Probability measure-valued polynomial diffusions

Christa Cuchiero[*][§]   Martin Larsson[†][¶]   Sara Svaluto-Ferro[‡][§][¶]

Abstract

We introduce a class of probability measure-valued diffusions, coined *polynomial*, of which the well-known Fleming–Viot process is a particular example. The defining property of finite dimensional polynomial processes considered in [8, 21] is transferred to this infinite dimensional setting. This leads to a representation of conditional marginal moments via a finite dimensional linear PDE, whose spatial dimension corresponds to the degree of the moment. As a result, the tractability of finite dimensional polynomial processes are preserved in this setting. We also obtain a representation of the corresponding extended generators, and prove well-posedness of the associated martingale problems. In particular, uniqueness is obtained from the duality relationship with the PDEs mentioned above.

Contents

[*]Vienna University of Economics and Business, Austria. E-mail: christa.cuchiero@wu.ac.at
[†]ETH Zurich, Switzerland. E-mail: martin.larsson@math.ethz.ch
[‡]University of Vienna, Austria. E-mail: sara.svaluto-ferro@univie.ac.at

1 Introduction

In this paper we develop probability measure-valued versions of a class of processes known as polynomial diffusions, which have – due to their inherent tractability – broad applications in population genetics, interacting particle systems, and finance; see e.g. [14, 33, 19]. The result is a class of stochastic processes that model randomly evolving probability measures, including examples such as Fleming–Viot processes [23, 16], as well as conditional laws of jump-diffusions on (subsets of) $\mathbb{R}^d$.

Finite dimensional polynomial diffusions form a rich class that includes Kimura diffusions [25], Wishart correlation matrices [1], and affine processes [13], just to name a few subclasses. See e.g. [8, 21, 22, 9] for further details and examples. This suggests transferring their defining property and tractability features to infinite dimensional processes. Such processes also appear as limits of empirically well-suited finite-dimensional polynomial models, whose limiting behavior is of key interest in population dynamics, but also in other areas, such as capital distribution curve modeling; see e.g. [31]. In a nutshell our contribution can be summarized as follows:

- it is the first time that polynomial processes in *infinite dimension* are considered;

- important examples of polynomial probability measure-valued diffusions, e.g. the Fleming–Viot and conditional law processes mentioned above, have not previously been considered from this angle. In fact, analogous definitions and methods apply to more general measure-valued diffusions such as super-Brownian motion and the Dawson–Watanabe superprocess;

- a remarkably large class of processes is characterized via the polynomial property; in this paper, we find necessary and sufficient conditions for measure valued diffusions to be polynomial and to take values in the space of probability measures;

- to achieve this, we rely on the maximum principle of the respective operators; we obtain new optimality conditions which are applicable for *all* martingale problems on the space of probability measures.

More precisely, the infinite dimensional setup that we consider here are polynomial diffusions X taking values in the space of probability measures on a locally compact

Polish space E. We define them as path-continuous solutions of martingale problems for certain operators L acting on classes of cylinder polynomials, i.e. functions p of the form

$$p(\nu) = \phi\left(\int_E g_1(x)\nu(dx), \ldots, \int_E g_m(x)\nu(dx)\right),$$

where ϕ is a polynomial in m variables, $g_1, \ldots, g_m$ are continuous and bounded, and the argument ν is a probability measure. Any such function p can be regarded as a homogeneous polynomial in the probability measure ν, and admits a natural notion of degree as discussed in Section 2. The defining property of a probability measure-valued polynomial diffusion is that Lp is again a homogeneous polynomial with the same degree as p (or is the zero polynomial). The precise definitions are actually somewhat more general; the details are in Section 4.

A consequence is that moments admit tractable representations. Specifically, for a probability measure-valued polynomial diffusion X starting at $X_0 = \nu$, we establish in Section 5.1 (under suitable conditions) the moment formula

$$\mathbb{E}\left[\int_{E^k} g(x_1, \ldots, x_k) X_t(dx_1) \cdots X_t(dx_k)\right]$$
$$= \int_{E^k} u(t, x_1, \ldots, x_k)\nu(dx_1) \cdots \nu(dx_k), \tag{1.1}$$

where $u(t, x_1, \ldots, x_k)$ solves the linear partial integro-differential equation (PIDE)

$$\frac{\partial u}{\partial t} = \overline{L}_k u \quad \text{in } (0, \infty) \times E^k \tag{1.2}$$

with initial data $u(0, x_1, \ldots, x_k) = g(x_1, \ldots, x_k)$, and where $\overline{L}_k$ is a linear operator derived from the generator L of X, acting on (a subspace of) $C(E^k)$. The k-dimensional PIDE (1.2) is significantly simpler than the Kolmogorov equation, whose state space in this context consists of measures on E. Indeed, (1.2) corresponds to the Feynman-Kac PIDE associated to an E^k-valued Markov process. When E consists of finitely many points, we recover the finite dimensional case where (1.2) reduces to a linear ODE associated to a certain Markov chain with values in E^k, whose solution is computed by matrix exponentiation. These PIDEs fall exactly in the setup considered in [3], who develop numerical solution procedures based on neural networks. These methods do not suffer from the curse of dimensionality, and give the whole function $(x_1, \ldots, x_k) \mapsto u(t, x_1, \ldots, x_k)$.

The moment formula forms a particular instance of *duality*, which is often used to prove uniqueness for measure-valued martingale problems. This is the case also here, and we obtain uniqueness under broad circumstances. Being solutions of PIDEs, our dual "processes" are deterministic, in contrast to other commonly used duals such as the Kingman coalescent in the Fleming–Viot case; see e.g. [12]. Note that moment formulas for Fleming–Viot type process are classical (see e.g. [12] or [11, Section 2.8]); we show here that they are actually available much more broadly.

Existence of measure-valued processes is often proved via large population limits of carefully constructed particle systems; see e.g. [11, Section 2] for a semigroup approach, or [16, 15] for an approach via martingale problems. We also work with martingale problems, but rather than using approximations by finite particle systems, we obtain existence directly via the positive maximum principle. This relies on new optimality conditions for polynomials of measure arguments developed in Section 3. As a result, we can describe large parametric families of specifications. In particular, we obtain a full characterization of probability measure-valued polynomial diffusions whose generator L has a sufficiently large domain. This yields extensions of the so-called *Fleming–Viot*

process with weighted sampling discussed in [11, Section 5.7.8], where the sampling-replacement rate is allowed to depend on the type. On the other hand, by restricting the domain of L we obtain a richer class, including e.g. the model of exchangeable diffusions considered in [32]; see also [11, Section 5.8.1]. Our existence and well-posedness results are in Section 5.2.

Arguably, the bulk of applications of measure-valued processes come from population genetics. But tractable specifications like those developed here are for instance also of interest in non-parametric Bayesian statistics (see e.g. [29, 30] who consider distributions of functionals of random probability measures), age distribution and longevity risk modeling (see e.g. [5]), or high-dimensional financial modeling. Let us sketch a situation from stochastic portfolio theory (see [18, 20] for an introduction to this subject.) Let Z be a process with values in the unit simplex $\Delta^d = \{z \in [0,1]^d \colon z_1 + \ldots + z_d = 1\}$, representing the capitalization weights of d stocks. For tractability, it is natural select Z to be a polynomial diffusion on Δ^d as in [7]. To compute basic moment statistics of the capitalization weights, one uses a moment formula similar to (1.1). For homogeneous polynomials $q(z_1, \ldots, z_d)$, it takes the form

$$\mathbb{E}[q(Z_t)] = \sum_{\boldsymbol{\alpha}} u_t(\boldsymbol{\alpha})\, z_1^{\alpha_1} \cdots z_d^{\alpha_d},$$

where $Z_0 = z \in \Delta^d$, and the sum extends over all multi-indices $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)$ with $|\boldsymbol{\alpha}| = \alpha_1 + \cdots + \alpha_d = k := \deg(q)$. There are $N := \binom{k+d-1}{k}$ such multi-indices, and the $\mathbb{R}^N$-valued function $u_t = (u_t(\boldsymbol{\alpha}) \colon |\boldsymbol{\alpha}| = k)$ solves the linear ODE

$$\frac{\partial u}{\partial t} = L_k u, \tag{1.3}$$

whose initial condition is the coefficient vector of q, and where L_k here is an $N \times N$ matrix derived from the generator of Z. For small or moderate dimensions d and degrees k, solving (1.3) is feasible. However, d is typically on the order of 10^3, which renders (1.3) computationally taxing even for small k, since the ODE dimension is $N \sim d^k$.

Now, consider instead a *linear factor model* $\widetilde{Z} = (\widetilde{Z}^1, \ldots, \widetilde{Z}^d)$ for the capitalization weights. This means that $\widetilde{Z}^i = \int_E g_i(x) X_t(dx)$ for some nonnegative functions $g_1, \ldots, g_d$ that sum to one, and a probability measure-valued polynomial diffusion X with, say, $E = [0,1]$. In this case,

$$\mathbb{E}[q(\widetilde{Z}_t)] = \mathbb{E}[p(X_t)]$$

for some measure polynomial $p(\nu)$ of degree $k = \deg(q)$. This expectation can be computed using the moment formula (1.1), which amounts to solving the PDE (1.2) up to time t. Discretizing the space domain E^k using n points in each dimension yields a complexity of order n^k. This can be made orders of magnitude smaller than the complexity d^k of solving (1.3). Importantly, n is a parameter that is chosen based on accuracy requirements, while d is an input to the problem. This illustrates how probability measure-valued polynomial diffusions can enhance tractability in high-dimensional models. On top of this, as projections of an infinite-dimensional process, these linear factor models constitute a much richer class than polynomial models on subsets of Δ^d.

The remainder of the paper is organized as follows. After reviewing some basic notation and definitions in the following subsection, we turn to polynomials of measure arguments in Section 2, and prove optimality conditions for such polynomials in Section 3. In Section 4 we define polynomial operators and study their form in the diffusion case. Section 5 contains the moment formula as well as our main results on well-posedness of the martingale problem. Applications and examples are treated in Section 6. Some proofs and supplementary material are gathered in appendices.

1.1 Notation and basic definitions

Throughout this paper, E is a locally compact Polish space endowed with its Borel σ-algebra. The following notation is used.

- $M_+(E)$ denotes the finite measures on E, $M_1(E) \subset M_+(E)$ the probability measures, and $M(E) = M_+(E) - M_+(E)$ the signed measures of bounded variation (i.e., of the form $\nu_+ - \nu_-$ with $\nu_+, \nu_- \in M_+(E)$). All three are topologized by weak convergence, which turns $M_+(E)$ and $M_1(E)$ into Polish spaces. For $\mu, \nu \in M(E)$ we write $\mu \leq \nu$ if $\nu - \mu \in M_+(E)$ and $|\nu|$ for $\nu_+ + \nu_-$.

- $C(E), C_b(E), C_0(E), C_c(E)$ have the usual meaning of continuous (and bounded, and vanishing at infinity, and compactly supported) real functions on E. The topology on the latter three is that of uniform convergence, and $\|\cdot\|$ denotes the supremum norm.

- If E is noncompact, then $E^\Delta = E \cup \{\Delta\}$ is the one-point compactification, itself a compact Polish space. If E is compact we write $E^\Delta = E$, which mitigates the need to consider the compact and noncompact cases separately. We also define

$$C_\Delta(E^k) := \big\{ f|_{E^k} \ : \ f \in C((E^\Delta)^k) \big\},$$

a closed subspace of $C_b(E^k)$. The spaces $C_\Delta(E)$ and $C(E^\Delta)$ can be identified, and we occasionally regard elements of the former as elements of the latter, and vice versa. When E is compact, we have $C(E) = C_b(E) = C_0(E) = C_c(E) = C_\Delta(E)$ and we then simply write $C(E)$. Note that the constant function 1 lies in $C_\Delta(E)$, but of course not in $C_0(E)$. This is one reason the spaces $C_\Delta(E^k)$ are useful; other reasons are discussed in Remarks 2.6 and 4.6.

- $\widehat{C}_\Delta(E^k)$ is the closed subspace of $C_\Delta(E^k)$ consisting of symmetric functions f, i.e., $f(x_1, \ldots, x_k) = f(x_{\sigma(1)}, \ldots, x_{\sigma(k)})$ for all $\sigma \in \Sigma_k$, the permutation group on k elements. $\widehat{C}_0(E^k)$ and $\widehat{C}(E^k)$ are defined similarly. For any $g \in \widehat{C}_\Delta(E^k), h \in \widehat{C}_\Delta(E^\ell)$ we denote by $g \otimes h \in \widehat{C}_\Delta(E^{k+\ell})$ the symmetric tensor product, given by

$$
\begin{aligned}
&(g \otimes h)(x_1, \ldots, x_{k+\ell}) \\
&= \frac{1}{(k+\ell)!} \sum_{\sigma \in \Sigma_{k+\ell}} g\big(x_{\sigma(1)}, \ldots, x_{\sigma(k)}\big) h\big(x_{\sigma(k+1)}, \ldots, x_{\sigma(k+\ell)}\big).
\end{aligned} \tag{1.4}
$$

For a linear subspace $D \subseteq C_\Delta(E)$ we set $D \otimes D := \operatorname{span}\{g \otimes g \ : \ g \in D\}$. We emphasize that only symmetric tensor products are used in this paper.

Two key notions are the *positive maximum principle* and *conservativity* for certain linear operators. In general, for a Polish space $\mathcal{X}$ and a subset $\mathcal{S} \subseteq \mathcal{X}$, these notions are defined as follows. An operator $\mathcal{A} \colon D \to C_b(\mathcal{X})$ with domain $D \subseteq C_b(\mathcal{X})$ is said to satisfy the positive maximum principle on $\mathcal{S}$ if

$$f \in D, \ x \in \mathcal{S}, \ \sup_{\mathcal{S}} f = f(x) \geq 0 \quad \text{implies} \quad \mathcal{A}f(x) \leq 0.$$

If $\mathcal{S}$ locally compact, $\mathcal{A}$ is called $\mathcal{S}$-conservative if there exist functions $f_n \in D \cap C_0(\mathcal{S})$ such that $\lim_{n\to\infty} f_n = 1$ on E and $\lim_{n\to\infty} (\mathcal{A}f_n)^- = 0$ on E^Δ, both in the bounded pointwise sense; c.f. Chapter 4.2 in [17]. For us, $\mathcal{S}$ will be E, E^Δ, $M_1(E)$, or $M_1(E^\Delta)$.

It is well-known that the positive maximum principle, combined with conservativity, is essentially equivalent to the existence of $\mathcal{S}$-valued solutions to the martingale problem for $\mathcal{A}$; see for instance Theorem 4.5.4 of [17]. We use this extensively, and review the relevant results in Section D. Here an important issue is that while $M_1(E)$ is compact when E is compact, $M_1(E)$ is not even locally compact when E is noncompact.

2 Polynomials of measure arguments

In this section we develop basic properties of polynomials of measure arguments. The notation and results introduced here play a central role throughout this paper. Throughout this section, E is a locally compact Polish space.

2.1 Monomials and polynomials

A *monomial* on $M(E)$ is an expression of the form

$$\langle g, \nu^k \rangle = \int_{E^k} g(x_1, \ldots, x_k) \nu(dx_1) \cdots \nu(dx_k)$$

for some $k \in \mathbb{N}_0$, where $g \in \widehat{C}_\Delta(E^k)$ is referred to as the *coefficient* of the monomial; see e.g. [11, Chapter 2]. We identify $\widehat{C}_\Delta(E^0)$ with $\mathbb{R}$, so that for $k = 0$ we have $\langle g, \nu^0 \rangle = g \in \mathbb{R}$. It is clear that the map $\nu \mapsto \langle g, \nu^k \rangle$ is homogeneous of degree k, and that $g \mapsto \langle g, \nu^k \rangle$ is linear. Furthermore, one has the identity $\langle g, \nu^k \rangle \langle h, \nu^\ell \rangle = \langle g \otimes h, \nu^{k+\ell} \rangle$, where the symmetric tensor product $g \otimes h$ is defined in (1.4).

A polynomial on $M(E)$ is now defined as a (finite) linear combination of monomials,

$$p(\nu) = \sum_{k=0}^m \langle g_k, \nu^k \rangle, \tag{2.1}$$

with coefficients $g_k \in \widehat{C}_\Delta(E^k)$. The degree of the polynomial $p(\nu)$, denoted by $\deg(p)$, is the largest k such that g_k is not the zero function, and $-\infty$ if p is the zero polynomial. The representation (2.1) is unique; see Corollary 2.4 below.

Example 2.1. Let $E = \{1, \ldots, d\}$ be a finite set. Then every element $\nu \in M(E)$ is of the form

$$\nu = z_1 \delta_1 + \cdots + z_d \delta_d, \qquad (z_1, \ldots, z_d) \in \mathbb{R}^d,$$

where δ_i is the Dirac mass concentrated at $\{i\}$. Monomials take the form

$$\langle g, \nu^k \rangle = \sum_{i_1, \ldots, i_k} g(i_1, \ldots, i_k) z_{i_1} \cdots z_{i_k},$$

where the summation ranges over $E^k = \{1, \ldots, d\}^k$. Therefore, as $g(\cdot)$ ranges over all symmetric functions on E^k, we recover all homogeneous polynomials of total degree k in the d variables $z_1, \ldots, z_d$. In particular, in view of Corollary 2.5 later, this relation provides a one to one correspondence between polynomials on the unit simplex Δ^d, namely

$$\Delta^d := \Big\{ z \in \mathbb{R}^d \colon \sum_{i=1}^d z_i = 1, \ z_i \geq 0 \Big\},$$

and polynomials on $M_1(E)$.

The following function space will play an important role.

Definition 2.2. *Let*

$$P := \{ \nu \mapsto p(\nu) \colon p \text{ is a polynomial on } M(E) \}$$

denote the algebra of all polynomials on $M(E)$ regarded as real-valued maps, equipped with pointwise addition and multiplication.

2.2 Continuity and smoothness of polynomials

Just like ordinary polynomials, the elements of P are smooth. This is made precise in Lemma 2.3 below. In its statement, we use a directional derivative of functions on $M(E)$ that is well-known since the work of [23]. A function $f\colon M(E) \to \mathbb{R}$ is called differentiable at ν in direction δ_x for $x \in E$ if

$$\partial_x f(\nu) := \lim_{\varepsilon \to 0} \frac{f(\nu + \varepsilon \delta_x) - f(\nu)}{\varepsilon}$$

exists. We write $\partial p(\mu)$ for the map $x \mapsto \partial_x p(\mu)$, and use the notation

$$\partial^k_{x_1 x_2 \cdots x_k} f(\nu) := \partial_{x_1} \partial_{x_2} \cdots \partial_{x_k} f(\nu)$$

for iterated derivatives. We write $\partial^k p(\nu)$ for the corresponding map from E^k to $\mathbb{R}$. Observe that for $p \in P$ of the form $p(\nu) = \langle g, \nu \rangle$ we get $\partial_x p(\nu) = \lim_{\varepsilon \to 0}(\int g(y) \varepsilon \delta_x(dy)) \varepsilon^{-1} = g(x)$ for each $x \in E$.

The following lemma asserts basic properties of polynomials, in particular that polynomials on $M(E)$ can be uniquely extended to polynomials on $M(E^\Delta)$, which will often be the object of interest for our purposes.

Lemma 2.3. *(i) Each $p \in P$ is continuous on $M_+(E)$, sequentially continuous on $M(E)$, and can be uniquely extended to a polynomial on $M(E^\Delta)$.[1]*

(ii) Let $p \in P$ be a monomial of the form $p(\nu) = \langle g, \nu^k \rangle$. Then, for every $x \in E$ and $\nu \in M(E)$,

$$\partial_x p(\nu) = k\langle g(\,\cdot\,, x), \nu^{k-1} \rangle,$$

where $g(\,\cdot\,, x)$ is the function $(x_1, \ldots, x_{k-1}) \mapsto g(x_1, \ldots, x_{k-1}, x)$. If $k = 0$, the right-hand side should be read as zero.

(iii) For each $p \in P$ and $x \in E$ the map $\partial_x p\colon \nu \mapsto \partial_x p(\nu)$ lies in P.

(iv) For each $p \in P$ and $\nu \in M(E)$, the map $\partial p(\nu)\colon x \mapsto \partial_x p(\nu)$ lies in $C_\Delta(E)$.

(v) The identity

$$\partial_x(pq)(\nu) = p(\nu)\partial_x q(\nu) + q(\nu)\partial_x p(\nu)$$

holds for all $p, q \in P$, $x \in E$, $\nu \in M(E)$.

(vi) The Taylor representation

$$p(\nu + \mu) = \sum_{\ell=0}^{k} \frac{1}{\ell!} \langle \partial^\ell p(\nu), \mu^\ell \rangle,$$

holds for all $p \in P$ and $\nu, \mu \in M(E)$, where k denotes the degree of p.

Proof. (i): For $h \in C_\Delta(E)^{\otimes k}$ we can write $h = \sum_{\ell=1}^{L} \lambda_\ell h_\ell^{\otimes k}$ for some $h_\ell \in C_\Delta(E)$ and $\lambda_\ell \in \mathbb{R}$. Since $\langle h_\ell, \nu \rangle$ is continuous by definition of weak convergence,

$$\langle h, \nu^k \rangle = \sum_{\ell=1}^{L} \lambda_\ell \langle h_\ell^{\otimes k}, \nu^k \rangle = \sum_{\ell=1}^{L} \lambda_\ell \langle h_\ell, \nu \rangle^k$$

is continuous as well. Note then that by linearity in (2.1) it is enough to prove the result for $p(\nu) = \langle g, \nu^k \rangle$ and $g \in \widehat{C}_\Delta(E^k)$. Choose $h \in C_\Delta(E)^{\otimes k}$ such that $\|g - h\| \le \varepsilon$ and let

---

[1] It can be shown that sequential continuity cannot be strengthened to continuity.

$\nu_n \in M(E)$ form a convergent sequence with limit $\nu \in M(E)$. Observe that, by the Banach–Steinhaus theorem, $\sup_n |\nu_n|(E) < \infty$. Then

$$\left|\langle g, \nu_n^k \rangle - \langle g, \nu^k \rangle\right| \leq \left|\langle h, \nu_n^k \rangle - \langle h, \nu^k \rangle\right| + \varepsilon\left(\sup_n |\nu_n|(E)^k + |\nu|(E)^k\right) \to C\varepsilon$$

for some $C \geq 0$. Since ε is arbitrary, this proves sequential continuity of p on $M(E)$. In particular we get continuity on $M_+(E)$ since this is a Polish space. The last part follows from the observation that every function in $C_\Delta(E)$ can be uniquely extended to a function in $C(E^\Delta)$.

(ii): Using the symmetry of g, a direct calculation yields

$$p(\nu + \varepsilon\delta_x) - p(\nu) = \varepsilon k \int g(x_1, \ldots, x_{k-1}, x) \prod_{j=1}^{k-1} \nu(dx_j) + o(\varepsilon).$$

The expression for $\partial_x p(\nu)$ follows.

For the remaining part of the proof it suffices to consider monomials $p(\nu) = \langle g, \nu^k \rangle$ for $g \in \widehat{C}_\Delta(E^k)$ due to the linearity in (2.1).

(iii): Fix $x \in E$ and note that $kg(\,\cdot\,, x) \in \widehat{C}_\Delta(E^{k-1})$. The claim follows by (ii).

(iv): For $p(\nu) = \langle g, \nu^k \rangle$, $|\partial_x p(\nu)| = |\langle kg(\,\cdot\,, x), \nu^{k-1} \rangle| \leq k\|g\||\nu|(E)^{k-1} < \infty$. Continuity of $x \mapsto \partial_x p(\nu)$ follows from the dominated convergence theorem and the fact that E is Polish, and thus a sequential space.

(v): For monomials $p(\nu) = \langle g, \nu^k \rangle$ and $q(\nu) = \langle h, \nu^\ell \rangle$, we have $pq(\nu) = \langle g \otimes h, \nu^{k+\ell} \rangle$. Since for all $x \in E$ and $\nu \in M(E)$

$$(k + \ell)\langle g \otimes h(\,\cdot\,, x), \nu^{k+\ell-1} \rangle = k\langle g(\,\cdot\,, x), \nu^{k-1} \rangle\langle h, \nu^\ell \rangle + \ell\langle g, \nu^k \rangle\langle h(\,\cdot\,, x), \nu^{\ell-1} \rangle,$$

the claim follows by (ii).

(vi): Observing that for $p(\nu) := \langle g, \nu^k \rangle$

$$p(\nu + \mu) = \sum_{\ell=0}^{k} \binom{k}{\ell} \int g(x_1, \ldots, x_k) \prod_{i=\ell+1}^{k} \nu(dx_i) \prod_{i=1}^{\ell} \mu(dx_i)$$

the result follows by (ii). $\qquad\square$

From Lemma 2.3(ii) one can deduce the uniqueness of the representation (2.1).

Corollary 2.4. *Suppose $p(\nu) = \sum_{k=0}^{m} \langle g_k, \nu^k \rangle$ equals zero for all $\nu \in M(E)$. Then $g_k = 0$ for all k.*

Proof. Let $x_1, \ldots, x_m \in E$ be arbitrary and differentiate m times using Lemma 2.3(ii) to get $m! g_m(x_1, \ldots, x_m) = \partial_{x_1 x_2 \cdots x_m} p(\nu) = 0$. Thus $g_m = 0$. Now repeat this successively for $g_{m-1}, g_{m-2}, \ldots, g_0$. $\qquad\square$

The following property turns out to be particularly useful in the context of the moment formula. In the finite-dimensional setting, the result states that every polynomial on the unit simplex has a homogeneous representative.

Corollary 2.5. *Every polynomial on $M(E)$ has a unique homogeneous representative on $M_1(E)$. That is, for every $p \in P$ with $\deg(p) = m$ there is a unique $g \in \widehat{C}_\Delta(E^m)$ such that*

$$p(\nu) = \langle g, \nu^m \rangle \qquad \text{for all } \nu \in M_1(E).$$

Proof. Corollary 2.4 yields a unique set of coefficients $g_0, \ldots, g_m$ with $g_k \in \widehat{C}_\Delta(E^k)$ and $p(\nu) = \sum_{k=0}^{m} \langle g_k, \nu^k \rangle$. The result now follows by setting $g := \sum_{k=0}^{m} g_k \otimes 1^{\otimes(m-k)}$. $\qquad\square$

Remark 2.6. If we choose to work with coefficients in $\widehat{C}_0(E^k)$ instead of $\widehat{C}_\Delta(E^k)$ we would obtain the same class of polynomials on $M_1(E)$. This is because every $g \in \widehat{C}_\Delta(E^k)$ equals $\sum_{i=0}^{k} g_i \otimes 1^{\otimes(k-i)}$ for some $g_i \in \widehat{C}_0(E^i)$, and therefore $\langle g, \nu^k \rangle = \sum_{i=0}^{k} \langle g_i, \nu^i \rangle$ for all $\nu \in M_1(E)$. Indeed, the g_i are given iteratively by

$$g_0 := g(\Delta, \ldots, \Delta) \quad \text{and} \quad g_i := \binom{k}{i}\Big(g(\Delta, \ldots, \Delta, \cdot) - \sum_{j=0}^{i-1} g_j \otimes 1^{\otimes(i-j)}\Big).$$

However, not every such polynomial admits a homogenous representative on $M_1(E)$ in the sense of Corollary 2.5, unless E is compact. An example is $1 + \langle g, \nu \rangle$ with $g \in C_0(E)$ nonzero. The existence of homogeneous representatives leads to significant notational simplifications when E is not compact (see Remark 4.6 for more details). This is the main reason for working with the spaces $\widehat{C}_\Delta(E^k)$.

2.3 Polynomials with regular coefficients

For a polynomial p, the derivative map $x \mapsto \partial_x p(\nu)$ is only as regular as the coefficients of p. This leads us to consider subspaces of polynomials with more regular coefficients. Let $D \subseteq C_\Delta(E)$ be a dense linear subspace containing the constant function 1 and define

$$P^D := \text{span}\big\{1, \ \langle g, \nu \rangle^k \colon k \geq 1, \ g \in D\big\}. \tag{2.2}$$

Thus P^D is the subalgebra of P consisting of all (finite) linear combinations of the constant polynomial and "rank-one" monomials $\langle g \otimes \cdots \otimes g, \nu^k \rangle = \langle g, \nu \rangle^k$ with $g \in D$. Equivalently, P^D consists of all polynomials $p(\nu) = \phi(\langle g_1, \nu \rangle, \ldots, \langle g_k, \nu \rangle)$ with $k \in \mathbb{N}$, $g_1, \ldots, g_k \in D$, and ϕ a polynomial on $\mathbb{R}^k$.

Lemma 2.7. *For any $p \in P^D$ and $\nu \in M(E)$, we have $\partial^k p(\nu) \in D^{\otimes k}$. Moreover P^D is dense in $C(M_1(E^\Delta))$. Here the elements of P^D are viewed as functions on $M_1(E^\Delta)$ by first extending them to $M(E^\Delta)$ using Lemma 2.3 (i) and then restricting them to $M_1(E^\Delta)$.*

Proof. For $p(\nu) := \phi(\langle g, \nu \rangle)$ where ϕ is polynomial we have $\partial^k p(\nu) = \phi^{(k)}(\langle g, \nu \rangle) g^{\otimes k} \in D^{\otimes k}$. Thus the first part of the result holds for all such p, and by linearity for all $p \in P^D$. For the second part, continuity of polynomials follows by Lemma 2.3(i). Stone–Weierstrass and the fact that D is densely contained in $C(E^\Delta)$ yield the density. $\square$

3 Optimality conditions

We now develop optimality conditions for polynomials of measure arguments, which are instrumental when working with the positive maximum principle on $M_1(E^\Delta)$. Our first result, Theorem 3.1, extends the classical first and second order Karush–Kuhn–Tucker conditions for functions on the finite-dimensional simplex (see e.g. [4]). It is derived by perturbing an optimizer $\nu_* \in M_1(E^\Delta)$ by shifting small amounts of mass to arbitrary points in E^Δ. Our second result, Theorem 3.4, is obtained by deforming the optimizer ν_* using a group of isometries of $C_\Delta(E)$. The resulting condition is genuinely infinite-dimensional; see Lemma 3.6. We will use the operator Ψ, which maps any function $g \colon E \times E \to \mathbb{R}^k$ to the function $\Psi(g) \colon E \times E \to \mathbb{R}^k$ given by

$$\Psi(g)(x, y) = \frac{1}{2}\left(g(x, x) + g(y, y) - 2g(x, y)\right). \tag{3.1}$$

Note that we use Lemma 2.3(i) to extend polynomials from $M_1(E)$ to $M_1(E^\Delta)$.

Theorem 3.1. *Let $p \in P$ and $\nu_* \in M_1(E^\Delta)$ satisfy $p(\nu_*) = \max_{M_1(E^\Delta)} p$. Then the following first and second order optimality conditions hold:*

(i) $\langle \partial p(\nu_*), \mu \rangle = \sup_E \partial p(\nu_*)$, *for all* $\mu \in M_1(E^\Delta)$ *such that* $\mathrm{supp}(\mu) \subseteq \mathrm{supp}(\nu_*)$. *In particular,*

$$\partial_x p(\nu_*) = \sup_E \partial p(\nu_*) \quad \textit{for all } x \in \mathrm{supp}(\nu_*). \tag{3.2}$$

(ii) $\langle \partial^2 p(\nu_*), \mu^2 \rangle \leq 0$ *for all signed measures* $\mu \in M(E^\Delta)$ *such that* $\langle 1, \mu \rangle = 0$ *and* $\mathrm{supp}(|\mu|) \subseteq \mathrm{supp}(\nu_*)$. *In particular,*

$$\Psi\big(\partial^2 p(\nu_*)\big)(x,y) \leq 0 \quad \textit{for all } x, y \in \mathrm{supp}(\nu_*). \tag{3.3}$$

Proof. (i): Pick any $x \in \mathrm{supp}(\nu_*)$ and $y \in E^\Delta$. For each $n \in \mathbb{N}$, let A_n be the ball of radius $1/n$ centered at x, intersected with $\mathrm{supp}(\nu_*)$. Then $\nu_*(A_n) > 0$, and the probability measures $\mu_n := \nu_*(\cdot \cap A_n)/\nu_*(A_n)$ converge weakly to δ_x as $n \to \infty$. Choose $\varepsilon_n \in (0, \nu_*(A_n))$. Then $\nu_* \geq \varepsilon_n \mu_n$ since for all $B \in \mathcal{B}(E)$

$$\nu_*(B) - \varepsilon_n \frac{\nu_*(B \cap A_n)}{\nu_*(A_n)} \geq \frac{\nu_*(B \cap A_n)}{\nu_*(A_n)}(\nu_*(A_n) - \varepsilon_n) \geq 0.$$

Hence $\nu_n := \nu_* + \varepsilon_n(\delta_y - \mu_n)$ is a probability measure. Maximality of ν_* and Lemma 2.3(vi) now give

$$0 \geq p(\nu_n) - p(\nu_*) = \varepsilon_n \langle \partial p(\nu_*), \delta_y - \mu_n \rangle + o(\varepsilon_n).$$

Dividing by ε_n, sending n to infinity, and using that $x \mapsto \partial_x p(\nu_*)$ is bounded and continuous, we obtain $\partial_x p(\nu_*) \geq \partial_y p(\nu_*)$. We deduce (3.2), which immediately implies (i).

(ii): In addition to the above, suppose y is in $\mathrm{supp}(\nu_*)$. Since we also have that $\mathrm{supp}(|\mu_n|) \subseteq \mathrm{supp}(\nu_*)$, we get $\langle \partial p(\nu_*), \delta_y - \mu_n \rangle = 0$ due to (i). Maximality of ν_* and Lemma 2.3(vi) then give

$$0 \geq p(\nu_n) - p(\nu_*) = \frac{1}{2}\varepsilon_n^2 \langle \partial^2 p(\nu_*), (\delta_y - \mu_n)^2 \rangle + o(\varepsilon_n^2),$$

and therefore $\langle \partial^2 p(\nu_*), (\delta_y - \delta_x)^2 \rangle \leq 0$. More generally, consider measures of the form

$$\nu_n := \nu_* + \varepsilon_n \left(\sum_{i=1}^m \lambda_i \delta_{y_i} - \sum_{i=1}^m \gamma_i \mu_{i,n} \right)$$

for some points $y_i \in \mathrm{supp}(\nu_*)$, convex weights $\lambda_1, \ldots, \lambda_m$ and $\gamma_1, \ldots, \gamma_m$, and $\mu_{i,n}$ constructed as μ_n above with x replaced by $x_i \in \mathrm{supp}(\nu_*)$. Letting ε_n decrease to zero sufficiently rapidly, the above argument gives $\langle \partial^2 p(\nu_*), \mu^2 \rangle \leq 0$ for the signed measure

$$\mu = \sum_{i=1}^m \lambda_i \delta_{y_i} - \sum_{i=1}^m \gamma_i \delta_{x_i}.$$

Passing to the weak closure yields (ii) with the additional restriction that the positive and negative parts of μ are probability measures. The general case is obtained by scaling. Finally, since $\langle \partial^2 p(\nu_*), (\delta_y - \delta_x)^2 \rangle = 2\Psi(\partial^2 p(\nu_*))(x,y)$ we obtain (3.3). $\square$

Remark 3.2. Note the similarity between Theorem 3.1 and the classical Karush–Kuhn–Tucker conditions on the finite-dimensional simplex Δ^d. Let $f \in C^2(\mathbb{R}^d)$ and $x^* \in \Delta^d$ satisfy $f(x^*) = \max_{\Delta^d} f$. Then the first and second order (necessary) Karush–Kuhn–Tucker conditions on Δ^d hold:

(i) For each $v \in \Delta^d$ such that $v_i = 0$ whenever $x_i^* = 0$, $\nabla f(x^*)^\top v = \max_{j \in \{1, \ldots, d\}} \frac{\partial f}{\partial x_j}(x^*)$.

(ii) For each $v \in \mathbb{R}^d$ such that $\mathbf{1}^\top v = 0$ and $v_i = 0$ whenever $x_i^* = 0$, $v^\top \nabla^2 f(x^*) v \leq 0$, where $\mathbf{1} := (1, \ldots, 1)^\top$.

Remark 3.3. Taking again $E = \{1, \ldots, d\}$ as example, the appearance of Ψ in (3.3) can be understood as follows. Suppose $z \in \Delta^d$ maximizes a function $f \in C^2(\mathbb{R}^d)$ over Δ^d. Then for every i, j such that $z_i > 0$ and $z_j > 0$, we must have $(e_i - e_j)^\top \nabla^2 f(z)(e_i - e_j) \leq 0$, where e_i is the i-th canonical unit vector. Indeed, otherwise $z \pm \varepsilon(e_i - e_j)$ would lie in Δ^d and give a higher function value for small $\varepsilon > 0$. More explicitly, we must have

$$\partial_{ii}^2 f(z) + \partial_{jj}^2 f(z) - 2\partial_{ij}^2 f(z) \leq 0,$$

where the left hand side is equal to $2\Psi(\partial^2 f(z))(i, j)$ on $E = \{1, \ldots, d\}$.

For the remainder of this section, $D \subseteq C_\Delta(E)$ is a linear subspace, and P^D is defined by (2.2).

Our next optimality condition is more subtle, in that it becomes trivial in the finite-dimensional case; see Lemma 3.6. The basic observation is that a group of isometries T_t of $C_\Delta(E)$ induces a flow of measures $\mu_t \in M_+(E^\Delta)$ via the formula $\langle g, \mu_t \rangle = \langle T_t g, \mu \rangle$ for every $g \in C_\Delta(E)$, where $\mu \in M_+(E^\Delta)$ is fixed. The value of a polynomial in its maximizer ν_* cannot be less than its value in $\nu_* - \mu + \mu_t$, for any t, and this leads to an optimality condition in terms of the group generator A.

For example, if $E = \mathbb{R}$, the generator could be $Ag = \tau g'$ for some $\tau \in C_\Delta^1(\mathbb{R})$. The isometries would then be $T_t g := g(\phi(t, \cdot))$, where ϕ solves $\frac{d}{dt}\phi(t, x) = \tau(\phi(t, x))$ with initial condition $\phi(0, x) = x$. The corresponding flow of measures would consist of the pushforwards of μ with respect to $\phi(t, \cdot)$. For more details see Lemma 6.1.

The tensor notation $A \otimes A$ is used to denote the linear operator from $D \otimes D$ to $\widehat{C}_\Delta(E^2)$ determined by

$$(A \otimes A)(g \otimes g) := (Ag) \otimes (Ag)$$

for a given linear operator $A \colon D \to C_\Delta(E)$.

Theorem 3.4. Fix $p \in P^D$ and $\nu_* \in M_1(E^\Delta)$ with $p(\nu_*) = \max_{M_1(E^\Delta)} p$. Let A be the generator of a strongly continuous group of positive isometries of $C_\Delta(E)$, and assume the domain of A contains both D and $A(D)$. Then

$$\langle A^2(\partial p(\nu_*)), \mu \rangle + \langle (A \otimes A)(\partial^2 p(\nu_*)), \mu^2 \rangle \leq 0$$

for every $\mu \in M_+(E^\Delta)$ with $\mu \leq \nu_*$.

Proof. Let $\{T_t\}_{t \in \mathbb{R}}$ be the group generated by A. For any $\mu \in M_+(E^\Delta)$, the group induces a flow of measures $\mu_t \in M(E^\Delta)$ via the formula $\langle g, \mu_t \rangle = \langle T_t g, \mu \rangle$ for $g \in C_\Delta(E)$. The positivity and isometry property of T_t implies that μ_t is nonnegative and has constant total mass $\mu_t(E^\Delta) = \mu(E^\Delta)$. Therefore, assuming henceforth that $\mu \leq \nu_*$, it follows that $\nu_* + \mu_t - \mu$ is a probability measure. Since $\|T_t g - g\| = O(t)$ for every $g \in D$, we have $\langle g, (\mu_t - \mu)^k \rangle = O(t^k)$ for every $g \in D^{\otimes k}$. Maximality of ν_* and Lemma 2.3(vi) then give

$$0 \geq p(\nu_* + \mu_t - \mu) - p(\nu_*) = \langle \partial p(\nu_*), \mu_t - \mu \rangle + \frac{1}{2}\langle \partial^2 p(\nu_*), (\mu_t - \mu)^2 \rangle + o(t^2)$$

$$= \langle (T_t - \mathrm{id})\partial p(\nu_*), \mu \rangle + \frac{1}{2}\langle (T_t \otimes T_t - 2\, T_t \otimes \mathrm{id} + \mathrm{id} \otimes \mathrm{id})\partial^2 p(\nu_*), \mu^2 \rangle + o(t^2). \quad (3.4)$$

We claim that both A and $-A$ satisfy the positive maximum principle on E^Δ. Indeed, for $f \in D$ and $x \in E^\Delta$ with $f(x) = \max_{E^\Delta} f \geq 0$, the positivity and isometry property give

$$T_t f(x) \leq T_t f^+(x) \leq \|T_t f^+\| = \|f^+\| = f(x). \quad (3.5)$$

We thus have $Af(x) = \lim_{t\downarrow 0}(T_t f(x) - f(x))/t \leq 0$ as well as $-Af(x) = \lim_{t\downarrow 0}(T_{-t}f(x) - f(x))/t \leq 0$, proving the claim. Since $\partial_x p(\nu_*) = \sup_E \partial p(\nu_*)$ for all $x \in \mathrm{supp}(\nu^*)$ due to Theorem 3.1, it follows that $A(\partial p(\nu_*))(x) = 0$ for all such x. As a result, using that $\mathrm{supp}(\mu) \subseteq \mathrm{supp}(\nu_*)$ and that the domain of A contains $A(D)$, we get

$$\langle (T_t - \mathrm{id})\partial p(\nu_*), \mu \rangle = \langle (T_t - \mathrm{id} - tA)\partial p(\nu_*), \mu \rangle = \frac{1}{2}t^2\langle A^2(\partial p(\nu_*)), \mu \rangle + o(t^2). \qquad (3.6)$$

Furthermore, using that

$$(T_t \otimes T_t - 2\, T_t \otimes \mathrm{id} + \mathrm{id} \otimes \mathrm{id})(g \otimes g) = (T_t g - g) \otimes (T_t g - g)$$

for all $g \in D$, we deduce that

$$\langle (T_t \otimes T_t - 2\, T_t \otimes \mathrm{id} + \mathrm{id} \otimes \mathrm{id})g, \mu^2 \rangle = t^2\langle (A \otimes A)g, \mu^2 \rangle + o(t^2) \qquad (3.7)$$

for all $g \in D \otimes D$. Inserting (3.6) and (3.7) into (3.4), dividing by t^2, and sending t to zero yields

$$0 \geq \frac{1}{2}\langle A^2(\partial p(\nu_*)), \mu \rangle + \frac{1}{2}\langle (A \otimes A)\partial^2 p(\nu_*), \mu^2 \rangle.$$

This completes the proof. $\qquad\square$

Remark 3.5. We claim that for A as in Theorem 3.4, the operator A^2 satisfies the positive maximum principle on E^Δ. Indeed, let $f \in D$ and $x \in E^\Delta$ with $f(x) = \max_{E^\Delta} f \geq 0$. Then, as in (3.5) and with the same notation, we have $T_t f(x) \leq f(x)$, and $Af(x) = 0$ since both A and $-A$ satisfy the positive maximum principle on E^Δ. Hence $A^2 f(x) = \lim_{t\downarrow 0}(T_t f(x) - f(x) - Af(x))/t \leq 0$, which proves the claim.

The following lemma illustrates the pure infinite-dimensional nature of the condition provided in Theorem 3.4.

Lemma 3.6. *Let A be the generator of a strongly continuous group of positive isometries of $C_\Delta(E)$. If the domain of A is all of $C_\Delta(E)$, then $A = 0$. This is in particular the case if A is bounded or E consists of finitely many points.*

Proof. Both A and $-A$ satisfy the positive maximum principle on E, and $A1 = 0$. Therefore Lemma C.2 implies that A and $-A$ are both of the form (C.1) with $B = \pm A$. As a result,

$$0 = Ag(x) - Ag(x) = \int (g(\xi) - g(x))(\nu_A + \nu_{-A})(x, d\xi)$$

for all $x \in E$ and $g \in C(E^\Delta)$. This implies that $1_{\{x\}^c}(\xi)\nu_A(x, d\xi)$ and $1_{\{x\}^c}(\xi)\nu_{-A}(x, d\xi)$ are zero for all $x \in E$ and hence that $A = 0$. Since each linear operator on a finite-dimensional vector space is bounded, and the domain of a bounded operator on $C_\Delta(E)$ can be extended to all of $C_\Delta(E)$, the second part follows. $\qquad\square$

4 Polynomial operators

Let E be a locally compact Polish space. We now define polynomial operators, which constitute a class of possibly unbounded linear operators acting on polynomials. They are not defined on all of P in general, but only on the subspace P^D for some dense subspace $D \subseteq C_\Delta(E)$; see (2.2). An analog of this notion has appeared previously in connection with finite-dimensional polynomial processes; see e.g. [8, 21, 9].

Definition 4.1. *Fix $\mathcal{S} \subseteq M(E)$. A linear operator $L: P^D \to P$ is called $\mathcal{S}$-polynomial if for every $p \in P^D$ there is some $q \in P$ such that $q|_\mathcal{S} = Lp|_\mathcal{S}$ and*

$$\deg(q) \leq \deg(p).$$

Given a linear operator $L\colon P^D \to P$, its associated carré-du-champ operator is the symmetric bilinear map $\Gamma\colon P^D \times P^D \to P$ defined by

$$\Gamma(p, q) = L(pq) - pLq - qLp. \tag{4.1}$$

The carré-du-champ operator gives information about the quadratic variation of the martingales appearing in the martingale problem for the operator L. It also gives information about path continuity of solutions to such martingale problems. We return to this issue in Lemma 5.2, which roughly speaking states that path continuity holds precisely when the carré-du-champ operator Γ is a derivation, which is defined as follows.

Definition 4.2. *Fix $\mathcal{S} \subseteq M(E)$. A symmetric bilinear map $\Gamma\colon P^D \times P^D \to P$ is called an $\mathcal{S}$-derivation if for all $p, q, r \in P^D$, $\Gamma(pq, r) = p\Gamma(q, r) + q\Gamma(p, r)$ on $\mathcal{S}$.*

For a finite-dimensional diffusion it is known that its generator is polynomial if and only if the drift and diffusion coefficients are polynomial of first and second degree, respectively; see [8] and [21]. The following result is the generalization of this fact to the probability-valued setting. The proof is given in Section A.

Theorem 4.3. *Let $L\colon P^D \to P$ be a linear operator. Then L is $M_1(E)$-polynomial and its carré-du-champ operator Γ is an $M_1(E)$-derivation if and only if*

$$Lp(\nu) = \big\langle B(\partial p(\nu)), \nu \big\rangle + \frac{1}{2}\big\langle Q(\partial^2 p(\nu)), \nu^2 \big\rangle, \quad \nu \in M_1(E), \tag{4.2}$$

for some linear operators $B\colon D \to C_\Delta(E)$ and $Q\colon D \otimes D \to \widehat{C}_\Delta(E^2)$.

In this case, B and Q are uniquely determined by L.

An analogue of Theorem 4.3 holds for L being $\mathcal{S}$-polynomial, where $\mathcal{S}$ is an arbitrary subset of $M(E)$; see Theorem A.1.

Example 4.4 (The Fleming-Viot generator)**.** Let $E = \mathbb{R}$ and $D = C_\Delta^2(\mathbb{R})$. The Fleming–Viot diffusion was introduced in [23] and subsequently studied by several other authors. This process takes values in $M_1(\mathbb{R})$, and its generator L acts on polynomials $p \in P^D$ by

$$Lp(\nu) = \int_E B(\partial p(\nu))(x)\nu(dx) + \frac{1}{2}\int_{E^2} \partial_{xy}^2 p(\nu)\nu(dx)(\delta_x(dy) - \nu(dy)),$$

where $Bg := \frac{1}{2}\sigma^2 g''$ for some $\sigma \in \mathbb{R}$. This is an $M_1(\mathbb{R})$-polynomial operator of the form (4.2), where $Q = \Psi$ as defined in (3.1). For more details, see Chapter 10.4 of [17].

Corollary 2.5 states that any polynomial on $M_1(E)$ has a unique homogeneous representative. Therefore, an operator L satisfying (4.2) actually maps any monomial $\langle g, \nu^k \rangle$ to a unique monomial $\langle h, \nu^k \rangle$ on $M_1(E)$. This induces an operator L_k acting on the corresponding coefficients by $L_k g := h$. The operators $L_1, L_2, \ldots$ are the key objects needed to compute conditional moments of polynomial diffusions corresponding to L.

Definition 4.5. *Let $L\colon P^D \to P$ satisfy (4.2). The k-th dual operator of L is defined as the unique linear operator $L_k\colon D^{\otimes k} \to \widehat{C}_\Delta(E^k)$ determined by*

$$Lp(\nu) = \langle L_k g, \nu^k \rangle, \qquad \nu \in M_1(E), \tag{4.3}$$

for every $p(\nu) = \langle g, \nu^k \rangle$ with $g \in D^{\otimes k}$.

Because of (4.2), the k-th dual operator L_k can be written

$$L_k = kB \otimes \mathrm{id}^{\otimes(k-1)} + \frac{k(k-1)}{2}Q \otimes \mathrm{id}^{\otimes(k-2)}, \tag{4.4}$$

where the tensor notation $B_1 \otimes \ldots \otimes B_N$ is used to denote the linear operator from $D^{\otimes k}$ to $\widehat{C}_\Delta(E^k)$ determined by $(B_1 \otimes \ldots \otimes B_N)(g^{\otimes k}) := B_1(g^{\otimes n_1}) \otimes \ldots \otimes B_N(g^{\otimes n_N})$ for given

linear operators $B_i \colon D^{\otimes n_i} \to \widehat{C}_\Delta(E^{n_i})$ with $n_1 + \cdots + n_N = k$. More explicitly, we have

$$L_k = B_k + Q_k$$

where B_k and Q_k are defined by

$$B_k g := \sum_{i=1}^{k} B^{(i)} g \quad \text{and} \quad Q_k g := \frac{1}{2} \sum_{i,j=1}^{k} Q^{(ij)} g \tag{4.5}$$

for $B^{(i)} g(x) := Bg(\ldots, x_{i-1}, \cdot, x_{i+1}, \ldots)(x_i)$ and

$$Q^{(ij)} g(x) := Q\big(g(\ldots, x_{i-1}, \cdot, x_{i+1}, \ldots, x_{j-1}, \cdot, x_{j+1}, \ldots)\big)(x_i, x_j).$$

Remark 4.6. Observe that without the existence of an homogeneous representative (guaranteed by Corollary 2.5), expression (4.3) would read

$$Lp(\nu) = \langle L_k^k g, \nu^k \rangle + \langle L_k^{k-1} g, \nu^{k-1} \rangle + \cdots + L_k^0 g, \qquad \nu \in M_1(E),$$

and the k-th dual operator would thus consist in a $(k+1)$-tuple of operators $L_k^k, \ldots, L_k^0$. In the context of the moment formula, as stated in Theorem 5.3 below, the PIDE of (5.2) would then translate to a system of $(k+1)$ PIDEs. If one is interested in studying jump-diffusions taking value in other subspaces of $M(E)$, as e.g. $M_+(E)$, a homogeneous representative can no longer be found and one has to deal with systems of PIDEs to compute the moments.

5 Existence and uniqueness of polynomial diffusions on $M_1(E)$

Let E be a locally compact Polish space, D a dense linear subspace of $C_\Delta(E)$ containing the constant function 1, and $L \colon P^D \to P$ a linear operator. In this section we study existence and uniqueness of $M_1(E)$-valued polynomial diffusions, and derive the moment formula.

An $M_1(E)$-valued process X with càdlàg paths defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ is called a *solution to the martingale problem for L* with initial condition $\nu \in M_1(E)$ if $X_0 = \nu$ $\mathbb{P}$-a.s. and

$$N_t^p = p(X_t) - p(X_0) - \int_0^t Lp(X_s) ds \tag{5.1}$$

defines a martingale for every $p \in P^D$. Uniqueness of solutions to the martingale problem is always understood in the sense of law. The martingale problem for L is *well–posed* if for every $\nu \in M_1(E)$ there exists a unique $M_1(E)$-valued solution to the martingale problem for L with initial condition ν. We are interested in solutions with continuous paths (with respect to the topology of weak convergence) corresponding to polynomial operators.

Definition 5.1. *Let L be $M_1(E)$-polynomial. Any continuous solution to the martingale problem for L is called a* probability-valued polynomial diffusion.

The following lemma relates path continuity of solutions to the martingale problem to the carré-du-champ operator being a derivation. This explains why we consider derivations in Theorem 4.3.

Lemma 5.2. *If the carré-du-champ operator Γ associated to L is an $M_1(E)$-derivation, then any solution to the martingale problem for L has continuous paths. Conversely, if for every initial condition $\nu \in M_1(E)$ there is a solution to the martingale problem for L with continuous paths, then the carré-du-champ operator Γ associated to L is an $M_1(E)$-derivation.*

Proof. Let X be a solution to the martingale problem for L. By Proposition 2 in [2], the real-valued process $p(X)$ is continuous for every $p \in P^D$, in particular for every linear monomial $p(\nu) = \langle h, \nu \rangle$ with $h \in D$. Since D is dense in $C_\Delta(E)$, we can conclude that X is continuous with respect to the topology of weak convergence on $M_1(E)$.

Conversely, if X is a solution to the martingale problem for L with continuous paths, then, by Lemma 2.3(i), the map $t \mapsto p(X_t)$ is continuous for all $p \in P^D$. The result now follows by Proposition 1 in [2]. $\square$

5.1 Moment formula and uniqueness in law

Polynomial diffusions are of interest in applications because they generally satisfy a *moment formula*, which allows moments of the process to be computed tractably. If E is a finite set, the moment formula always holds, but technical conditions, in particular on the dual operators, are needed in the general case. For details regarding operators and semigroups, we refer e.g. to [17].

Theorem 5.3. *Suppose L satisfies (4.2) and fix $k \in \mathbb{N}$. Assume that the k-th dual operator L_k is closable, and let g be in the domain of its closure $\overline{L}_k$. Suppose that there is a solution $u \colon \mathbb{R}_+ \times E^k \to \mathbb{R}$ of*

$$\frac{\partial u}{\partial t}(t, x) = \overline{L}_k u(t, \cdot)(x), \qquad (t, x) \in \mathbb{R}_+ \times E^k, \tag{5.2}$$
$$u(0, x) = g(x), \qquad x \in E^k,$$

and suppose that $\sup_{t \in [0,T]} \|\overline{L}_k u(t, \cdot)\| < \infty$ for all $T \in \mathbb{R}_+$. In particular, $u(t, \cdot)$ is assumed to be in the domain of $\overline{L}_k$ for all $t \geq 0$. Then for any continuous solution X to the martingale problem for L, one has the moment formula

$$\mathbb{E}\left[\langle g, X_T^k \rangle \mid \mathcal{F}_t\right] = \langle u(T - t, \cdot), X_t^k \rangle. \tag{5.3}$$

Proof. We will follow the proof of Theorem 4.4.11 in [17] and extend it to obtain also the formula for the *conditional* moments. Fix $T \in \mathbb{R}_+$, $t \in [0, T]$, and $A \in \mathcal{F}_t$. Define for all $(s_1, s_2) \in [0, T - t] \times [0, T - t]$ define $f(s_1, s_2) := \mathbb{E}[\langle u(s_1, \cdot), X_{t+s_2}^k \rangle 1_A]$. Fix $s_2 \in [0, T - t]$. Equation (5.2) and the fundamental theorem of calculus then yield

$$f(s_1, s_2) - f(0, s_2) = \mathbb{E}[\langle u(s_1, \cdot) - u(0, \cdot), X_{t+s_2}^k \rangle 1_A] = \int_0^{s_1} \mathbb{E}[\langle \overline{L}_k u(s, \cdot), X_{t+s_2}^k \rangle 1_A] ds.$$

Fix then $s_1 \in [0, T - t]$. Since $u(t, \cdot)$ is in the domain of $\overline{L}_k$ for all $t \in \mathbb{R}_+$, (5.1) yields

$$f(s_1, s_2) - f(s_1, 0) = \mathbb{E}[\mathbb{E}[\langle u(s_1, \cdot), X_{t+s_2}^k \rangle - \langle u(s_1, \cdot), X_t^k \rangle | \mathcal{F}_t] 1_A]$$
$$= \int_0^{s_2} \mathbb{E}[\langle \overline{L}_k u(s_1, \cdot), X_{t+s}^k \rangle 1_A] ds.$$

Since $\sup_{s_1, s_2 \in [0, T-t]} \left| \mathbb{E}[\langle \overline{L}_k u(s_1, \cdot), X_{t+s_2}^k \rangle 1_A] \right| \leq \sup_{s_1 \in [0, T]} \|\overline{L}_k u(s_1, \cdot)\| < \infty$, we can then conclude that both $f(\cdot, s_2)$ and $f(s_1, \cdot)$ are absolutely continuous with bounded derivatives. Lemma 4.4.10 in [17] then yields $f(T - t, 0) - f(0, T - t) = 0$, and the result follows. $\square$

In order to avoid confusion, for the rest of the section we denote by u_g the solution of (5.2) with initial condition $u_g(0, \cdot) = g$.

In most of the cases of interest (see Remark 5.7(iii) below) the operator L_k satisfies the positive maximum principle on E^k, for each $k \in \mathbb{N}$. If this is the case, the existence of a solution u_g of (5.2) satisfying the conditions of Theorem 5.3 for sufficiently many g, is essentially equivalent to the fact that $\overline{L}_k$ generates a strongly continuous positve contraction semigroup on $\widehat{C}_\Delta(E^k)$ or in other words that it is the generator of a Feller process on E^k. We state this precisely in the following remark.

Remark 5.4. Let L satisfy (4.2) and let X denote a solution to the corresponding martingale problem with initial condition $X_0 = \nu \in M_1(E)$. Assume that the corresponding k-th dual operator L_k satisfies the positive maximum principle on $(E^\Delta)^k$ (which in particular implies that L_k is closable), for each $k \in \mathbb{N}$.

Let D_0 be a dense subset of the domain of $\overline{L}_k$ and suppose that the conditions of Theorem 5.3 hold true for all $g \in D_0$. By Proposition 1.3.4 of [17], if we additionally have that $t \mapsto \overline{L}_k u_g(t, \cdot)$ is continuous, then $\overline{L}_k$ is the generator of a strongly continuous contraction semigroup $\{Y_t^k\}_{t \geq 0}$ on $\widehat{C}_\Delta(E^k)$ and $Y_t^k g = u_g(t, \cdot)$. In this case the moment formula reads as

$$\mathbb{E}\big[\langle g, X_T^k \rangle \mid \mathcal{F}_t\big] = \langle Y_{T-t}^k g, X_t^k \rangle, \quad \text{for all } g \in \widehat{C}_\Delta(E^k).$$

Conversely, if $\overline{L}_k$ is the generator of a strongly continuous contraction semigroup $\{Y_t^k\}_{t \geq 0}$ on $\widehat{C}_\Delta(E^k)$, then for all g in the domain of $\overline{L}_k$ the map $u_g(t, x) := Y_t^k g(x)$ satisfies the conditions of Theorem 5.3. By the Hille–Yosida theorem, this is for instance the case if the range of $\lambda - L_k$ is dense in $\widehat{C}_\Delta(E^k)$ for some $\lambda > 0$. In this case, Corollary 4.2.8 in [17] yields a solution $Z^{(k)}$ (without loss of generality defined on the same probability space as X) to the martingale problem for L_k with values in $(E^\Delta)^k$ and satisfying $Y_t^k g(x) = \mathbb{E}[g(Z_t^{(k)})|Z_0^{(k)} = x]$. The moment formula then yields

$$\mathbb{E}[g(Z_t^{(k)})|Z_0^{(k)} \sim \nu^k] = \mathbb{E}[\langle g, X_t^k \rangle]. \tag{5.4}$$

This gives an alternative interpretation to (5.3), namely that the PIDE in (5.2) is the Feynman-Kac PIDE associated to the k-dimensional process Markov process $Z^{(k)}$. Note that in the case of a finite state space E, (5.2) reduces to an ODE and $\overline{L}_k$ is automatically the generator of k-dimensional Markov chain.

As in the finite-dimensional case, the moment formula gives well-posedness of the martingale problem.

Corollary 5.5. *Suppose L satisfies (4.2), and let X be a continuous solution to the martingale problem for L with initial condition $\nu \in M_1(E)$. If the moment formula (5.3) holds for all $g \in D^{\otimes k}$ and $k \in \mathbb{N}$, then the law of X is uniquely determined by L and ν.*

Proof. By the moment formula (5.3) we have $\mathbb{E}[\langle g, X_T^k \rangle] = \langle u_g(T, \cdot), \nu^k \rangle$ for all $k \in \mathbb{N}$ and $g \in D^{\otimes k}$. Since $g \mapsto u_g$ is determined by L, Lemma 2.7 yields that the one-dimensional distributions of X are uniquely determined by L and ν. The conclusion follows by Theorem 4.4.2 in [17]. $\square$

5.2 Existence and well-posedness

Our first main result of this section gives abstract sufficient conditions for existence of solutions to the martingale problem. Applications of this result are discussed in Section 6. Recall that E is throughout a locally compact Polish space.

Theorem 5.6. *Let $D \subseteq C_\Delta(E)$ be a dense linear subspace containing the constant function 1. Let $L \colon P^D \to P$ be a linear operator satisfying (4.2), where*

(i) *B is E-conservative and satisfies $B1 = 0$,*

(ii) *Q is given by*

$$Q(g) = \alpha \Psi(g) + \sum_{i=1}^n (A_i \otimes A_i)(g), \qquad g \in D \otimes D,$$

where $\alpha \colon E^2 \to \mathbb{R}$ is a nonnegative symmetric function and, for $i = 1, \ldots, n$, A_i is the generator of a strongly continuous group of positive isometries of $C_\Delta(E)$, and the domain of A_i contains both D and $A_i(D)$,

(iii) $B - \frac{1}{2}\sum_{i=1}^{n} A_i^2$ *satisfies the positive maximum principle on E^Δ.*

Then L is $M_1(E)$-polynomial and its martingale problem has a solution with continuous paths for every initial condition $\nu \in M_1(E)$. If in addition the moment formula (5.3) holds for all $g \in D^{\otimes k}$ and $k \in \mathbb{N}$, then the martingale problem for L is well–posed.

Note that (4.2) imposes the implicit condition on α that $\alpha\Psi(g)$ must lie in $\widehat{C}_\Delta(E^2)$ for every $g \in D \otimes D$. If $D = C_\Delta(E)$, then α is necessarily bounded, as is seen from Theorem 5.9 below. However, this does not hold for general $D \subseteq C_\Delta(E)$, as one can see by considering $E = \mathbb{R}$, $D \subseteq C_\Delta^1(\mathbb{R})$, and $\alpha(x,y) = |x-y|^{-1}1_{\{x \neq y\}}$.

Proof. Theorem 4.3 shows that L is $M_1(E)$-polynomial. Lemma D.2 yields existence of a solution to the martingale problem for any initial condition (necessarily with continuous paths due to Lemma 5.2) once we check that L satisfies the positive maximum principle on $M_1(E^\Delta)$. Let therefore $\nu_* \in M_1(E^\Delta)$ be a maximizer of $p \in P^D$ over $M_1(E^\Delta)$. The optimality conditions in Theorem 3.1 yield

$$\partial_x p(\nu_*) = \sup_E \partial p(\nu_*) \quad \text{and} \quad \Psi\big(\partial^2 p(\nu_*)\big)(x,y) \leq 0, \quad x,y \in \operatorname{supp}(\nu_*).$$

Therefore, since $B - \frac{1}{2}\sum_{i=1}^{n} A_i^2$ satisfies the positive maximum principle and α is non-negative, we get

$$Lp(\nu_*) \leq \frac{1}{2}\sum_{i=1}^{n}\left(\langle A_i^2(\partial p(\nu_*)), \nu_*\rangle + \langle (A_i \otimes A_i)(\partial^2 p(\nu_*)), \nu_*^2\rangle\right).$$

The optimality condition in Theorem 3.4 now yields $Lp(\nu_*) \leq 0$. This proves the positive maximum principle and thus the existence statement. The assertions regarding the moment formula and well–posedness follow from Theorem 5.3 and Corollary 5.5. $\qquad\square$

Remark 5.7. (i) With regard to item (iii) in Theorem 5.6, note that a linear operator $\mathcal{G} : D \to C_\Delta(E)$ satisfies the positive maximum principle on E^Δ if and only if $\mathcal{G}$ satisfies the positive maximum principle on E and $\mathcal{G}g(\Delta) \geq 0$ for every nonnegative $g \in C_0(E) \cap D$. In many cases of interest, for instance $E \subseteq \mathbb{R}^d$ and $D \subseteq \mathbb{R} + C_c(E)$, the positive maximum principle on E implies the positive maximum principle on E^Δ.

(ii) Let us also remark that the k-th dual operator $\mathcal{G}_k$ of $\langle \mathcal{G}(\partial p(\nu)), \nu\rangle$ satisfies the positive maximum principle on $(E^\Delta)^k$ if it holds for $\mathcal{G}$ on E^Δ. Indeed, if $x^* \in (E^\Delta)^k$ is a maximum of g, then x_i^* is a maximum of $g(\ldots, x_{i-1}^*, \cdot, x_{i+1}^*, \ldots)$. Hence $\mathcal{G}_k$ given by

$$\mathcal{G}_k g = k\mathcal{G} \otimes \operatorname{id}^{\otimes(k-1)}g = \sum_{j=1}^{k}\mathcal{G}^{(j)}g,$$

where we use the same notation as in (4.5), clearly satisfies the positive maximum principle on $(E^\Delta)^k$.

(iii) Consider the setting and the assumptions of Theorem 5.6 and define

$$G_k := k\Big(B - \frac{1}{2}\sum_{i=1}^{n} A_i^2\Big) \otimes \operatorname{id}^{\otimes(k-1)}, \qquad C_k := \frac{k(k-1)}{2}(\alpha\Psi) \otimes \operatorname{id}^{\otimes(k-2)},$$

$$T_k := k\Big(\frac{1}{2}\sum_{i=1}^{n} A_i^2\Big) \otimes \operatorname{id}^{\otimes(k-1)} + \frac{k(k-1)}{2}\Big(\sum_{i=1}^{n}(A_i \otimes A_i)\Big) \otimes \operatorname{id}^{\otimes(k-2)}.$$

Note that by (4.4) we have $L_k = G_k + C_k + T_k$. We claim that G_k, C_k, T_k, and hence L_k, satisfy the positive maximum principle on $(E^\Delta)^k$.

By item (iii) in Theorem 5.6, $B - \frac{1}{2}\sum_{i=1}^n A_i^2$ satisfies the positive maximum principle on E^Δ, whence by (ii) it holds also for G_k on $(E^\Delta)^k$. The form of Ψ and the nonnegativity of α guarantee that this holds also for C_k. Finally, since $T_k = \sum_{i=1}^n \frac{1}{2}(\sum_{j=1}^n A_i^{(j)})^2$ where $A_i^{(j)}g(x) = A_i g(\dots, x_{j-1}, \cdot, x_{j+1}, \dots)(x_j)$, Remark 3.5 yields the positive maximum principle on $(E^\Delta)^k$ also for T_k and thus for L_k.

The following result gives a useful condition for uniqueness when all the operators A_i are zero. Due to Lemma 3.6 this happens, for instance, if $D = C_\Delta(E)$ and in particular if E consists of finitely many points. An example where uniqueness holds when those operators are not all zero is given in Example 6.7.

Lemma 5.8. *Consider setting and assumptions of Theorem 5.6, and assume that $A_i = 0$ for all i. Assume additionally that α is bounded and B is closable and its closure is the generator of a strongly continuous contraction semigroup on $C_\Delta(E)$. Then the moment formula (5.3) holds for all $g \in \widehat{C}_\Delta(E^k)$ and $k \in \mathbb{N}$.*

Since B satisfies the positive maximum principle on E^Δ due to Theorem 5.6(iii), the Hille–Yosida theorem guarantees that the conditions of the lemma are satisfied whenever $\lambda - B$ has dense range in $C_\Delta(E)$ for some $\lambda > 0$.

Proof. Let $\{Y_t^1\}_{t\geq 0}$ be the semigroup corresponding to $\overline{B}$. Fix any $k \in \mathbb{N}$ and let B_k and Q_k be as in (4.5). It is straightforward to check that B_k is the restriction to $D^{\otimes k}$ of the generator of the strongly continuous contraction semigroup $\{\overline{(Y_t^1)^{\otimes k}}\}_{t\geq 0}$ on $\widehat{C}_\Delta(E^k)$. Moreover, one has the estimate

$$\|Q_k g\| \leq k(k-1)\|\alpha\|\|g\|, \quad g \in \widehat{C}_\Delta(E^k),$$

whence Q_k is a bounded operator. It follows as in Theorem 1.7.1 and Corollary 1.7.2 in [17] that $L_k = B_k + Q_k$ is closable and its closure is the generator of a strongly continuous contraction semigroup on $\widehat{C}_\Delta(E^k)$. By Remark 5.4 and Theorem 5.3 the result follows. $\qquad\square$

While Theorem 5.6 only gives sufficient conditions for existence, the result is sharp. Indeed, we now show that if $D = C_\Delta(E)$, no other polynomial specifications exist. For instance, this is the case if E is a finite set. The following theorem, which is our second main result of this section, makes this precise. The proof is given in Section B.

Theorem 5.9. *Let $D = C_\Delta(E)$ and let $L\colon P^D \to P$ be a linear operator. Then L is $M_1(E)$-polynomial, its martingale problem is well posed, and all solutions have continuous paths, if and only if L satisfies (4.2) with*

$$Bg = \int (g(\xi) - g(\cdot))\,\nu_B(\cdot, d\xi) \quad \text{and} \quad Qg = \alpha\Psi(g), \qquad (5.5)$$

where ν_B is a nonnegative, finite kernel from E to E, and $\alpha\colon (E^\Delta)^2 \to \mathbb{R}$ is nonnegative, symmetric, bounded, and continuous on $(E^\Delta)^2 \setminus \{x = y\}$. In this case, for each $k \in \mathbb{N}$ the k-th dual operator L_k satisfies the hypothesis of Theorem 5.3, and the moment formula (5.3) holds for all $g \in \widehat{C}_\Delta(E^k)$. Moreover, B and Q, and hence each L_k, are bounded operators.

As in Theorem 5.6, condition (4.2) imposes implicit conditions on the different parameters. This is the case for the measure ν_B, which in particular needs to satisfy $\int (g(\xi) - g(\cdot))\nu_B(\cdot, d\xi) \in C_\Delta(E)$ for all $g \in C_\Delta(E)$. This condition is clearly satisfied if the map from E to $M_+(E)$ given by $x \mapsto \nu_B(x, \cdot)$ is continuous. However the converse fails to be true as one can see by considering the following kernel

$$\nu_B(x, d\xi) = \delta_{\phi(x)} 1_{\{\phi(x) \neq x\}},$$

for some continuous $\phi\colon E \to E$ such that $\phi \neq \mathrm{id}$.

Corollary 5.10. *Let $D \subseteq C_\Delta(E)$ be a dense linear subspace containing the constant function 1 and let L satisfy* (4.2) *with B and Q as in Theorem 5.9. Then L is $M_1(E)$-polynomial, its martingale problem is well–posed, and all solutions have continuous paths. Moreover, the moment formula* (5.3) *holds for all $g \in D^{\otimes k}$ and $k \in \mathbb{N}$.*

Proof. Since by Theorem 5.9 each L_k is bounded, the operator L can be uniquely extended to $P^{C_\Delta(E)}$. The result then follows by the same theorem. $\qquad\square$

The last main result of this section characterizes probability-valued polynomial martingales. An $M_1(E)$-valued process X is called a martingale if $\langle g, X \rangle$ is a martingale for every $g \in C_\Delta(E)$. Note that, unlike Theorem 5.6, the conditions are both necessary and sufficient, regardless of the choice of domain D.

Theorem 5.11. *Let $D \subseteq C_\Delta(E)$ be a dense linear subspace containing the constant function 1. Let $L: P^D \to P$ be a linear operator. Then L is $M_1(E)$-polynomial, its martingale problem has a solution for any initial condition, and every solution is a martingale with continuous paths, if and only if L satisfies* (4.2) *with*

$$B = 0 \quad and \quad Q = \alpha \Psi$$

for some nonnegative symmetric function $\alpha: E^2 \to \mathbb{R}$. In this case, if in addition α is bounded, the martingale problem is well–posed.

Proof. To prove the forward implication, first note that Lemma 5.2 and Theorem 4.3 imply that L satisfies (4.2). To see that $B = 0$, pick any $g \in D$ and $x \in E$, and let X be a solution to the martingale problem with initial condition δ_x. Since $\langle g, X \rangle$ is a martingale, we have $\langle Bg, X \rangle = 0$ and hence $Bg(x) = \langle Bg, X_0 \rangle = 0$. The form of Q will follow from Lemma C.3. To verify its hypotheses, fix $g \in D$ and $\nu \in M_1(E)$, and define $p \in P^D$ by $p(\mu) := -(\langle g, \nu \rangle - \langle g, \mu \rangle)^2$. Then $\partial^2 p(\nu) = -2g \otimes g$, $p \le 0$, and $p(\nu) = 0$, so the positive maximum principle yields

$$-\langle Q(g \otimes g), \nu^2 \rangle = Lp(\nu) \le 0.$$

Next, fix $g \in D$ and $\nu \in M_1(E)$ such that g is constant on the support of ν. Define $p \in P^D$ by $p(\mu) := \langle g, \mu \rangle^2 - \langle g^2, \mu \rangle$. Then, again, $\partial^2 p(\nu) = 2g \otimes g$, and Jensen's inequality yields $p \le 0$ and $p(\nu) = 0$. Consequently,

$$\langle Q(g \otimes g), \nu^2 \rangle = Lp(\nu) \le 0.$$

The form of Q thus follows from Lemma C.3.

To prove the reverse implication, observe that existence of solutions to the martingale problem, along with path continuity, follows from Corollary 5.10, as does well–posedness if in addition α is bounded. Since $B = 0$, it is clear that $\langle g, X \rangle$ is a martingale for every $g \in D$ and every solution X to the martingale problem. This implies that X is a martingale. $\qquad\square$

6 Examples and applications

6.1 Finite underlying space

Let $E = \{1, \ldots, d\}$. Then $C_\Delta(E) = C(E)$ is finite-dimensional, so any dense linear subspace must equal the whole space. We therefore take $D = C(E)$. In this setting, any $M_1(E)$-valued process X is of the form $X_t = \sum_{i=1}^d Z_t^i \delta_i$ for some Δ^d-valued process $Z = (Z^1, \ldots, Z^d)$. When X is a polynomial diffusion, Theorem 5.9 describes its generator L in terms of a kernel ν_B from E to E and a nonnegative symmetric function $\alpha: E^2 \to \mathbb{R}$.

As we now show, the process Z then also solves a martingale problem whose generator can be written down explicitly.

In view of Example 2.1, any polynomial f on Δ^d can be represented as

$$f(z) = p(z_1 \delta_1 + \cdots + z_d \delta_d)$$

for some $p \in P^D$. We may then define an operator A acting on such polynomials f by the formula

$$Af(z) := Lp(z_1 \delta_1 + \cdots + z_d \delta_d).$$

Since $f(Z) = p(X)$ and $Af(Z) = Lp(X)$, it is clear that Z is a solution to the martingale problem for A with polynomials f as test functions. Conversely, if a solution Z to this martingale problem is given, a solution to the martingale problem for L is obtained by setting $X := \sum_{i=1}^d Z^i \delta_i$.

Next, a computation shows that A has the form

$$
\begin{aligned}
Af(z) = {} & \sum_{i,j=1}^d \nu_B(i, \{j\}) z_i \left(\frac{\partial f}{\partial z_j}(z) - \frac{\partial f}{\partial z_i}(z) \right) \\
& + \frac{1}{2} \sum_{i,j=1}^d \alpha(i,j) z_i z_j \left(\frac{\partial^2 f}{\partial z_i^2}(z) + \frac{\partial^2 f}{\partial z_j^2}(z) - 2 \frac{\partial^2 f}{\partial z_i \partial z_j}(z) \right).
\end{aligned}
\tag{6.1}
$$

This can alternatively be written $Af(z) = b(z)^\top \nabla f(z) + \frac{1}{2} \operatorname{Tr}\left(a(x) \nabla^2 f(x) \right)$, where the coefficients b and a are given by

$$
\begin{aligned}
b_k(z) &:= \sum_{i=1}^d \left(\nu_B(i, \{k\}) z_i - \nu_B(k, \{i\}) z_k \right), & k &= 1, \ldots, d, \\
a_{k\ell}(z) &:= -\frac{1}{2} \alpha(k, \ell) z_k z_\ell, & k, \ell &= 1, \ldots, d, \ k \neq \ell,
\end{aligned}
$$

and $a_{kk}(z) = -\sum_{\ell \neq k} a_{k\ell}(z)$. Here well-posedness was obtained in [21], which we thus recover as a special case. In particular, Z is a polynomial diffusion on Δ^d in the sense of [21, Definition 2.1]. Furthermore, Theorem 5.9 yields the moment formula for X, which reduces to the corresponding formula for Z given in [21, Theorem 3.1].

6.2 Underlying space $E \subseteq \mathbb{R}^d$

Let $E \subseteq \mathbb{R}^d$ be a closed subset and set

$$D := \{ f|_E \colon f \in \mathbb{R} + C_c^\infty(\mathbb{R}^d) \}.$$

Our goal is to analyze Theorem 5.6 in this setting. If E is not all of $\mathbb{R}^d$, the dynamics of the spatial motion is restricted. Intuitively, its diffusion component must be tangential to the boundary of E. This is encoded as follows.

$$
\Sigma^d(E) := \left\{ \tau \in C^1_\Delta(\mathbb{R}^d, \mathbb{R}^{d \times d}) \colon
\begin{array}{l}
g \in D, \ x \in E, \ g(x) = \max_E g \\
\text{imply } \tau(x)^\top \nabla g(x) = 0
\end{array}
\right\}.
\tag{6.2}
$$

Here $C^1_\Delta(\mathbb{R}^d, \mathbb{R}^{d \times d})$ consists of the matrix-valued functions with components in $C^1_\Delta(\mathbb{R}^d) = C_\Delta(\mathbb{R}^d) \cap C^1(\mathbb{R}^d)$.

Lemma 6.1. *Fix $\tau \in \Sigma^d(E)$ with columns $\tau_1, \ldots, \tau_d$. The operators $A_i \colon D \to C_\Delta(E)$ given by*

$$A_i g := \tau_i^\top \nabla g, \quad g \in D, \tag{6.3}$$

satisfy the conditions of Theorem 3.4. That is, each A_i is the generator of a strongly continuous group of positive isometries of $C_\Delta(E)$, and its domain contains both D and $A_i(D)$.

Note that A_i is well-defined by (6.3) in the sense that $A_i g$ only depends on g through its values on E. This is a direct consequence of the definition (6.2) of $\Sigma^d(E)$.

Proof. By Proposition 2.5 in [10], for each $i = 1, \ldots, d$, there exists a map $(t, x) \mapsto \phi_i(t, x)$ from $\mathbb{R} \times E$ to E such that

$$\frac{\partial}{\partial t} \phi_i(t, x) = \tau_i(\phi_i(t, x)), \qquad \phi_i(0, x) = x,$$

and the flow property $\phi_i(s + t, x) = \phi_i(s, \phi_i(t, x))$ holds since $\tau_i \in C^1_\Delta(\mathbb{R}^d, \mathbb{R}^d)$. This implies that $T_{i,t} g(x) := g(\phi_i(t, x))$, $t \in \mathbb{R}$, defines a strongly continuous group of positive isometries of $C_\Delta(E)$ with generator A_i. It is clear that the domain of A_i contains D, and it also contains $A_i(D)$ since the components of τ_i lie in $C^1_\Delta(\mathbb{R}^d)$. $\qquad\square$

Theorem 6.2. *Let $L \colon P^D \to P$ be a linear operator satisfying (4.2), where*

  (i) *B is E-conservative and $B1 = 0$,*

 (ii) *Q is given by*

$$Q(g \otimes g) = \alpha \Psi(g \otimes g) + \mathrm{Tr}\left((\tau^\top \nabla g) \otimes (\tau^\top \nabla g)^\top\right) \qquad g \in D,$$

   where $\tau \in \Sigma^d(E)$ and $\alpha \colon E^2 \to \mathbb{R}$ is a nonnegative symmetric function,

 (iii) *$B - \sum_{i=1}^d (\tau_i^\top \nabla)^2$ satisfies the positive maximum principle on E, where $\tau_1, \ldots, \tau_d$ are the columns of τ.*

Then conditions (i)–(iii) of Theorem 5.6 hold.

Proof. This follows directly from Lemma 6.1, up to the fact that in (iii) we need to verify that the positive maximum principle holds on E^Δ, not just on E. Since $D \subseteq C_c(E)$, this follows from Remark 5.7(i). $\qquad\square$

The rest of the section is devoted to the case $d = 1$ and $E = \mathbb{R}$. In view of Lemma C.1, the operator B should satisfy the positive maximum principle on $E = \mathbb{R}$. It is well-known, see e.g. [6] or [24], that under this condition B is a Lévy type operator, i.e.

$$Bg = bg' + \frac{1}{2} ag'' + \int \left(g(\cdot + \xi) - g - \chi(\xi)g'\right) F(\cdot, d\xi), \qquad g \in D, \tag{6.4}$$

for some continuous functions a, b with $a \geq 0$, a truncation function χ, and a kernel $F(\cdot, d\xi)$ from $\mathbb{R}$ to $\mathbb{R}$ such that $\int |\xi|^2 \wedge 1\, F(\cdot, d\xi) < \infty$. Every operator of this form satisfies $B1 = 0$ and the positive maximum principle on $\mathbb{R}$. The following result expresses Theorem 6.2 in this setting.

Corollary 6.3. *Let $L \colon P^D \to P$ be a linear operator satisfying (4.2), where B is given by (6.4) with $a := \sigma^2 + \tau^2$ for some continuous functions σ and τ, and Q is given by*

$$Q(g \otimes g)(x, y) = \frac{1}{2} \alpha(x, y)(g(x) - g(y))^2 + \tau(x)\tau(y)g'(x)g'(y), \qquad g \in D,$$

where $\alpha \in \widehat{C}_\Delta(\mathbb{R}^2)$ is nonnegative and $\tau \in C^1_\Delta(\mathbb{R})$. Assume also that B is $\mathbb{R}$-conservative. Then conditions (i)–(iii) of Theorem 5.6 hold true.

The coefficient α quantifies the diffusive exchange of mass between different points in the support of $X_t(dx)$. This is perhaps most clearly seen when $E = \{1, \ldots, d\}$; see Section 6.1. The role of τ is different, as it governs random fluctuations of the support of $X_t(dx)$. The following example illustrates this point.

Example 6.4. Consider an operator L of the form given in Corollary 6.3 with $\alpha = 0$, $Bg = \frac{1}{2}g''$, and $\tau = 1$ (hence $\sigma = 0$). The resulting operator Q is given by $Q(g \otimes g) = g' \otimes g'$. A solution to the martingale problem for L is given by $X = \delta_W$, where W is a Brownian motion. Indeed, applying Itô's formula to $\langle g, X_t \rangle^k = g(W_t)^k$ for any $g \in D$ and $k \in \mathbb{N}_0$ establishes that (5.1) is a martingale for any $p \in P^D$.

In this example, as well as in Corollary 6.3, a nonzero τ in the specification of Q is coupled with a corresponding diffusive component in the specification (6.4) of B. The following result shows that this is a general phenomenon.

Proposition 6.5. *Let $L \colon P^D \to P$ be a linear operator satisfying* (4.2) *with B given by* (6.4). *Suppose that L satisfies the positive maximum principle on $\mathbb{R}$. If $a = 0$, then $Q = \alpha \Psi$ for some nonnegative symmetric function $\alpha \colon \mathbb{R}^2 \to \mathbb{R}$.*

Proof. Lemma 6.6 below with $\lambda = 0, 1, 1/2$, along with Lemma C.1, imply that the conditions of Lemma C.3(i) are satisfied. The result follows. □

The next lemma constitutes the main tool to prove Proposition 6.5. But it also has other consequences. In particular, it implies that $Q(g \otimes g)(x, y)$ depends on g just through $g(x), g(y), g'(x)$, and $g'(y)$, provided that L satisfies the positive maximum principle on $M_1(E)$. This illustrates that the form of Q as given in Theorem 6.2 is very general.

Lemma 6.6. *Let $L \colon P^D \to P$ be a linear operator satisfying* (4.2) *with B given by* (6.4). *Suppose that L satisfies the positive maximum principle on $\mathbb{R}$. Then, for all $\lambda \in [0, 1]$, $g \in D$, and $x, y \in \mathbb{R}$ such that $g(x) = g(y)$, we have that*

$$\langle Q(g \otimes g), \nu_\lambda^2 \rangle \le \langle (ag')^2, \nu_\lambda \rangle, \qquad \nu_\lambda = \lambda \delta_x + (1 - \lambda) \delta_y.$$

Proof. Fix $g \in D$ such that $g(x) = g(y)$. Since, by Lemma C.1, $B1 = 0$ and $Q(g \otimes 1) = 0$ it is enough to consider the case $g(x) = g(y) = 1$. The result will follow from Lemma C.4. Indeed, if we let $(p_n)_{n \in \mathbb{N}}$ and $(f_n)_{n \in \mathbb{N}}$ be the sequences described there, by the positive maximum principle of L on $\mathbb{R}$ we get

$$0 \ge Lp_n(\nu_\lambda) = \langle Bf_n, \nu_\lambda \rangle + \frac{1}{2} \langle Q(g \otimes g), \nu_\lambda^2 \rangle$$

and letting n go to ∞ we can conclude the proof.

To verify the hypotheses of Lemma C.4, observe that Lemma C.1 yields

$$\langle Q(g \otimes g), \nu_\lambda^2 \rangle \ge 0 \qquad \text{for all } \lambda \in [0, 1].$$

Fix some $g \in D$ and $x, y \in \mathbb{R}$ such that $g(z) = g'(z) = 0$ for $z \in \{x, y\}$, and suppose that $\|g\| = 1$. Let $F_n : [0, 1] \to \mathbb{R}$ be the function defined in Lemma B.1. Consider then the sequence of polynomials given by

$$p_n(\nu) = \langle g, \nu \rangle^2 F_n\big(\langle H, \nu \rangle\big) - \frac{1}{n} \langle H, \nu \rangle,$$

where, for some compactly supported function $\rho \in C_\Delta^\infty(\mathbb{R})$ such that $\rho = 1$ on some neighborhood of x and y and $\rho(\mathbb{R}) \subseteq [0, 1]$,

$$H(z) = C|z - x|^2 |z - y|^2 \rho(z) + (1 - \rho(z)).$$

Observe that the conditions on g guarantee that for C big enough $|g| \le H$ and thus $|\langle g, \nu \rangle| \le \langle H, \nu \rangle$ for all $\nu \in M_1(\mathbb{R})$. For $\mathrm{supp}(\rho)$ small enough we also have that $\|H\| \le 1$. Lemma B.1 then yields $\langle g, \nu \rangle^2 F_n\big(\langle H, \nu \rangle\big) \le \frac{1}{n} \langle H, \nu \rangle$ for all $\nu \in M_1(\mathbb{R})$, and therefore $p_n \le 0$ on $M_1(\mathbb{R})$. This automatically implies that p_n has a maximum at ν_λ for all $\lambda \in [0, 1]$. Proceeding as in the proof of Theorem 5.9 we then obtain that $\langle Q(g \otimes g), \nu_\lambda^2 \rangle = 0$ for any $g \in D$ such that $g(x) = g(y) = 1$ and $g'(x) = g'(y) = 0$. Choosing $\lambda = 0, 1, 1/2$ we get the result. □

The following example gives a simple condition for well-posedness. We let G_k, C_k, and T_k be as in Remark 5.7(iii).

Example 6.7. Consider the setting of Corollary 6.3. Suppose that σ^2 is bounded away from zero, let the jump kernel $F(\,\cdot\,, d\xi)$ in (6.4) be zero, and assume that the parameters b and σ^2 are Lipschitz continuous and bounded. Then, by Theorem 8.1.6 of [17], B is $\mathbb{R}$-conservative and the closure of $G_k + T_k$ generates a strongly continuous semigroup on $\widehat{C}_\Delta(E^k)$ for each $k \in \mathbb{N}$. Since C_k is bounded, $\overline{L}_k$ generates a strongly continuous contraction semigroup on $\widehat{C}_\Delta(E^k)$ as well (see e.g. Theorem 1.7.1 in [17] for more details). Since Remark 5.7(iii) shows that L_k satisfies the positive maximum principle, Remark 5.4 and Theorem 5.3 yield the moment formula for all $g \in D^{\otimes k}$. Well-posedness thus follows from Theorem 5.6.

6.3 Conditional laws of jump-diffusions are polynomial

In this section we deal with particle systems driven by some idiosyncratic noise (Brownian motion and jumps) and one common Brownian motion. We show that for essentially *all* such jump diffusions the conditional law with respect to the common Brownian motion is polynomial.

Throughout $E = \mathbb{R}$ and $D \subseteq \mathbb{R} + C_c^\infty(\mathbb{R})$. Let b, σ, τ and $F(\,\cdot\,, d\xi)$ be as in Corollary 6.3 with the additional integrability conditon $\int |\xi|^2 \wedge |\xi| F(\,\cdot\,, d\xi) < \infty$. For these parameters and $\alpha = 0$ we define L to be the corresponding polynomial operator as of Corollary 6.3.

Moreover, let $(Z^i)_{i \in \mathbb{N}}$ be a weak solution of the system

$$dZ_t^i = b(Z_t^i)dt + \sigma(Z_t^i)dW_t^i + \tau(Z_t^i)dW_t^0 + \int \xi\big(\mathfrak{p}^i(dt, d\xi) - F(Z_t^i, d\xi)dt\big), \qquad (6.5)$$

with $Z_0^i = x \in \mathbb{R}$, where W^0 is a Brownian motion and $(W^1, \mathfrak{p}^1), (W^2, \mathfrak{p}^2), \ldots$ is a sequence of couples of Brownian motions and random measures with compensator $F(\,\cdot\,, d\xi)$. We assume that each couple is independent of the other couples and of W^0. Note that the generator of each Z^i is given by B as defined in (6.4).

Assume now that $Z^1, Z^2, \ldots$ are exchangeable and set

$$X_t = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{Z_t^i}.$$

By De Finetti's theorem (see e.g. Theorem 4.1 in [27] or, for a general overview, also Section 12.3 in [26]) we get that $(Z_t^i)_{i \in \mathbb{N}}$ are conditionally i.i.d. with respect to the invariant σ-algebra $\mathcal{F}_t^\infty = \sigma(X_s, s \le t)$ and that X can be expressed as

$$X_t = \mathbb{P}(Z_t^1 \in \,\cdot\,|\mathcal{F}_t^\infty). \qquad (6.6)$$

This implies in particular that for all $g \in D^{\otimes k}$ and $k \in \mathbb{N}$ it holds

$$\langle g, X_t^k \rangle = \mathbb{E}[g(Z_t^1, \ldots, Z_t^k)|\mathcal{F}_t^\infty]. \qquad (6.7)$$

Note that (see e.g. Theorem 2.3 in [28]) that under the additional assumption of pathwise uniqueness for the solution of (6.5), we get that

$$X_t = \mathbb{P}(Z_t^1 \in \,\cdot\,|\mathcal{F}_t^0), \qquad \text{where} \qquad \mathcal{F}_t^0 = \sigma(W_s^0, s \le t),$$

since $\mathcal{F}_t^0 = \mathcal{F}_t^\infty$ in this case.

In the following proposition we now show that X is polynomial by proving that it solves the martingale problem for the polynomial operator L specified above.

Proposition 6.8. *Let X be given by* (6.6). *Then X solves the martingale problem for L with initial condition δ_x.*

Proof. Let $g \in D^{\otimes k}$ and set $Z := (Z^1, \ldots, Z^k)$. Then we get that

$$N_t^{g,k} := g(Z_t) - g(x, \ldots, x) - \int_0^t L_k g(Z_s) ds$$

is a bounded $(\mathcal{F}_t)_{t \geq 0}$-martingale, where, in accordance with (4.4),

$$L_k = kB \otimes \mathrm{id}^{\otimes(k-1)} + \frac{k(k-1)}{2} \Sigma^\tau \otimes \Sigma^\tau \otimes \mathrm{id}^{\otimes(k-2)}$$

for $\Sigma^\tau g := \tau g'$. Since $\mathcal{F}_t^\infty \subseteq \mathcal{F}_t$ this implies that $\mathbb{E}[N_t^{g,k} | \mathcal{F}_t^\infty]$ is an $(\mathcal{F}_t^\infty)_{t \geq 0}$-martingale and hence setting $p(\nu) := \langle g, \nu^k \rangle$ we can compute using (6.7)

$$\mathbb{E}[p(X_t) | \mathcal{F}_s^\infty] - p(X_s) = \mathbb{E}\left[\int_s^t L_k g(Z_u) du \Big| \mathcal{F}_s^\infty \right]$$

$$= \mathbb{E}\left[\int_s^t \mathbb{E}[L_k g(Z_u) | \mathcal{F}_u^\infty] du \Big| \mathcal{F}_s^\infty \right] = \mathbb{E}\left[\int_s^t Lp(X_u) du \Big| \mathcal{F}_s^\infty \right]$$

proving that X is a solution to the martingale problem for L. □

A Proof of Theorem 4.3 and a generalization

We first prove Theorem 4.3. Assume first L is of the stated form. Then for monomials $p(\nu) = \langle g, \nu \rangle^k$ with $g \in D$, $k \in \mathbb{N}$ and $\nu \in M_1(E)$ one has

$$Lp(\nu) = \langle B(\partial p(\nu)), \nu \rangle + \frac{1}{2} \langle Q(\partial^2 p(\nu)), \nu^2 \rangle$$

$$= k \langle g, \nu \rangle^{k-1} \langle Bg, \nu \rangle + \frac{1}{2} k(k-1) \langle g, \nu \rangle^{k-2} \langle Q(g \otimes g), \nu^2 \rangle,$$

which is a polynomial in ν of degree at most k. Moreover, $L1 = 0$. By linearity, this shows that L is $M_1(E)$-polynomial. Next, a direct calculation yields

$$\Gamma(p, q)(\nu) = \left\langle Q(\partial p(\nu) \otimes \partial q(\nu)), \nu^2 \right\rangle \qquad \text{for all } \nu \in M_1(E),$$

which is easily seen to be an $M_1(E)$-derivation due to the product rule give in Lemma 2.3(v).

Conversely, assume L is $M_1(E)$-polynomial and Γ is an $M_1(E)$-derivation. Consider arbitrary first degree monomials $q(\nu) = \langle g, \nu \rangle$ and $r(\nu) = \langle h, \nu \rangle$, $g, h \in D$. The $M_1(E)$-polynomial property and Corollary 2.5 yield

$$Lq(\nu) = \langle Bg, \nu \rangle \qquad \text{for all } \nu \in M_1(E),$$

for some map $B : D \to C_\Delta(E)$ that are easily seen to be linear due to the linearity of L. Furthermore, the $M_1(E)$-polynomial property, definition (4.1) of Γ, and Corollary 2.5 imply that

$$\Gamma(q, r)(\nu) = \langle Q(g \otimes h), \nu^2 \rangle \qquad \text{for all } \nu \in M_1(E),$$

where Q inherits symmetry and linearity from Γ and take values in $\widehat{C}_\Delta(E^2)$. Thus, by taking linear combinations, we can and do extend them to operators on $D \otimes D$.

Explicit calculation now shows that Lp is of the form (4.2) for $p = q$ and $p = q^2$. Furthermore, since Γ is an $M_1(E)$-derivation we have $\Gamma(1,1) = 2\Gamma(1,1)$, hence $\Gamma(1,1) = 0$, and therefore $L1 = L(1^2) = 0 + 2L1$. Thus $L1 = 0$, so that (4.2) holds also for $p = 1$.

We now make more substantial use of the fact that Γ is an $M_1(E)$-derivation in order to extend (4.2) to higher degree monomials. We proceed by induction on k, and assume

Lp is of the form (4.2) for all $p = q^l$, $l \leq k$. So far we have proved this for $k = 2$. The definition (4.1) of Γ and the fact that it is an $M_1(E)$-derivation give the identity on $M_1(E)$

$$L(q^{k+1}) = 2qL(q^k) - q^2 L(q^{k-1}) + q^{k-1}\Gamma(q, q)$$

for $k \geq 2$. Due to the induction assumption, the right-hand side can be computed explicitly using (4.2). The result is

$$(k+1)q(\nu)^k \langle Bg, \nu \rangle + \frac{1}{2}(k+1)kq(\nu)^{k-1} \langle Q(g \otimes g), \nu^2 \rangle,$$

which is equal to $\langle B(\partial p(\nu)), \nu \rangle + \frac{1}{2}\langle Q(\partial^2 p(\nu)), \nu^2 \rangle$ with $p = q^{k+1}$, for all $\nu \in M_1(E)$. This concludes the induction step. It follows by induction that (4.2) holds for all monomials $\langle g, \nu \rangle^k$, and by linearity for all $p \in P^D$. Finally, the uniqueness assertion is immediate from the way B and Q were obtained above. This completes the proof of Theorem 4.3. $\square$

We now state a generalization of Theorem 4.3, where $M_1(E)$ is replaced by a general state space. We let E be a locally compact Polish space, $D \subseteq C_\Delta(E)$ be a dense linear subspace, and fix $\mathcal{S} \subseteq M(E)$.

Theorem A.1. *Let $L : P^D \to P$ be a linear operator. Then L is $\mathcal{S}$-polynomial and its carré-du-champs operator Γ is an $M_1(E)$-derivation if and only if L admits a representation*

$$Lp(\nu) = B_0(\partial p(\nu)) + \left\langle B_1(\partial p(\nu)), \nu \right\rangle$$
$$+ \frac{1}{2}\left(Q_0(\partial^2 p(\nu)) + \left\langle Q_1(\partial^2 p(\nu)), \nu \right\rangle + \left\langle Q_2(\partial^2 p(\nu)), \nu^2 \right\rangle \right), \quad \nu \in \mathcal{S}$$

for some linear operators $B_0 : D \to \mathbb{R}$, $B_1 : D \to C_\Delta(E)$, $Q_0 : D \otimes D \to \mathbb{R}$, $Q_1 : D \otimes D \to C_\Delta(E)$, $Q_2 : D \otimes D \to \widehat{C}_\Delta(E^2)$. If $\mathcal{S}$ contains an open subset of $M(E)$, these operators are uniquely determined by L.

Proof. The proof of this result follows the proof of Theorem 4.3. $\square$

B Proof of Theorem 5.9

Assume L satisfies (4.2) with B and Q as in (5.5), where ν_B is a nonnegative, finite kernel from E to E, and $\alpha : (E^\Delta)^2 \to \mathbb{R}$ is nonnegative, symmetric, bounded, and continuous on $(E^\Delta)^2 \backslash \{x = y\}$. Clearly Q is bounded with operator norm $2\|\alpha\|$. Identifying $C_\Delta(E)$ and $C(E^\Delta)$, we infer from Lemma C.2 that B is bounded, satisfies $B1 = 0$ as well as the positive maximum principle on E^Δ, and that $\{e^{tB}\}_{t \geq 0}$ is a strongly continuous contraction semigroup. By considering any sequence of functions $g_n \in C_0(E)$ with $0 \leq g_n(x) \uparrow 1$ for all $x \in E$, and using that $\nu_B(x, \{\Delta\}) = 0$ for all $x \in E$, one sees that B is E-conservative. Theorem 5.6 then yields that L is $M_1(E)$-polynomial and its martingale problem has an solution with continuous paths for every initial condition $\nu \in M_1(E)$. Well–posedness follows by Lemma 5.8.

We now prove the opposite implication. Assume L is $M_1(E)$-polynomial, its martingale problem is well–posed, and all solutions have continuous paths. Theorem 4.3 and Lemma 5.2 imply that L satisfies (4.2), and then also the positive maximum principle on $M_1(E)$ due to Lemma D.1.

By Lemma C.1 B satisfies the positive maximum principle on E and Lemma C.2 thus shows that B has the form in (5.5) for some nonnegative, finite kernel ν_B from E^Δ to E^Δ. Additionally, B is bounded, satisfies the positive maximum principle on E^Δ, and is the generator of the strongly continuous contraction semigroup $\{e^{tB}\}_{t \geq 0}$. We must prove that $\nu_B(x, \{\Delta\}) = 0$ for all $x \in E$; this will allow us to view ν_B as a kernel from E to E.

Assume by contradiction that there is some $x \in E$ with $\nu_B(x, \{\Delta\}) > 0$. Let Z be the Markov process associated to the semigroup $\{e^{tB}\}_{t \geq 0}$. Then, by approximating $1_{\{\cdot \in \Delta\}}$ by a sequence of bounded continuous functions g^n and applying relation (5.4), we find

$$0 < \mathbb{P}[Z_t \in \Delta | Z_0 = x] = \mathbb{E}[1_{\{Z_t \in \Delta\}} | Z_0 = x] = \lim_{n \to \infty} \mathbb{E}[g^n(Z_t) | Z_0 = x]$$
$$= \lim_{n \to \infty} \mathbb{E}[\langle g^n(\cdot), X_t \rangle | X_0 = \delta_x] = \mathbb{E}[\langle 1_{\{\cdot \in \Delta\}}, X_t \rangle | X_0 = \delta_x]$$

for all $t \geq 0$. This contradicts the fact that X_t is $M_1(E)$-valued and proves that B is of the stated form.

The form of Q will follow from Lemma C.3. To verify its hypotheses, note that by Lemma C.1 $\langle Q(g \otimes g), \nu^2 \rangle \geq 0$. Next, fix some $g \in D$ and $\nu \in M_1(E)$ such that $g = 0$ on the support of ν, and suppose that $\|g\| = 1$. For each $n \in \mathbb{N}$, define the polynomial

$$p_n(\mu) = \langle g, \mu \rangle^2 F_n\left(\langle |g|, \mu \rangle\right) - \frac{1}{n}\langle |g|, \mu \rangle,$$

where F_n is as in Lemma B.1. Since $D = C_\Delta(E)$, we have $p_n \in P^D$. Moreover, since $F_n(z)zn \leq 1$ for all $z \in [0, 1]$, we get

$$\langle g, \mu \rangle^2 F_n\left(\langle |g|, \mu \rangle\right) \leq \frac{1}{n}\langle |g|, \mu \rangle, \quad \mu \in M_1(E),$$

and therefore $p_n \leq 0$ on $M_1(E)$. Since $g = 0$ on the support of ν, $p_n(\nu) = 0$. Applying the positive maximum principle and using the form (4.2) of L, as well as $\langle g, \nu \rangle = \langle |g|, \nu \rangle = 0$ and $F_n(0) = 1$ we obtain

$$0 \geq Lp_n(\nu) = -\frac{1}{n}\langle B(|g|), \nu \rangle + \langle Q(g \otimes g), \nu^2 \rangle$$

for all n, whence $\langle Q(g \otimes g), \nu^2 \rangle \leq 0$. By scaling, this actually holds for any $g \in D$ and $\nu \in M_1(E)$ such that $g = 0$ on the support of ν. If g equals some other constant $c \in \mathbb{R}$ on the support of ν, we still get

$$\langle Q(g \otimes g), \nu^2 \rangle = \langle Q((g - c) \otimes (g - c)), \nu^2 \rangle \leq 0$$

using that $Q(g \otimes 1) = 0$ by Lemma C.1. Thus Lemma C.3(ii) holds, and we conclude that $Q = \alpha \Psi$ for some nonnegative symmetric function $\alpha \colon E^2 \to \mathbb{R}$. It remains to use that $\alpha \Psi(g) \in \widehat{C}_\Delta(E^2)$ to show that this function can be extended to a bounded continuous function on $(E^\Delta)^2 \setminus \{x = y\}$.

Continuity is clear. For proving boundedness, choose a sequence of pairs $(x_n, y_n) \in (E^\Delta)^2 \setminus \{x = y\}$ such that $\alpha(x_n, y_n) \xrightarrow{n \to \infty} \infty$. Since we can assume without loss of generality that $\alpha(x_i, y_i) > 0$, $x_i \neq x_j$, $x_i \neq y_j$, and $y_i \neq y_j$ for all $i, j \in \mathbb{N}$, we can construct $g \in C_\Delta(E)$ such that

$$(g(x_n) - g(y_n))^4 = \alpha(x_n, y_n)^{-1}.$$

This yields $\alpha(x_n, y_n)\Psi(g \otimes g)(x_n, y_n) = \alpha(x_n, y_n)^{1/2}$ proving that $\alpha \Psi(g \otimes g)$ is unbounded and providing the necessary contradiction. $\square$

Lemma B.1. *Define $F_n(z) := \frac{n-1}{n}(1 - z)^n + \frac{1}{n}$ for all $z \in [0, 1]$. Then*

$$F_n(z) \in [0, 1], \qquad F_n(z)zn \leq 1, \qquad and \qquad F_n(z)\sqrt{zn} \leq 1,$$

for all $z \in [0, 1]$.

C Auxiliary lemmas

Let E be a locally compact Polish space.

Lemma C.1. *Let $D \subseteq C_\Delta(E)$ be a dense linear subspace containing the constant function 1, and let $L\colon P^D \to P$ be a linear operator satisfying (4.2) and the positive maximum principle on $M_1(E)$. Then B satisfies the positive maximum principle on E, $B1 = 0$, $\langle Q(g \otimes g), \nu^2 \rangle \geq 0$, and $Q(g \otimes 1) = 0$ for all $g \in D$ and $\nu \in M_1(E)$.*

Proof. By (4.2) we get $L1 = 0$. Note also that for any $g \in D$ and $x \in E$ such that $g(x) = \max_E g \geq 0$, the polynomial $p(\nu) = \langle g, \nu \rangle$ lies in P^D and satisfies $p(\delta_x) = \max_{M_1(E)} p \geq 0$. Thus $Bg(x) = Lp(\delta_x) \leq 0$. Furthermore, taking $p(\nu) = \langle 1, \nu \rangle$ we get $p \equiv 1$ on $M_1(E)$ and hence $B1(x) = Lp(\delta_x) = 0$ for all $x \in E$. Fix then g and ν as in the lemma and define $p \in P^D$ by $p(\mu) = -(\langle g, \nu \rangle - \langle g, \mu \rangle)^2$. Then $p \leq 0$, $p(\nu) = 0$, $\partial p(\nu) = 0$, and $\partial^2 p(\nu) = -2g \otimes g$, so the positive maximum principle yields $-\langle Q(g \otimes g), \nu^2 \rangle = Lp(\nu) \leq 0$. Furthermore, taking $p(\nu) = \langle g \otimes 1, \nu^2 \rangle - \langle g, \nu \rangle$ we get $p \equiv 0$ on $M_1(E)$ and hence $0 = \langle g, \nu \rangle \langle B1, \nu \rangle + \langle Q(g \otimes 1), \nu^2 \rangle = \langle Q(g \otimes 1), \nu^2 \rangle$ for all $\nu \in M_1(E)$, proving the claim. $\square$

Lemma C.2. *Let $B\colon C(E^\Delta) \to C(E^\Delta)$ be a linear operator. Then $B1 = 0$ and B satisfies the positive maximum principle on E if and only if there is a nonnegative, finite kernel ν_B from E to E^Δ such that*

$$Bg(x) = \int (g(\xi) - g(x))\nu_B(x, d\xi) \tag{C.1}$$

for all $x \in E$ and $g \in C(E^\Delta)$. In this case, B is bounded and satisfies the positive maximum principle on E^Δ, and $\{e^{tB}\}_{t \geq 0}$ is a strongly continuous contraction semigroup. Moreover, there is some nonnegative (finite) measure $\nu_B(\Delta, \cdot)$ such that (C.1) holds also for $x = \Delta$.

Proof. Assume there is a nonnegative, finite kernel ν_B from E to E^Δ such that (C.1) holds for all $x \in E$ and $g \in C(E^\Delta)$. Then clearly $B1 = 0$. Suppose $g \in C(E^\Delta)$, $x \in E$, and $g(x) = \max_E g \geq 0$. Then $g(x) = \max_{E^\Delta} g$, so that $g(\xi) - g(x) \leq 0$ for all $\xi \in E^\Delta$ and hence $Bg(x) \leq 0$. Thus B satisfies the positive maximum principle on E, which proves sufficiency.

To prove necessity, assume $B1 = 0$ and B satisfies the positive maximum principle on E. By Lemmas 4.2.1 and 1.2.11 in [17], the restriction $B|_{C_0(E)}$ is dissipative, hence closable, and even closed since it is globally defined on $C_0(E)$. By the closed graph theorem $B|_{C_0(E)}$ is bounded, and then so is B since $B1 = 0$. Pick any $g \in C(E^\Delta)$ with $g(\Delta) = \max_{E^\Delta} g \geq 0$. Then $g - g(\Delta) \leq 0$, so there exist functions $h_n \in C_c(E)$ with $h_n \leq 0$ and $h_n \to g - g(\Delta)$ uniformly. Then $Bh_n \to B(g - g(\Delta)) = Bg$ uniformly as well. Taking x_n such that $h_n(x_n) = 0$ and $x_n \to \Delta$, we obtain $Bg(\Delta) = \lim_{n \to \infty} Bh_n(x_n) \leq 0$. We have thus proved that B is bounded and satisfies the positive maximum principle on E^Δ. As a result, Lemma 4.2.1 and Theorem 1.7.1 in [17] yield that $\{e^{tB}\}_{t \geq 0}$ is a strongly continuous contraction semigroup.

It remains to exhibit a kernel ν_B from E^Δ to E^Δ such that (C.1) holds for all $x \in E^\Delta$ and $g \in C(E^\Delta)$. To this end, fix $x \in E^\Delta$ and define $h \in C(E^\Delta)$ by $h(y) := d(x, y)$, where $d(\cdot, \cdot)$ is a compatible metric for the Polish space E^Δ. Since B satisfies the positive maximum principle on E^Δ, the map

$$C(E^\Delta) \to \mathbb{R}, \qquad g \mapsto B(gh)(x)$$

is a positive linear functional. By the Riesz–Markov representation theorem, there is a measure $\mu(x, \cdot) \in M_+(E^\Delta)$ such that $B(gh)(x) = \int_{E^\Delta} g(\xi)\mu(x, d\xi)$ for all $g \in C(E^\Delta)$. We

define

$$\nu_B(x, d\xi) := 1_{E^\Delta \setminus \{x\}}(\xi) \frac{1}{h(\xi)} \mu(x, d\xi),$$

which is permissible since $h(y) > 0$ for all $y \neq x$. For every $g \in C_c(E^\Delta \setminus \{x\})$ we have $g/h \in C(E^\Delta)$, and therefore

$$Bg(x) = B\left(\frac{g}{h}\, h\right)(x) = \int_{E^\Delta} \frac{g(\xi)}{h(\xi)} \mu(x, d\xi) = \int_{E^\Delta} g(\xi)\nu_B(x, d\xi).$$

Since B is bounded, the identity $Bg(x) = \int_{E^\Delta} g(\xi)\nu_B(x, d\xi)$ extends by continuity to all $g \in C(E^\Delta)$ with $g(x) = 0$. Thus, using also that $B1 = 0$,

$$Bg(x) = B(g - g(x))(x) = \int_{E^\Delta} (g(\xi) - g(x))\nu_B(x, d\xi).$$

Repeating this for every $x \in E^\Delta$ yields that ν_B satisfies (C.1) for all $x \in E^\Delta$ and $g \in C(E^\Delta)$. To see that $\nu_B(x, E^\Delta) < \infty$, note that $\int_{E^\Delta} g(\xi)\nu_B(x, d\xi) \leq \|B\|$ whenever $g \in C(E^\Delta)$ satisfies $0 \leq g \leq 1$ and $g(x) = 0$. Measurability of $\nu_B(\,\cdot\,, A)$ for every Borel set $A \subseteq E^\Delta$ follows from a monotone class argument, so that ν_B is indeed a kernel from E^Δ to E^Δ. $\qquad\square$

Lemma C.3. *Let $D \subseteq C_\Delta(E)$ be a dense linear subspace containing the constant function 1, and let $Q\colon D \otimes D \to \widehat{C}_\Delta(E^2)$ be a linear operator. The following conditions are equivalent:*

(i) $Q(g \otimes g)(x, y) \geq 0$ for all $g \in D$, $x, y \in E$, with equality if $g(x) = g(y)$.

(ii) $\langle Q(g \otimes g), \nu^2 \rangle \geq 0$ for all $g \in D$ and $\nu \in M_1(E)$, with equality if g is constant on the support of ν.

If either condition is satisfied, then Q is of the form $Q = \alpha\Psi$ for some nonnegative symmetric function $\alpha\colon E^2 \to \mathbb{R}$.

Proof. It is clear that (i) implies (ii). For the converse, first note that for any $x \in E$ and $g \in D$, trivially g is constant on the support of δ_x. Thus $Q(g \otimes g)(x, x) = \langle Q(g \otimes g), \delta_x^2 \rangle = 0$. Taking $\nu = \frac{1}{2}(\delta_x + \delta_y)$ for any $x, y \in E$ then yields $Q(g \otimes g)(x, y) = \langle Q(g \otimes g), \nu^2 \rangle \geq 0$, with equality if $g(x) = g(y)$ since g is then constant on the support of ν. This proves that (ii) implies (i).

It remains to obtain the stated form of Q under the assumption that (i) holds. If E is a singleton then $Q = 0$, so we may assume that E contains at least two points. Fix $x, y \in E$ with $x \neq y$. Due to (i), the map $(g, h) \mapsto Q(g \otimes h)(x, y)$ is bilinear and positive semidefinite, and therefore satisfies the Cauchy–Schwarz inequality

$$|Q(g \otimes h)(x, y)| \leq \sqrt{Q(g \otimes g)(x, y)}\,\sqrt{Q(h \otimes h)(x, y)}.$$

Along with (i) this implies that $Q(g \otimes h)(x, y)$ depends on g and h only through their values at x and y. Moreover, since D is dense in $C_\Delta(E)$, for every $a \in \mathbb{R}^2$ there exists $g \in D$ such that $a = (g(x), g(y))$. Thus there is a unique map $T\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ such that

$$Q(g \otimes h)(x, y) = T(a, b) \quad \text{where} \quad a = \begin{pmatrix} g(x) \\ g(y) \end{pmatrix}, \; b = \begin{pmatrix} h(x) \\ h(y) \end{pmatrix}.$$

The map T inherits bilinearity and positive semidefiniteness. Since $Q(g \otimes 1)(x, y) = 0$ due to the Cauchy–Schwarz inequality and (i), we also have $T(a, b) = 0$ for $b = (1, 1)$. This implies that $T(a, b) = \frac{1}{2}\alpha(x, y)(a_1 - a_2)(b_1 - b_2)$ for some $\alpha(x, y) \in \mathbb{R}_+$. Thus,

$$Q(g \otimes h)(x, y) = \frac{1}{2}\alpha(x, y)(g(x) - g(y))(h(x) - h(y)) = \alpha(x, y)\Psi(g \otimes h)(x, y).$$

Defining $\alpha(x,x)$ arbitrarily, we obtain the map $\alpha\colon E^2 \to \mathbb{R}$, which is symmetric due to the symmetry of $Q(g \otimes h)$. $\qquad\square$

Consider now the setting of Lemma 6.6, i.e. $E = \mathbb{R}$ and $D = \mathbb{R} + C_c^\infty(\mathbb{R})$.

Lemma C.4. *Consider two operators $B : D \to C_\Delta(\mathbb{R})$ and $Q : D \otimes D \to \widehat{C}_\Delta(\mathbb{R}^2)$ such that B is as in* (6.4) *and Q satisfies*

$$Q(h \otimes h)(x,y) \geq 0 \text{ for all } h \in D, \text{ with equality if } h(x) = h(y) \text{ and } h'(x) = h'(y) = 0.$$

Then, for each $g \in D$ and $x,y \in \mathbb{R}$ such that $g(x) = g(y) = 1$ there exists a sequence $(p_n)_{n\in\mathbb{N}} \subseteq P^D$ such that

$$p_n(\nu_\lambda) = \max_{M_1(\mathbb{R})} p_n, \quad \partial p_n(\nu_\lambda) = f_n, \quad \text{and} \quad \langle Q(\partial^2 p_n(\nu_\lambda)), \nu_\lambda^2 \rangle = \langle Q(g \otimes g), \nu_\lambda^2 \rangle$$

for all $n \in \mathbb{N}$ and $\lambda \in [0,1]$, where $\nu_\lambda = \lambda\delta_x + (1-\lambda)\delta_y$ and $(f_n)_{n\in\mathbb{N}}$ satisfies

$$\lim_{n\to\infty} -2Bf_n(z) = (a(z)g'(z))^2, \qquad z \in \{x,y\}.$$

Proof. Fix $g \in D$ such that $g(x) = g(y) = 1$. Let $F_n : [0,1] \to \mathbb{R}$ as in Lemma B.1 and fix a compactly supported function $\rho \in C_c^\infty(\mathbb{R})$ such that $\rho = 1$ on some neighborhoods of x and y and $\rho(\mathbb{R}) \subseteq [0,1]$. Set then

$$\overline{g}_n(z) = 1 + g'(x)(z-x)\frac{(z-y)^2}{(x-y)^2}F_{n^4}\left(\frac{|z-x|^2}{C_x}\right) + g'(y)(z-y)\frac{(z-x)^2}{(x-y)^2}F_{n^4}\left(\frac{|z-y|^2}{C_y}\right),$$

where $C_x = 2\sup_{z\in\text{supp}(\rho)}(z-x)^2$. Setting $g_n = 1 + (\overline{g}_n - 1)\rho$ we get $g_n \in \mathbb{R} + C_c^\infty(\mathbb{R}) = D$. For n even, define now the polynomial

$$p_n(\nu) = \frac{1}{n(n-1)}\left(\langle g_n, \nu\rangle^n - \langle g_n^n, \nu\rangle\right).$$

Since $p_n(\nu_\lambda) = 0$ and by Jensen inequality $p_n \leq 0$, we can conclude that ν_λ maximizes p_n for all n even and $\lambda \in [0,1]$. Observe that

$$\partial p_n(\nu_\lambda) = \frac{1}{n-1}\left(g_n - \frac{1}{n}g_n^n\right) =: f_n \quad \text{and} \quad \partial^2 p_n(\nu_\lambda) = g_n \otimes g_n.$$

Proceeding as in the proof of Lemma C.3, we can use the assumptions on Q to prove that $Q(g \otimes h)(x,y)$ depends on g and h only through their values and the values of their derivatives at x and y. Since $g_n(z) = g(z) = 1$ and $g_n'(z) = g'(z)$ for all n even and $z \in \{x,y\}$, this implies that $\langle Q(g_n \otimes g_n), \nu_\lambda^2\rangle = \langle Q(g \otimes g), \nu_\lambda^2\rangle$. Finally, the representation of B given by (6.4) yields

$$-2Bf_n(z) = \left(a(z)g'(z)\right)^2 - 2\int \frac{1}{n-1}\left(g_n(z+\xi) - \frac{1}{n}g_n(z+\xi)^n\right) - \frac{1}{n}\,F(z,d\xi),$$

for all $z \in \{x,y\}$. Since by the dominated convergence theorem the integral term converges to 0 for n going to ∞, this concludes the proof. $\qquad\square$

D Existence for martingale problems

The purpose of this section is to establish the (essential) equivalence between the existence of a solution to the martingale problem for L and the positive maximum principle for L. Here, E is a locally compact Polish space, D a dense linear subspace of $C_\Delta(E)$ containing the constant function 1, and $L\colon P^D \to P$ a linear operator satisfying (4.2).

The first lemma asserts that the positive maximum principle is implied if a solution to the martingale problem exists.

Lemma D.1. *If there exists a solution X to the martingale problem for L for each initial condition in $M_1(E)$, then L satisfies the positive maximum principle on $M_1(E)$.*

The proof of Lemma D.1 is standard and we thus omit it. See for instance the proof of Lemma 2.3 in [21].

The next lemma is an adaptation of a classical result from [17]. For the application of this result it is crucial that L is an operator on the space of bounded continuous functions on a locally compact, separable, metrizable space. Since this is not the case for $M_1(E)$ if E is noncompact, we work on $M_1(E^\Delta)$, which is a compact Polish space with respect to the topology of weak convergence.

The result of [17] can then be applied and we just have to check that if the initial condition of an $M_1(E^\Delta)$ solution X assigns mass 1 to E, then $X_t(E) = 1$ almost surely for each $t \geq 0$, so that the solution actually takes values in $M_1(E)$.

Lemma D.2. *Suppose that L satisfies the positive maximum principle on $M_1(E^\Delta)$. If B is E-conservative, then there exists a solution to the martingale problem for L for every initial condition in $M_1(E)$.*

Proof. Recall that because of Lemma 2.3(i), the operator L can be seen as an operator on the space polynomials on $M_1(E^\Delta)$. The first part of the proof consists then in proving that if L satisfies the positive maximum principle on $M_1(E^\Delta)$ then there exists an $M_1(E^\Delta)$-valued solution to the martingale problem for L for every initial condition in $M_1(E^\Delta)$. This result is a consequence of Theorem 4.5.4 in [17] and the successive Remark 4.5.5. We now explain how the necessary conditions hold true. Observe that $M_1(E^\Delta)$ is a compact separable metrizable space and, by Lemma 2.7, that $P^D(M_1(E^\Delta)) := \{p|_{M_1(E^\Delta)} : p \in P^D\}$ is a dense subset of the space of continuous functions on $M_1(E^\Delta)$. Moreover, the positive maximum principle implies that $Lp|_{M_1(E^\Delta)} = Lq|_{M_1(E^\Delta)}$ for all $p, q \in P^D$ such that $p|_{M_1(E^\Delta)} = q|_{M_1(E^\Delta)}$. We may thus regard L as an operator on the space of continuous functions on $M_1(E^\Delta)$ with domain $P^D(M_1(E^\Delta))$.

For the second part, recall that by definition of E-conservativity there exist functions $g_n \in D \cap C_0(E)$ such that $\lim_{n\to\infty} g_n = 1$, and $\lim_{n\to\infty}(Bg_n)^- = 0$ bounded pointwise on E and E^Δ, respectively. By the dominated convergence theorem, (5.1), and Fatou's lemma we can compute

$$\mathbb{E}[X_t(E)] = \lim_{n\to\infty} \mathbb{E}[\langle g_n, X_t \rangle] = \lim_{n\to\infty} \left(\langle g_n, \nu \rangle + \mathbb{E}\left[\int_0^t \langle Bg_n, X_s \rangle ds \right] \right) \geq \nu(E) = 1.$$

Finally, note that a càdlàg process X on $M_1(E^\Delta)$ such that $X_t(E) = 1$ almost sure is càdlàg also with respect to the topology of weak convergence on $M_1(E)$. $\square$

References

[1] A. Ahdida and A. Alfonsi. A mean-reverting SDE on correlation matrices. *Stochastic Processes and their Applications*, 123(4):1472–1520, 2013. MR-3016230

[2] D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de Probabilités XIX 1983/84*, pages 177–206. Springer, 1985. MR-0889476

[3] C. Beck, S. Becker, P. Grohs, N. Jaafari, and A. Jentzen. Solving stochastic differential equations and Kolmogorov equations by means of deep learning. *Preprint arXiv:1806.00421*, 2018.

[4] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1995. MR-3444832

[5] A. Boumezoued, H. Hardy, N. El Karoui, and S. Arnold. Cause-of-death mortality: What can be learned from population dynamics? *Insurance: Mathematics and Economics*, 78:301–315, 2018. MR-3761084

[6] P. Courrège. Sur la forme intégro-différentielle des opérateurs de C_k^∞ dans C satisfaisant au principe du maximum. *Séminaire Brelot-Choquet-Deny (Théorie du Potentiel)*, 10(2), 1965. MR-0114057

[7] C. Cuchiero. Polynomial processes in stochastic portfolio theory. *Stochastic processes and their applications*, Forthcoming, 10.1016/j.spa.2018.06.007, 2018.

[8] C. Cuchiero, M. Keller-Ressel, and J. Teichmann. Polynomial processes and their applications to mathematical finance. *Finance and Stochastics*, 16:711–740, 2012. MR-2972239

[9] C. Cuchiero, M. Larsson, and S. Svaluto-Ferro. Polynomial jump-diffusions on the unit simplex. *Annals of Applied Probability*, 28(4):2451–2500, 2018. MR-3843834

[10] G. Da Prato and H. Frankowska. Invariance of stochastic control systems with deterministic arguments. *Journal of Differential Equations*, 200(1):18–52, 2004. MR-2046316

[11] D. Dawson. Measure-valued Markov processes. In *École d'Été de Probabilités de Saint-Flour XXI–1991*, Lecture Notes in Mathematics, pages 1–260. Springer, 1993. MR-1242575

[12] D. Dawson and K. Hochberg. Wandering Random Measures in the Fleming-Viot Model. *Annals of Probability*, 10(3):554–580, 1982. MR-0659528

[13] D. Duffie, D. Filipović, and W. Schachermayer. Affine processes and applications in finance. *Annals of applied probability*, pages 984–1053, 2003. MR-1994043

[14] A. Etheridge. *Some Mathematical Models from Population Genetics: École D'Été de Probabilités de Saint–Flour XXXIX-2009*. Lecture Notes in Mathematics. Springer, 2011. MR-2759587

[15] S. N. Ethier and T. G. Kurtz. *The Infinitely-Many-Alleles Model with Selection as a Measure-Valued Diffusion*, pages 72–86. Springer Berlin Heidelberg, Berlin, Heidelberg, 1987. MR-0893637

[16] S. N. Ethier and T. G. Kurtz. Fleming–Viot processes in population genetics. *SIAM Journal on Control and Optimization*, 31(2):345–386, 1993. MR-1205982

[17] S. N. Ethier and T.G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley Series in Probability and Statistics. Wiley, 2 edition, 2005. MR-0838085

[18] R. Fernholz. *Stochastic Portfolio Theory*. Applications of Mathematics. Springer-Verlag, New York, 2002. MR-1894767

[19] R. Fernholz and I. Karatzas. Relative arbitrage in volatility-stabilized markets. *Annals of Finance*, 1(2):149–177, 2005.

[20] R. Fernholz and I. Karatzas. Stochastic portfolio theory: an overview. *Handbook of numerical analysis*, 15:89–167, 2009. MR-1894767

[21] D. Filipović and M. Larsson. Polynomial diffusions and applications in finance. *Finance and Stochastics*, 20(4):931–972, 2016. MR-3551857

[22] D. Filipović and M. Larsson. Polynomial jump-diffusion models. *ArXiv e-prints*, 2017. URL https://arxiv.org/abs/1711.08043.

[23] W. H. Fleming and M. Viot. Some measure-valued Markov processes in population genetics theory. *Indiana Univ. Math. J.*, 28(5):817–843, 1979. MR-0542340

[24] W. Hoh. Pseudo differential operators generating Markov processes. *Habilitationsschrift, Univeristät Bielefeld*, 1998.

[25] M. Kimura. Diffusion models in population genetics. *Journal of Applied Probability*, 1(2):177–232, 1964. MR-0172727

[26] A. Klenke. *Probability Theory: A Comprehensive Course*. Universitext. Springer London, 2 edition, 2013. MR-3112259

[27] P. Kotelenez and T. Kurtz. Macroscopic limits for stochastic partial differential equations of McKean–Vlasov type. *Probability Theory and Related Fields*, 146(1):189, 2008. MR-2550362

[28] T. Kurtz and J. Xiong. Particle representations for a class of nonlinear SPDEs. *Stochastic Processes and their Applications*, 83, 1999. MR-1705602

[29] E. Regazzini, A. Guglielmi, and G. Di Nunno. Theory and numerical analysis for exact distributions of functionals of a Dirichlet process. *Ann. Statist.*, 30(5):1376–1411, 2002. MR-1936323

[30] E. Regazzini, A. Lijoi, and I. Prünster. Distributional results for means of normalized random measures with independent increments. *Ann. Statist.*, 31(2):560–585, 2003. Dedicated to the memory of Herbert E. Robbins. MR-1983542

[31] M. Shkolnikov. Large volatility-stabilized markets. *Stochastic Processes and their Applications*, 123(1):212–228, 2013. MR-2988116

[32] J. Vaillancourt. On the existence of random McKean–Vlasov limits for triangular arrays of exchangeable diffusions. *Stochastic Analysis and Applications*, 6(4):431–446, 1988. MR-0964251

[33] J. Vaillancourt. Interacting Fleming-Viot processes. *Stochastic processes and their applications*, 36(1):45–57, 1990. MR-1075600