it is enough to show that uniqueness holds for $\mathbf{MP}(\mathcal{A},\delta_x)$ for every $x\in[0,1]$. We may and will restrict the domain of $\mathcal{A}$ to $\mathrm{dom}(\mathcal{A}) = \boldsymbol{C}_c^2([0,1])$.

Let $X(.)$ be the canonical process on $\boldsymbol{C} \doteq \boldsymbol{C}([0,\infty),[0,1])$, let $(\mathcal{B}_t)$ be the canonical filtration. For $x \in [0,1]$, let $\mathbf{P}_x$ be a probability measure on $\mathcal{B}(\boldsymbol{C})$ such that $((\boldsymbol{C},\mathcal{B}(\boldsymbol{C}),\mathbf{P}_x),(\mathcal{B}_t),X(.)) \in \mathbf{MP}(\mathcal{A},\delta_x)$. Let Y be the canonical process on $\boldsymbol{D} \doteq \boldsymbol{D}([0,\infty),\mathbb{N})$. For $n \in \mathbb{N}$, let $\tilde{\mathbf{P}}_n$ be a probability measure on $\mathcal{B}(\boldsymbol{D})$ such that Y is an $\mathbb{N}$-valued Markov process with $Y(0) = n$ $\tilde{\mathbf{P}}_n$-almost surely and rate matrix

$$q(k,l) = \begin{cases} \gamma\binom{k}{2} & \text{if } l = k-1, \\ -\gamma\binom{k}{2} & \text{if } l = k, \\ 0 & \text{otherwise.} \end{cases} \tag{5.2}$$

Thus, under $\tilde{\mathbf{P}}_n$, Y starts in state n and jumps downward in steps of one (with rates given by $q(k,k-1)$) until it reaches state one. Define a duality function $H\colon [0,1]\times\mathbb{N} \to \mathbb{R}$ by $H(x,n) \doteq x^n$. Notice that $(H(.,n))_{n\in\mathbb{N}}$ is measure-determining for $\mathcal{P}([0,1])$. For $x \in [0,1]$, $n \in \mathbb{N}$, $t \geq 0$, set

$$\begin{aligned} m_{x,n}(t) &\doteq \mathbf{E}_{\mathbf{P}_x}\left[H(X(t),n)\right] = \mathbf{E}_{\mathbf{P}_x}\left[(X(t))^n\right], \\ g_{x,n}(t) &\doteq \mathbf{E}_{\tilde{\mathbf{P}}_n}\left[H(x,Y(t))\right] = \mathbf{E}_{\tilde{\mathbf{P}}_n}\left[x^{Y(t)}\right], \end{aligned}$$

and show that $m_{x,n}(.) = g_{x,n}(.)$. As in the proof of Theorem 3.4 (considering only Dirac distributions), this implies that uniqueness holds for $\mathbf{MP}(\mathcal{A},\delta_x)$, every $x \in [0,1]$.

Fix $x \in [0,1]$. Notice that X is a martingale under $\mathbf{P}_x$, hence $m_{x,1}(t) = \mathbf{E}_{\mathbf{P}_x}[X(0)] = x$. For $n \in \mathbb{N}\setminus\{1\}$, by Itô's formula,

$$\begin{aligned} (X(t))^n = (X(0))^n &+ \int_0^t n(X(s))^{n-1}\sqrt{\gamma(1-X(s))X(s)}dW(s) \\ &+ \frac{\gamma}{2}\int_0^t n(n-1)(X(s))^{n-2}(1-X(s))X(s)ds. \end{aligned}$$

Therefore, taking expectations with respect to $\mathbf{P}_x$ on both sides,

$$m_{x,n}(t) = x^n + 0 + \gamma\binom{n}{2}\int_0^t \left(m_{x,n-1}(s) - m_{x,n}(s)\right)ds.$$

It follows that the functions $m_{x,n}(.)$, $n \in \mathbb{N}$, solve the system of linear ordinary differential equations

$$\frac{d}{dt}m_{x,n}(t) = \begin{cases} 0 & \text{if } n = 1, \\ \gamma\binom{n}{2}\left(m_{x,n-1}(t) - m_{x,n}(t)\right) & \text{if } n \geq 2, \end{cases} \tag{5.3}$$

with initial condition $m_{x,n}(0) = x^n$, $n \in \mathbb{N}$. Clearly, system (5.3) together with the initial condition uniquely determines the functions $m_{x,n}(.)$. Notice that $g_{x,n}(0) = x^n$ for all $n \in \mathbb{N}$ since $Y(0) = n$ $\tilde{P}_n$-almost surely. It is therefore enough to show that the functions $g_{x,n}(.)$ solve system (5.3). For $n = 1$, since $Y(t) = 1$ $\tilde{P}_n$-almost surely for all $t \geq 0$, $g_{x,1}(t) = x$, hence $\frac{d}{dt} g_{x,n}(t) = 0$ for all $t \geq 0$. For $n \in \mathbb{N} \setminus \{1\}$, $t \geq 0$, $h \geq 0$, using the Markov property and the jump structure of Y,

$$\begin{aligned}
g_{x,n}(t+h) &= \mathbf{E}_{\tilde{\mathbf{P}}_n}\left[x^{Y(t+h)}\right] \\
&= \mathbf{E}_{\tilde{\mathbf{P}}_n}\left[\mathbf{E}_{\tilde{\mathbf{P}}_{Y(h)}}\left[x^{Y(t)}\right]\right] \\
&= \sum_{m=1}^{n} \tilde{\mathbf{P}}_n\left(Y(h) = m\right) \underbrace{\mathbf{E}_{\tilde{\mathbf{P}}_m}\left[x^{Y(t)}\right]}_{= g_{x,m}(t)},
\end{aligned}$$

hence, recalling (5.2),

$$\begin{aligned}
\frac{d}{dt} g_{x,n}(t) &= \lim_{h \to 0+} \frac{1}{h} \left(\sum_{m=1}^{n} \tilde{\mathbf{P}}_n\left(Y(h) = m\right) \left(g_{x,m}(t) - g_{x,n}(t)\right) \right) \\
&= \sum_{m=1}^{n} q(n,m) \left(g_{x,m}(t) - g_{x,n}(t)\right) \\
&= \gamma \binom{n}{2} \left(g_{x,n-1}(t) - g_{x,n}(t)\right).
\end{aligned}$$

It follows that the functions $g_{x,n}(.)$ satisfy (5.3) with the same initial condition as the functions $m_{x,n}(.)$.

5.2 Convergence to the McKean-Vlasov limit

Here we sketch a way of establishing a law of large numbers for the empirical measures of weakly interacting Itô diffusions; an early reference is Oelschläger [1984].

Let W_i, $i \in \mathbb{N}$, be independent d_1-dimensional Wiener processes. Let b, σ be functions defined on $\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$ taking values in $\mathbb{R}^d$ and $\mathbb{R}^{d \times d_1}$, respectively. Assume that b, σ are bounded and continuous, where $\mathcal{P}(\mathbb{R}^d)$ is equipped with the topology of weak convergence. For $N \in \mathbb{N}$, consider the system of stochastic differential equations

$$dX_i^N(t) = b\left(X_i^N(t), \mu^N(t)\right) dt + \sigma\left(X_i^N(t), \mu^N(t)\right) dW_i(t), \tag{5.4}$$

where $\mu^N(t)$ is the empirical measure of $X_1^N(t), \dots, X_N^N(t)$, that is,

$$\mu^N(t) \doteq \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i^N(t)}. \tag{5.5}$$

Notice that the mapping $(\mathbb{R}^d)^N \ni x \mapsto \frac{1}{N}\sum_{i=1}^N \delta_{x_i} \in \mathcal{P}(\mathbb{R}^d)$ is continuous. Define functions on $(\mathbb{R}^d)^N$ with values in $(\mathbb{R}^d)^N$ and $(\mathbb{R}^{d\times d_1})^N$, respectively, by

$$\begin{aligned} b_N(x) &\doteq \left(b\left(x_1, \frac{1}{N}\sum_{i=1}^N \delta_{x_i}\right), \dots, b\left(x_N, \frac{1}{N}\sum_{i=1}^N \delta_{x_i}\right)\right) \\ \sigma_N(x) &\doteq \left(\sigma\left(x_1, \frac{1}{N}\sum_{i=1}^N \delta_{x_i}\right), \dots, \sigma\left(x_N, \frac{1}{N}\sum_{i=1}^N \delta_{x_i}\right)\right). \end{aligned}$$

Since b, σ are assumed to be bounded and continuous, b_N, σ_N are bounded and continuous. Theorem 4.4 therefore guarantees that, given any initial distribution $\nu \in \mathcal{P}(\mathbb{R}^{N\cdot d})$, $\mathbf{SDE}(b_N, \sigma_N; \nu)$ is non-empty. Consequently, given any initial distribution, there is a (weak) solution of system (5.4) with that initial distribution. Let $\nu_0 \in \mathcal{P}(\mathbb{R}^d)$, and consider initial distributions $\nu = \nu_N$ of product form, that is, $\nu_N \doteq \otimes^N \nu_0$. Let $X^N(.)$ be the solution process of a (weak) solution with initial condition ν_N. For simplicity, we do not distinguish notationally between the possibly different stochastic bases and sequences of Wiener processes. For $t \geq 0$, let $\mu^N(t)$ be the empirical measure of $X^N(t) = (X_1^N(t), \dots, X_N^N(t))$ as given by (5.5). Notice that $\mu^N(t)$ is a $\mathcal{P}(\mathbb{R}^d)$-valued random variable.

One may ask what happens with $\mu^N(t)$ as N tends to infinity. Another reasonable question is, of course, to fix N and let t tend to infinity, or to let both parameters tend to infinity. Under certain conditions, the answer to the first question is that $\mu^N(t)$ tends to a non-random limit $\mu(t) \in \mathcal{P}(\mathbb{R}^d)$ with $\mu(t) = \mathrm{Law}(X(t))$, where $X(.)$ solves the "nonlinear" stochastic differential equation

$$dX(t) = b\left(X(t), \mathrm{Law}(X(t))\right) dt + \sigma\left(X(t), \mathrm{Law}(X(t))\right) dW(t) \tag{5.6}$$

with initial distribution $\mathrm{Law}(X(0)) = \nu_0$. Here it is assumed that (weak) existence and uniqueness in law hold for Equation (5.6). Thus, $\mu(.)$ satisfies the Kolmogorov forward equation associated with (5.6), that is,

$$\frac{d}{dt}\mu(t) = \mathcal{L}^*_{\mu(t)}\mu(t), \tag{5.7}$$

where $\mathcal{L}_\mu^*$ is the formal adjoint of the operator $\mathcal{L}_\mu$ given by

$$\mathcal{L}_\mu(f)(x) \doteq \sum_{i=1}^{d} b_i(x,\mu)\frac{\partial f}{\partial x_i}(x) + \frac{1}{2}\sum_{i,j=1}^{d}(\sigma\sigma^\mathsf{T})_{ij}(x,\mu)\frac{\partial^2 f}{\partial x_i \partial x_j}(x), \quad x \in \mathbb{R}^d,$$

with $\mathrm{dom}(\mathcal{L}) \subset \boldsymbol{C}^2(\mathbb{R}^d)$. Equation (5.7) is also called the McKean-Vlasov equation associated with the system of weakly interacting processes given by (5.4).

One way of establishing convergence of $\mu^N(t)$ to $\mu(t)$ as $N \to \infty$ is as follows. Fix a finite time horizon $T > 0$ and set

$$\hat{\mu}_T^N \doteq \frac{1}{N}\sum_{i=1}^{N} \delta_{(X_i^N(\cdot_{|[0,T]}), W_i(\cdot_{|[0,T]}))}.$$

Thus $\hat{\mu}_T^N$ is a random variable taking values in $\mathcal{P}(\boldsymbol{Z})$ with $\boldsymbol{Z} \doteq \boldsymbol{C}([0,T],\mathbb{R}^d)\times \boldsymbol{C}([0,T],\mathbb{R}^{d_1})$. Let Q_N denote its law, that is, $Q_N \doteq \mathbf{P}_N \circ (\mu_T^N)^{-1}$. The strategy is now similar to that of the proof of Theorem 4.4. First check that $(Q_N)_{N\in\mathbb{N}}$ is tight in $\mathcal{P}(\mathcal{P}(\boldsymbol{Z}))$. Let us suppose that this is the case. Let $(Q_{N(i)})$ be a convergent subsequence with limit Q. Let $\hat{\mu}$ be a random variable with law Q, that is, $\hat{\mu}$ is a $\boldsymbol{Z}$-valued random variable defined on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$ such that $\mathbf{P} \circ \hat{\mu} = Q$. The second step is now to show that, $\mathbf{P}$-almost surely, $\hat{\mu}$ is the law of a solution of (5.6). To this end, let (X, W) be the canonical process on $\boldsymbol{Z}$, let $(\mathcal{B}_t)$ be the canonical filtration. Then it is enough to show that for all $T \geq t > s \geq 0$, all $G \in \boldsymbol{C}_b(\boldsymbol{Z})$ $\mathcal{B}_s$-measurable, all $f \in \boldsymbol{C}_c^2(\mathbb{R}^d \times \mathbb{R}^{d_1})$, $\mathbf{P}$-almost surely,

$$\mathbf{E}_{\hat{\mu}}\left[f(X(t),W(t)) - f(X(s),W(s)) - \int_s^t \mathcal{A}_{\hat{\mu}}(f)(X(r),W(r))dr\right] = 0,$$

where[3]

$$\begin{aligned}\mathcal{A}_\mu(f)(x,z) \doteq{}& \sum_{i=1}^{d} b_i(x,\mu)\frac{\partial f}{\partial x_i}(x,z) + \frac{1}{2}\sum_{i,j=1}^{d}(\sigma\sigma^\mathsf{T})_{ij}(x,\mu)\frac{\partial^2 f}{\partial x_i\partial x_j}(x,z)\\ &+ \frac{1}{2}\sum_{k=1}^{d_1}\frac{\partial^2 f}{\partial z_k^2}(x,z) + \sum_{i=1}^{d}\sum_{k=1}^{d_1}\sigma_{ik}(x,\mu)\frac{\partial^2 f}{\partial x_i \partial z_k}(x,z).\end{aligned}$$

If uniqueness holds for the martingale problem for $\mathcal{A}_\mu$, $\mu \in \mathcal{P}(\mathbb{R}^d)$, then any limit point Q is equal to the Dirac measure concentrated at the (unique) law of the solution of Equation (5.6) with initial distribution ν_0.

[3] Apply the Itô formula to $f(X(t),W(t))$ assuming that $(\ldots,W,X)$ is a solution of Equation (5.6).

A Weak convergence of probability measures

Here we collect some facts about weak convergence of probability measures; cf. Chapter 13 and Section 21.7 in Klenke [2008] or Chapter 3 in Ethier and Kurtz [1986]. A standard reference on the topic is Billingsley [1999].

Let $\mathcal{X}$ be a *Polish space*, that is, $\mathcal{X}$ is a separable topological space that is metrizable by a complete metric. Any closed subset of Euclidean space under the standard topology is a Polish space. If $\mathcal{Y}$ is a Polish space, then the spaces of $\mathcal{Y}$-valued continuous trajectories $\boldsymbol{C}([0,T],\mathcal{Y})$, $\boldsymbol{C}([0,\infty),\mathcal{Y})$ are Polish under the topology of uniform convergence on compact time intervals, and the spaces of $\mathcal{Y}$-valued càdlàg trajectories $\boldsymbol{D}([0,T],\mathcal{Y})$, $\boldsymbol{D}([0,\infty),\mathcal{Y})$ are Polish under the Skorohod topology; see, for instance, Chapters 2 and 3 in Billingsley [1999]. Let $\mathcal{B}(\mathcal{X})$ be the Borel σ-algebra over $\mathcal{X}$. Denote by $\mathcal{P}(\mathcal{X})$ the space of probability measures on $\mathcal{B}(\mathcal{X})$.

Definition A.1. A sequence $(\theta_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathcal{X})$ is said to *converge weakly* to $\theta \in \mathcal{P}(\mathcal{X})$, in symbols $\theta_n \xrightarrow{w} \theta$, if

$$\int_{\mathcal{X}} f(x)\theta_n(dx) \overset{n\to\infty}{\longrightarrow} \int_{\mathcal{X}} f(x)\theta_n(dx) \text{ for all } f \in \boldsymbol{C}_b(\mathcal{X}).$$

The terminology weak convergence, rather than weak-$*$ convergence as in functional analysis, has historic roots. The limit of a weakly convergent sequence in $\mathcal{P}(\mathcal{X})$ is unique. Weak convergence induces a topology on $\mathcal{P}(\mathcal{X})$; under this topology, $\mathcal{X}$ being a Polish space, $\mathcal{P}(\mathcal{X})$ is a Polish space, too.

Let d be a complete metric compatible with the topology of $\mathcal{X}$; thus $(\mathcal{X}, d)$ is a complete and separable metric space. There are different choices for a complete metric on $\mathcal{P}(\mathcal{X})$ that is compatible with the topology of weak convergence. Two common choices are the Prohorov metric and the bounded Lipschitz metric, respectively. The *Prohorov metric* on $\mathcal{P}(\mathcal{X})$ is defined by

$$\rho(\theta,\nu) \doteq \inf\left\{\varepsilon > 0 : \theta(G) \leq \nu(G^{\varepsilon}) + \varepsilon \text{ for all closed } G \subset \mathcal{X}\right\}, \tag{A.1}$$

where $G^{\varepsilon} \doteq \{x \in \mathcal{X} : d(x,G) < \varepsilon\}$. Notice that ρ is indeed a metric. The *bounded Lipschitz metric* on $\mathcal{P}(\mathcal{X})$ is defined by

$$\tilde{\rho}(\theta,\nu) \doteq \sup\left\{ \left|\int f d\theta - \int f d\nu\right| : f \in \boldsymbol{C}_b(\mathcal{X}) \text{ such that } \|f\|_{bL} \leq 1 \right\}, \tag{A.2}$$

where $\|f\|_{bL} \doteq \sup_{x\in\mathcal{X}} |f(x)| + \sup_{x,y\in\mathcal{X}:x\neq y} \frac{|f(x)-f(y)|}{d(x,y)}$.

The following theorem gives a number of equivalent characterizations of weak convergence.

Theorem A.1 ("Portemanteau theorem"). *Let $(\theta_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathcal{X})$ be a sequence of probability measures, and let $\theta \in \mathcal{P}(\mathcal{X})$. Then the following are equivalent:*

(i) $\theta_n \xrightarrow{w} \theta$ *as* $n \to \infty$*;*

(ii) $\rho(\theta_n, \theta) \overset{n\to\infty}{\longrightarrow} 0$ *(Prohorov metric);*

(iii) $\tilde{\rho}(\theta_n, \theta) \overset{n\to\infty}{\longrightarrow} 0$ *(bounded Lipschitz metric);*

(iv) $\int f d\theta_n \overset{n\to\infty}{\longrightarrow} \int f d\theta$ *for all uniformly continuous* $f \in \boldsymbol{C}_b(\mathcal{X})$*;*

(v) $\int f d\theta_n \overset{n\to\infty}{\longrightarrow} \int f d\theta$ *for all Lipschitz continuous* $f \in \boldsymbol{C}_b(\mathcal{X})$*;*

(vi) $\int f d\theta_n \overset{n\to\infty}{\longrightarrow} \int f d\theta$ *for all* $f \in \boldsymbol{B}(\mathcal{X})$ *such that* $\theta(U_f) = 0$ *where* $U_f \doteq \{x \in \mathcal{X} : f$ *discontinuous at* $x\}$*;*

(vii) $\limsup_{n\to\infty} \theta_n(G) \leq \theta(G)$ *for all closed* $G \subset \mathcal{X}$*;*

(viii) $\liminf_{n\to\infty} \theta_n(O) \geq \theta(O)$ *for all open* $O \subset \mathcal{X}$*;*

(ix) $\lim_{n\to\infty} \theta_n(B) = \theta(B)$ *for all* $B \in \mathcal{B}(\mathcal{X})$ *such that* $\theta(\partial B) = 0$*, where* $\partial B \doteq \mathrm{cl}(B) \cap \mathrm{cl}(B^c)$ *denotes the boundary of the Borel set* B*.*

From Definition A.1 it is clear that weak convergence is preserved under continuous mappings. The mapping theorem for weak convergence requires continuity only with probability one with respect to the limit measure; this should be compared to characterizations (vi) and (ix) in Theorem A.1.

Theorem A.2 (Mapping theorem). *Let $(\theta_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathcal{X})$, $\theta \in \mathcal{P}(\mathcal{X})$. Let $\mathcal{Y}$ be a second Polish space, and let $\psi\colon \mathcal{X} \to \mathcal{Y}$ be a measurable mapping. If $\theta_n \xrightarrow{w} \theta$ and $\theta(\{x \in \mathcal{X} : \psi$ discontinuous at $x\}) = 0$, then $\psi \circ \theta_n \xrightarrow{w} \psi \circ \theta$.*

A standard method for proving that a sequence $(a_n)_{n\in\mathbb{N}}$ of elements of a complete metric space converges to a unique limit element a is to proceed as follows. First show that $(a_n)_{n\in\mathbb{N}}$ is relatively compact (i.e., its closure is compact). Then, taking any convergent subsequence $(a_{n(j)})_{j\in\mathbb{N}}$ with limit $\tilde{a}$, show that $\tilde{a} = a$. This establishes $a_n \to a$ as $n \in \mathbb{N}$. Relative compactness in $\mathcal{P}(\mathcal{X})$ with the topology of weak convergence when $\mathcal{X}$ is Polish can be characterized through uniform exhaustibility by compacts.

Definition A.2. Let I be a non-empty set. A family $(\theta_i)_{i\in I} \subset \mathcal{P}(\mathcal{X})$ is called *tight* (or *uniformly tight*) if for any $\varepsilon > 0$ there is a compact set $K_\varepsilon \subset \mathcal{X}$ such that

$$\inf_{i\in I} \theta_i(K_\varepsilon) \geq 1 - \varepsilon.$$

Theorem A.3 (Prohorov)**.** *Let I be a non-empty set, and let $(\theta_i)_{i\in I} \subset \mathcal{P}(\mathcal{X})$, where $\mathcal{X}$ is Polish. Then $(\theta_i)_{i\in I} \subset \mathcal{P}(\mathcal{X})$ is tight if and only if $(\theta_i)_{i\in I}$ is relatively compact in $\mathcal{P}(\mathcal{X})$ with respect to the topology of weak convergence.*

Depending on the structure of the underlying space $\mathcal{X}$, conditions for tightness or relative compactness can be derived. Let us consider here the case $\mathcal{X} = \boldsymbol{C}([0,\infty), \mathbb{R}^d)$ with the topology of uniform convergence on compact time intervals. With this choice, $\mathcal{X}$ is the canonical path space for ($\mathbb{R}^d$-valued) continuous processes. Let X be the canonical process on $\boldsymbol{C}([0,\infty), \mathbb{R}^d)$, that is, $X(t,\omega) \doteq \omega(t)$ for $t \geq 0$, $\omega \in \boldsymbol{C}([0,\infty), \mathbb{R}^d)$.

Theorem A.4. *Let I be a non-empty set, and let $(\theta_i)_{i\in I} \subset \mathcal{P}(\boldsymbol{C}([0,\infty), \mathbb{R}^d))$. Then $(\theta_i)_{i\in I}$ is relatively compact if and only if the following two conditions hold:*

(i) $(\theta_i \circ (X(0))^{-1}$ *is tight in* $\mathcal{P}(\mathbb{R}^d)$, *and*

(ii) *for every $\varepsilon > 0$, every $T \in \mathbb{N}$ there is $\delta > 0$ such that*

$$\sup_{i\in I} \theta_i \left(\left\{ \omega \in \boldsymbol{C}([0,\infty), \mathbb{R}^d) : \boldsymbol{w}_T(\omega, \delta) > \varepsilon \right\} \right) \leq \varepsilon,$$

where $\boldsymbol{w}_T(\omega,\delta) \doteq \sup_{s,t\in[0,T]:|t-s|\leq\delta} |\omega(t) - \omega(s)|$ *is the* modulus of continuity *of size δ over the time interval $[0,T]$.*

Theorem A.4 should be compared to the Arzelà-Ascoli criterion for relative compactness in $\boldsymbol{C}([0,\infty), \mathbb{R}^d)$. The next theorem gives a sufficient condition for relative compactness (or tightness) in $\mathcal{P}(\boldsymbol{C}([0,\infty), \mathbb{R}^d))$; the result should be compared to Kolmogorov's continuity theorem.

Theorem A.5 (Kolmogorov's sufficient condition)**.** *Let I be a non-empty set, and let $(\theta_i)_{i\in I} \subset \mathcal{P}(\boldsymbol{C}([0,\infty), \mathbb{R}^d))$. Suppose that*

(i) $(\theta_i \circ (X(0))^{-1}$ *is tight in* $\mathcal{P}(\mathbb{R}^d)$, *and*

(ii) *there are strictly positive numbers C, α, β such that for all $t, s \in [0,\infty)$, all $i \in I$,*

$$\mathbf{E}_{\theta_i}\left[|X(s) - X(t)|^\alpha \right] \leq C|t-s|^{1+\beta}.$$

Then $(\theta_i)_{i\in I}$ is relatively compact in $\mathcal{P}(\boldsymbol{C}([0,\infty), \mathbb{R}^d))$.

References

P. Billingsley. *Convergence of Probability Measures.* Wiley series in Probability and Statistics. John Wiley & Sons, New York, 2nd edition, 1999.

S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence.* Wiley Series in Probability and Statistics. John Wiley & Sons, New York, 1986.

A. Friedman. *Partial Differential Equations of Parabolic Type.* Prentice-Hall, Englewood Cliffs (NJ), 1964.

O. Kallenberg. On the existence of universal functional solutions to classical SDEs. *Ann. Probab.*, 24(1):196–205, 1996.

I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics.* Springer, New York, 2nd edition, 1991.

A. Klenke. *Probability Theory. A Comprehensive Course.* Springer, London, 2008.

K. Oelschläger. A martingale approach to the law of large numbers for weakly interacting stochastic processes. *Ann. Probab.*, 12(2):458–479, 1984.

D. W. Stroock and S. R. S. Varadhan. *Multidimensional Diffusion Processes*, volume 233 of *Grundlehren der mathematischen Wissenschaften.* Springer-Verlag, Berlin, 1979.

T. Yamada and S. Watanabe. On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.*, 11(1):155–167, 1971.