

Polynomial diffusions and applications in finance

Damir Filipović¹ · Martin Larsson²

Received: 26 May 2015 / Accepted: 21 December 2015 / Published online: 14 June 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract This paper provides the mathematical foundation for polynomial diffusions. They play an important role in a growing range of applications in finance, including financial market models for interest rates, credit risk, stochastic volatility, commodities and electricity. Uniqueness of polynomial diffusions is established via moment determinacy in combination with pathwise uniqueness. Existence boils down to a stochastic invariance problem that we solve for semialgebraic state spaces. Examples include the unit ball, the product of the unit cube and nonnegative orthant, and the unit simplex.

Keywords Polynomial diffusions · Polynomial diffusion models in finance · Stochastic invariance · Boundary attainment · Moment problem

Mathematics Subject Classification 60H30 · 60J60

JEL Classification G12 · G13

1 Introduction

This paper provides the mathematical foundation for polynomial diffusions on a large class of state spaces in \mathbb{R}^d . A polynomial diffusion is characterized by having a linear

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. 307465-POLYTE.

✉ M. Larsson
martin.larsson@math.ethz.ch

¹ EPFL and Swiss Finance Institute, Quartier UNIL-Dorigny, Extranef 218, 1015 Lausanne, Switzerland

² Department of Mathematics, ETH Zurich, Rämistrasse 101, 8092, Zurich, Switzerland

drift and quadratic diffusion function. In consequence, moments are given in closed form. Such processes represent an extension of the affine class. They play an important role in a growing range of applications in finance, including financial market models of interest rates, credit risk, stochastic volatility, and commodities and electricity.

An arbitrage-free financial market model is determined by a state price density, i.e., a positive semimartingale ζ defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$. The model price $\Pi(t, T)$ at time t of any time T cash-flow C_T is given by

$$\Pi(t, T) = \frac{1}{\zeta_t} \mathbb{E} [\zeta_T C_T \mid \mathcal{F}_t]. \quad (1.1)$$

We may interpret \mathbb{P} as the historical measure, or more generally as an auxiliary measure possibly different from, but equivalent to, the historical measure. A *polynomial diffusion model* consists of a polynomial diffusion X as factor process, along with a positive polynomial p on the state space. The state price density is specified by $\zeta_t = e^{-\alpha t} p(X_t)$, where α is a real parameter chosen to control the lower bound on implied interest rates. We let the time T cash-flow of a security be given by $C_T = q(X_T)$ for some polynomial q . The polynomial property of X along with the elementary fact that pq is a polynomial implies that $\Pi(t, T)$ becomes a rational function in X_t with coefficients given in closed form in terms of a matrix exponential. Polynomial diffusion models thus yield closed form expressions for any security with cash-flows specified as polynomial functions of X , which makes them universally applicable in finance. This includes financial market models for interest rates (with $C_T = 1$), credit risk in a doubly stochastic framework (with C_T the conditional survival probability), stochastic volatility (with C_T the spot variance), and commodities and electricity (with C_T the spot price).

While polynomial diffusions have appeared in the literature since Wong [48], so far no existence and uniqueness theory has been available beyond the scalar case. This paper fills this gap and thus provides the mathematical foundation for polynomial diffusion models in finance.

Our main uniqueness result (Theorem 4.2) is based on the classical theory of the moment problem. Since the mixed moments of all finite-dimensional marginal distributions of a polynomial diffusion are uniquely determined by its generator (Theorem 3.1 and Corollary 3.2), uniqueness follows whenever these moments determine the underlying distribution. This is often true, for instance in the affine case or when the state space is compact, or more generally if exponential moments exist; Theorem 3.3 provides sufficient conditions. There are, however, situations where the moment problem approach fails. We therefore provide two additional results based on Yamada–Watanabe type arguments, which give uniqueness in the one-dimensional case (Theorem 4.3) as well as when the process dynamics exhibits a certain hierarchical structure (Theorem 4.4). These uniqueness results do not depend on the geometry of the state space.

In order to study existence, we assume that the state space is a basic closed semi-algebraic set, i.e., the nonnegativity set of a finite family of polynomials. Existence reduces to a stochastic invariance problem that we solve under suitable geometric

and algebraic conditions on the state space (Theorem 5.3). We also study boundary attainment. In applications, it is frequently of interest to know whether the trajectories of a given process may hit the boundary of the state space. In particular, simulating trajectories becomes a much more delicate task if the boundary is attained; see Lord et al. [35]. We present sufficient conditions for both attainment and non-attainment that are tight (Theorem 5.7).

A semialgebraic state space is a natural choice for at least three reasons. First, positive semidefiniteness of the quadratic diffusion matrix boils down to nonnegativity constraints on polynomials. Second, polynomial diffusion models in finance involve polynomials that are required to be positive on the state space. And third, semialgebraic sets turn out to be an ideal setting for employing tools from real algebraic geometry to verify the hypotheses of our existence and boundary attainment results.

We give a detailed analysis of some specific semialgebraic state spaces that do and will play an important role in financial applications, and that illustrate the scope of polynomial diffusions. Specifically, we consider certain quadric sets including the unit ball $\{x \in \mathbb{R}^d : \|x\| \leq 1\}$, the product space $[0, 1]^m \times \mathbb{R}_+^n$, and the unit simplex $\{x \in \mathbb{R}_+^d : x_1 + \dots + x_d = 1\}$. We also elaborate on polynomial diffusion models in finance, and show how to specify novel stochastic models for interest rates, stochastic volatility, and stock markets.

Polynomial processes have been studied in various degrees of generality by several authors, for instance Wong [48], Mazet [38], Zhou [49], Forman and Sørensen [24], among others. The first systematic accounts treating the time-homogeneous Markov jump-diffusion case are Cuchiero [9] and Cuchiero et al. [10]. The use of polynomial diffusions in financial modeling goes back at least to the early 2000s. Zhou [49] used one-dimensional polynomial (jump-)diffusions to build short rate models that were estimated to data using a generalized method-of-moments approach, relying crucially on the ability to compute moments efficiently. A short rate model based on the Jacobi process was presented by Delbaen and Shirakawa [15], and Larsen and Sørensen [33] used the same process for exchange rate modeling. The multidimensional Jacobi process was studied by Gouriéroux and Jasiak [27], who constructed a stock price model with smooth transitions of drift and volatility regimes. More recently, polynomial diffusions have featured in the context of financial applications in several papers; see Filipović et al. [23, 21] for models of the term structure of variance swap rates and interest rates, respectively, models, and Cuchiero et al. [10] for variance reduction for option pricing and hedging, among other applications. There are several reasons for moving beyond the affine class. In particular, nontrivial dynamics on compact state spaces become a possibility, which together with the polynomial property fits well with polynomial expansion techniques; see also Filipović et al. [20]. Also on non-compact state spaces, one can achieve richer dynamics than in the affine case. Examples of non-affine polynomial processes include multidimensional Jacobi or Fisher–Wright processes (Ethier [18], Gouriéroux and Jasiak [27]), Pearson diffusions (Forman and Sørensen [24]), and Dunkl processes (Dunkl [17], Gallardo and Yor [25]).

The rest of the paper is structured as follows. In Sect. 2, we define polynomial diffusions. Section 3 is concerned with power and exponential moments. In Sect. 4, we discuss uniqueness. In Sect. 5, we treat existence and boundary attainment. Section 6 contains examples of semialgebraic state spaces. Section 7 outlines various

polynomial diffusion models in finance. For the sake of readability, most proofs are given in Appendices A–I. Some basic notions from algebraic geometry are reviewed in Appendix J.

We end this introduction with some notational conventions that will be used throughout this paper. For a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$, we write $\{f = 0\}$ for the set $\{x \in \mathbb{R}^d : f(x) = 0\}$. A *polynomial* p on \mathbb{R}^d is a map $\mathbb{R}^d \rightarrow \mathbb{R}$ of the form $\sum_{\alpha} c_{\alpha} x_1^{\alpha_1} \cdots x_d^{\alpha_d}$, where the sum runs over all multi-indices $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$ and only finitely many of the coefficients c_{α} are nonzero. Such a representation is unique. The *degree* of p is the number $\deg p = \max\{\alpha_1 + \cdots + \alpha_d : c_{\alpha} \neq 0\}$. We let $\text{Pol}(\mathbb{R}^d)$ denote the ring of all polynomials on \mathbb{R}^d , and $\text{Pol}_n(\mathbb{R}^d)$ the subspace consisting of polynomials of degree at most n . Let E be a subset of \mathbb{R}^d . A *polynomial on E* is the restriction $p = q|_E$ to E of a polynomial $q \in \text{Pol}(\mathbb{R}^d)$. Its degree is $\deg p = \min\{\deg q : p = q|_E, q \in \text{Pol}(\mathbb{R}^d)\}$. We let $\text{Pol}(E)$ denote the ring of polynomials on E , and $\text{Pol}_n(E)$ the subspace of polynomials on E of degree at most n . Both $\text{Pol}_n(\mathbb{R}^d)$ and $\text{Pol}_n(E)$ are finite-dimensional real vector spaces, but if there are nontrivial polynomials that vanish on E , their dimensions will be different. If E has a nonempty interior, then $\text{Pol}_n(\mathbb{R}^d)$ and $\text{Pol}_n(E)$ can be identified. The set of real symmetric $d \times d$ matrices is denoted \mathbb{S}^d , and the subset of positive semidefinite matrices is denoted \mathbb{S}_+^d .

2 Definition of polynomial diffusions

Throughout this paper, we fix maps $a : \mathbb{R}^d \rightarrow \mathbb{S}^d$ and $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ with

$$a_{ij} \in \text{Pol}_2(\mathbb{R}^d) \text{ and } b_i \in \text{Pol}_1(\mathbb{R}^d) \text{ for all } i, j \tag{2.1}$$

and a state space $E \subseteq \mathbb{R}^d$. Our goal is to investigate the following issues:

- (a) For a suitable class of state spaces E , find conditions on a, b, E that guarantee the existence of an E -valued solution to the stochastic differential equation

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t \tag{2.2}$$

for some d -dimensional Brownian motion W and some continuous function $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ with $\sigma \sigma^{\top} = a$ on E . We shall consider the class of basic closed semialgebraic sets E , defined using polynomial equalities and inequalities.

- (b) Find conditions for *uniqueness in law for E -valued solutions to (2.2)*. By this we mean that for any $x \in E$ and any E -valued solutions X and X' to (2.2) with $X_0 = X'_0 = x$, possibly with different driving Brownian motions, X and X' have the same law.
- (c) Find conditions for a solution to (2.2) to attain the boundary of E .
- (d) Find large parametric classes of a, b, E for which (2.2) admits a solution.

Investigating these issues is motivated by the fact that diffusions as in (2.2) admit closed form conditional moments and have broad applications in finance, as we shall see below.

We consider the partial differential operator \mathcal{G} given by

$$\mathcal{G}f = \frac{1}{2} \text{Tr}(a \nabla^2 f) + b^\top \nabla f. \tag{2.3}$$

In view of (2.1), \mathcal{G} maps $\text{Pol}_n(\mathbb{R}^d)$ to itself for each $n \in \mathbb{N}$. As we work on a state space $E \subseteq \mathbb{R}^d$, we now refine this property. We say that \mathcal{G} is *well defined on* $\text{Pol}(E)$ if $\mathcal{G}f = 0$ on E for any $f \in \text{Pol}(\mathbb{R}^d)$ with $f = 0$ on E . In this case, \mathcal{G} is well defined as an operator on $\text{Pol}(E)$. This always holds if E has a nonempty interior.

Definition 2.1 The operator \mathcal{G} is called *polynomial on* E if it is well defined on $\text{Pol}(E)$, and thus maps $\text{Pol}_n(E)$ to itself for each $n \in \mathbb{N}$. In this case, we call any E -valued solution to (2.2) a *polynomial diffusion on* E .

It is a simple matter to verify that any second order partial differential operator that maps $\text{Pol}_n(E)$ to itself for each $n \in \mathbb{N}$ is necessarily of the form (2.1) and (2.3) on E .

Lemma 2.2 Let $\tilde{\mathcal{G}}f = \frac{1}{2} \text{Tr}(\tilde{a} \nabla^2 f) + \tilde{b}^\top \nabla f$ be a partial differential operator for some maps $\tilde{a} : \mathbb{R}^d \rightarrow \mathbb{S}^d$ and $\tilde{b} : \mathbb{R}^d \rightarrow \mathbb{R}^d$. Assume $\tilde{\mathcal{G}}$ is well defined on $\text{Pol}(E)$. Then the following are equivalent:

- (i) $\tilde{\mathcal{G}}$ maps $\text{Pol}_n(E)$ to itself for each $n \in \mathbb{N}$.
- (ii) $\tilde{\mathcal{G}}$ maps $\text{Pol}_n(E)$ to itself for $n \in \{1, 2\}$.
- (iii) The components of \tilde{a} and \tilde{b} restricted to E lie in $\text{Pol}_2(E)$ and $\text{Pol}_1(E)$, respectively.

In this case, \tilde{a} and \tilde{b} restricted to E are uniquely determined by the action of $\tilde{\mathcal{G}}$ on $\text{Pol}_2(E)$.

Proof The implications (i) \Rightarrow (ii) and (iii) \Rightarrow (i) are immediate, and the implication (ii) \Rightarrow (iii) follows upon applying $\tilde{\mathcal{G}}$ to the monomials of degree one and two. In particular, this pins down \tilde{a} and \tilde{b} on E , and thus also establishes the last part of the lemma. □

In the one-dimensional case $d = 1$, one can classify all polynomial diffusions on intervals E . Indeed, one has $a(x) = a + \alpha x + Ax^2$ and $b(x) = b + \beta x$ for some scalars a, α, A, b, β , and $E = \{x \in \mathbb{R} : a(x) \geq 0\}$. See Forman and Sørensen [24] and Filipović et al. [23] for details.

The multidimensional case is less trivial. For example, let $d = 2$, $E = \mathbb{R} \times \{0\}$, and consider the operator $\mathcal{G}f(x, y) = \frac{1}{2} \partial_{xx} f(x, y) + \partial_y f(x, y)$. This operator is not well defined on $\text{Pol}(E)$, since the polynomial $f(x, y) = y$ vanishes on E , but $\mathcal{G}f(x, y) = 1$. On the other hand, \mathcal{G} is the generator of the diffusion $dX_t = (dB_t, dt)$, where B is a one-dimensional Brownian motion. This process immediately leaves E for any starting point $x \in E$. If, however, an E -valued solution to (2.2) exists for any starting point $x \in E$, then \mathcal{G} is well defined on $\text{Pol}(E)$. This follows from the following basic positive maximum principle.

Lemma 2.3 Consider $f \in C^2(\mathbb{R}^d)$ and suppose $\bar{x} \in E$ is a maximizer of f over E . If (2.2) admits an E -valued solution with $X_0 = \bar{x}$, then $\mathcal{G}f(\bar{x}) \leq 0$.

Proof Let X be an E -valued solution to (2.2) with $X_0 = \bar{x}$, and assume for contradiction that $\mathcal{G}f(\bar{x}) > 0$. By the definition of a global maximizer, $f(x) \leq f(\bar{x})$ for all $x \in E$. Let $\tau = \inf\{t \geq 0 : \mathcal{G}f(X_t) \leq 0\}$, and note that $\tau > 0$. Then for $t \in (0, \tau)$, we have $f(X_t) \leq f(\bar{x})$ and $\mathcal{G}f(X_t) > 0$, which implies

$$f(X_{t \wedge \tau}) - f(\bar{x}) - \int_0^{t \wedge \tau} \mathcal{G}f(X_s) ds < 0$$

for all $t > 0$. Thus the left-hand side is a local martingale starting from zero, strictly negative for all $t > 0$. This contradiction proves that $\mathcal{G}f(\bar{x}) \leq 0$. \square

Regarding uniqueness, it is crucial to restrict attention to E -valued solutions. To illustrate what can otherwise go wrong, consider the stochastic differential equation $dX_t = -2\sqrt{X_t^-} dt + 2\sqrt{X_t^+} dW_t$, which is well known to have a unique \mathbb{R}_+ -valued solution: the zero-dimensional squared Bessel process. However, this stochastic differential equation admits other solutions that do not remain in \mathbb{R}_+ , for example $X_t = Y_t \mathbf{1}_{\{t \leq \tau\}} - (t - \tau)^2 \mathbf{1}_{\{t > \tau\}}$, where Y is a zero-dimensional squared Bessel process with $Y_0 \geq 0$ and $\tau = \inf\{t : Y_t = 0\}$. Here τ is finite almost surely.

Note that in Definition 2.1, we require neither uniqueness of solutions to (2.2), nor that \mathcal{G} be the generator of a Markov process on E . There are two reasons for this. First, existence of E -valued solutions to (2.2) does not in itself imply that those solutions are Markovian. Second, in the context of Markov processes, the polynomial property holds if and only if the corresponding semigroup leaves $\text{Pol}_n(E)$ invariant for each $n \in \mathbb{N}$. However, this fact, properly phrased, does not require the Markov property. Only Itô calculus is needed. This observation is crucial for our approach to proving uniqueness. Finally, we remark that a polynomial diffusion that is also a Markov process is a “polynomial process” in the terminology of Cuchiero et al. [10], with vanishing killing rate and no jumps.

3 Power and exponential moments

Throughout this section, we assume that \mathcal{G} is polynomial on E and let X be an E -valued solution to (2.2) realized on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$.

For any $n \in \mathbb{N}$, we let $N = N(n, E)$ denote the dimension of $\text{Pol}_n(E)$. We fix a basis of polynomials h_1, \dots, h_N for $\text{Pol}_n(E)$ and write

$$H(x) = (h_1(x), \dots, h_N(x))^\top.$$

Then for each $p \in \text{Pol}_n(E)$, there exists a unique vector $\vec{p} \in \mathbb{R}^N$ such that

$$p(x) = H(x)^\top \vec{p}. \tag{3.1}$$

The restriction of \mathcal{G} to $\text{Pol}_n(E)$ has a unique matrix representation $G \in \mathbb{R}^{N \times N}$, characterized by the property that $G\vec{p}$ is the coordinate vector of $\mathcal{G}p$ whenever \vec{p} is the coordinate vector of p . That is, we have

$$\mathcal{G}p(x) = H(x)^\top G\vec{p}. \tag{3.2}$$

We now show that $\mathbb{E}[p(X_T) | \mathcal{F}_t]$ is indeed well defined as a polynomial function of X_t . Recall that we do not assume uniqueness of solutions to (2.2), and we do not require X to be Markov. The proof is given in Appendix B.

Theorem 3.1 *If $\mathbb{E}[\|X_0\|^{2n}] < \infty$, then for any $p \in \text{Pol}_n(E)$ with coordinate representation $\vec{p} \in \mathbb{R}^N$, we have*

$$\mathbb{E}[p(X_T) | \mathcal{F}_t] = H(X_t)^\top e^{(T-t)G}\vec{p}, \quad t \leq T.$$

The following result is a direct consequence of Theorem 3.1. Its statement and proof use standard multi-index notation: For a multi-index $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{N}_0^d$, we write $|\mathbf{k}| = k_1 + \dots + k_d$ and $x^\mathbf{k} = x_1^{k_1} \dots x_d^{k_d}$.

Corollary 3.2 *For any time points $0 \leq t_1 < \dots < t_m$ and for any multi-indices $\mathbf{k}(1), \dots, \mathbf{k}(m)$ such that*

$$\mathbb{E}[\|X_0\|^{2|\mathbf{k}(1)| + \dots + 2|\mathbf{k}(m)|}] < \infty,$$

the expectation $\mathbb{E}[X_{t_1}^{\mathbf{k}(1)} \dots X_{t_m}^{\mathbf{k}(m)}]$ is uniquely determined by \mathcal{G} and the law of X_0 .

Proof We prove the result for $m = 2$; the general case follows by iteration. Set $\mathbf{j} = \mathbf{k}(1)$, $\mathbf{k} = \mathbf{k}(2)$, and $n = |\mathbf{j}| + |\mathbf{k}|$. Since $\mathbb{E}[\|X_0\|^{2|\mathbf{k}|}] < \infty$, Theorem 3.1 yields $X_{t_1}^{\mathbf{j}} \mathbb{E}[X_{t_2}^{\mathbf{k}} | \mathcal{F}_{t_1}] = p(X_{t_1})$ for some polynomial $p \in \text{Pol}_n(E)$ whose coordinate representation \vec{p} only depends on G . Since $\mathbb{E}[\|X_0\|^{2n}] < \infty$, another application of Theorem 3.1 yields

$$\mathbb{E}[X_{t_1}^{\mathbf{j}} X_{t_2}^{\mathbf{k}}] = \mathbb{E}[\mathbb{E}[p(X_{t_1}) | \mathcal{F}_0]] = \mathbb{E}[H(X_0)^\top e^{t_1 G} \vec{q}].$$

This proves the corollary. □

We next provide conditions under which X_T admits finite exponential moments. This result will be used in connection with proving uniqueness in Theorem 4.2 below, but is also of interest on its own for applications in finance.¹ Its proof is given in Appendix C.

Theorem 3.3 *If*

$$\mathbb{E}[e^{\delta \|X_0\|}] < \infty \quad \text{for some } \delta > 0 \tag{3.3}$$

¹We thank Mykhaylo Shkolnikov for suggesting a way to improve an earlier version of this result.

and the diffusion coefficient satisfies the linear growth condition

$$\|a(x)\| \leq C(1 + \|x\|) \quad \text{for all } x \in E \quad (3.4)$$

for some constant C , then for each $t \geq 0$, there exists $\varepsilon > 0$ with $\mathbb{E}[e^{\varepsilon\|X_t\|}] < \infty$.

4 Uniqueness

Throughout this section, we assume that \mathcal{G} is polynomial on E . We present three results regarding uniqueness in law for E -valued solutions to (2.2). Recall that this notion of uniqueness pertains to deterministic initial conditions, as defined under (b) in Sect. 2.

The first result relies on the fact that the joint moments of all finite-dimensional marginal distributions of a polynomial diffusion are uniquely determined by \mathcal{G} ; see Corollary 3.2. Thus uniqueness in law follows if the finite-dimensional marginal distributions are the only ones with these moments. This property is known as *determinacy* in the literature on the moment problem, a classical topic in mathematics; references include Stieltjes [45], Akhiezer [3], Berg et al. [5], Schmüdgen [43], Stoyanov [46], Kleiber and Stoyanov [32] and many others.

Lemma 4.1 *Let X be an E -valued solution to (2.2). If for each $t \geq 0$, there exists $\varepsilon > 0$ with $\mathbb{E}[\exp(\varepsilon\|X_t\|)] < \infty$, then any E -valued solution to (2.2) with the same initial law as X has the same law as X . In particular, this holds if (3.3) and (3.4) are satisfied.*

Proof For any $t \geq 0$ and $i \in \{1, \dots, d\}$, the hypothesis yields $\mathbb{E}[\exp(\varepsilon|X_{i,t}|)] < \infty$ for some $\varepsilon > 0$. As a consequence, the moment-generating function of $X_{i,t}$ exists and is analytic in $(-\varepsilon, \varepsilon)$, hence equal to its power series expansion, and thus determined by the moments of $X_{i,t}$. By Curtiss [11, Theorem 1], the moment-generating function determines the law of $X_{i,t}$, which thus satisfies the determinacy property. Now, according to Petersen [40, Theorem 3], determinacy of the (one-dimensional) marginals of a measure on \mathbb{R}^m implies determinacy of the measure itself. It follows that determinacy holds for the law of each collection $(X_{t_1}, \dots, X_{t_m})$, $0 \leq t_1 < \dots < t_m$. By Corollary 3.2, the corresponding moments are the same for any E -valued solution to (2.2) with the same initial law as X . This proves the theorem. \square

If $X_0 = x$ is deterministic, then (3.3) holds and Lemma 4.1 directly yields our first result.

Theorem 4.2 *If the linear growth condition (3.4) is satisfied, then uniqueness in law for E -valued solutions to (2.2) holds.*

Theorem 4.2 assumes the linear growth condition (3.4) to ensure existence of exponential moments. While valid for all affine diffusions, as well as when E is compact, this condition excludes some interesting examples, in particular geometric

Brownian motion.² Uniqueness for the geometric Brownian motion holds of course, and can be established via the Yamada–Watanabe pathwise uniqueness theorem for one-dimensional diffusions. Our second result records this fact.

Theorem 4.3 *If the dimension is $d = 1$, then uniqueness in law for E -valued solutions to (2.2) holds.*

Proof Since \mathcal{G} is polynomial, the drift $b(x)$ in (2.2) is an affine function on E , and the dispersion restricted to E is of the form $\sigma(x) = \sqrt{\alpha + ax + Ax^2}$ for some real parameters α, a, A . Hence $b(x)$ is Lipschitz-continuous, and $\sigma(x)$ satisfies

$$(\sigma(x) - \sigma(y))^2 \leq \rho_n(|x - y|), \quad \text{for all } x, y \in E \text{ with } |x|, |y| \leq n,$$

where $\rho_n(z) = |a + 2nA|z$, for any $n \geq 1$. A localization argument in conjunction with Rogers and Williams [42, Theorem V.40.1] shows that pathwise uniqueness holds for any E -valued solution to (2.2). This in turn implies uniqueness in law; see Rogers and Williams [42, Theorem V.17.1]. □

Our third result, in combination with Theorems 4.2 and 4.3, yields uniqueness in a wide range of cases that are encountered in applications. The setup is the following. We assume that any E -valued solution to (2.2) can be partitioned as $X = (Y, Z)$, where Y is an autonomous m -dimensional diffusion with closed state space $E_Y \subseteq \mathbb{R}^m$, Z is n -dimensional, and $m + n = d$. That is, (Y, Z) solves the stochastic differential equation

$$dY_t = b_Y(Y_t) dt + \sigma_Y(Y_t) dW_t, \tag{4.1}$$

$$dZ_t = b_Z(Y_t, Z_t) dt + \sigma_Z(Y_t, Z_t) dW_t, \tag{4.2}$$

for polynomials $b_Y : \mathbb{R}^m \rightarrow \mathbb{R}^m$ and $b_Z : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ of degree one, continuous maps $\sigma_Y : \mathbb{R}^m \rightarrow \mathbb{R}^{m \times d}$ and $\sigma_Z : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times d}$, and where Y takes values in E_Y . The proof of the following theorem is given in Appendix D.

Theorem 4.4 *Assume that uniqueness in law for E_Y -valued solutions to (4.1) holds, and that σ_Z is locally Lipschitz in z , locally in y , on E . That is, for each compact*

²For geometric Brownian motion, there is a more fundamental reason to expect that uniqueness cannot be proved via the moment problem: it is well known that the lognormal distribution is not determined by its moments; see Heyde [29]. It thus becomes natural to pose the following question: *Can one find a process Y , essentially different from geometric Brownian motion, such that all joint moments of all finite-dimensional marginal distributions,*

$$\mathbb{E}[Y_{t_1}^{\alpha_1} \cdots Y_{t_m}^{\alpha_m}], \quad m \in \mathbb{N}, (\alpha_1, \dots, \alpha_m) \in \mathbb{N}^m, 0 \leq t_1 < \cdots < t_m < \infty,$$

coincide with those of geometric Brownian motion? We have not been able to exhibit such a process. Note that any such Y must possess a continuous version. Indeed, the known formulas for the moments of the lognormal distribution imply that for each $T \geq 0$, there is a constant $c = c(T)$ such that $\mathbb{E}[(Y_t - Y_s)^4] \leq c(t - s)^2$ for all $s \leq t \leq T, |t - s| \leq 1$, whence Kolmogorov’s continuity lemma implies that Y has a continuous version; see Rogers and Williams [42, Theorem I.25.2].

subset $K \subseteq E$, there exists a constant κ such that for all $(y, z, y', z') \in K \times K$,

$$\|\sigma_Z(y, z) - \sigma_Z(y', z')\| \leq \kappa \|z - z'\|. \tag{4.3}$$

Then uniqueness in law for E -valued solutions to (2.2) holds.

5 Existence and boundary attainment

In this section, we discuss existence of E -valued solutions to (2.2) and give conditions under which the boundary of the state space is attained. The results are stated and proved using some basic concepts from algebra and algebraic geometry. Appendix J provides a review of the required notions.

Existence of a solution to (2.2) with values in \mathbb{R}^d is well known to hold under linear growth conditions; see for instance Ikeda and Watanabe [31, Theorem IV.2.4]. The problem at hand thus boils down to finding conditions under which a solution to (2.2) takes values in E . This is a stochastic invariance problem. In Appendix A, we discuss necessary and sufficient conditions for nonnegativity of certain Itô processes, which is the basic tool we use for proving stochastic invariance.

We henceforth assume that the state space E is a *basic closed semialgebraic set*. Specifically, let \mathcal{P} and \mathcal{Q} be finite collections of polynomials on \mathbb{R}^d , and define

$$E = \{x \in M : p(x) \geq 0 \text{ for all } p \in \mathcal{P}\},$$

where

$$M = \{x \in \mathbb{R}^d : q(x) = 0 \text{ for all } q \in \mathcal{Q}\}. \tag{5.1}$$

In particular, if $\mathcal{Q} = \emptyset$ then $M = \mathbb{R}^d$. The following result provides simple necessary conditions for the invariance of E with respect to (2.2).

Theorem 5.1 *Suppose there exists an E -valued solution to (2.2) with $X_0 = x$, for any $x \in E$. Then*

- (i) $a \nabla p = 0$ and $\mathcal{G} p \geq 0$ on $E \cap \{p = 0\}$, for each $p \in \mathcal{P}$;
- (ii) $a \nabla q = 0$ and $\mathcal{G} q = 0$ on E , for each $q \in \mathcal{Q}$.

Proof Pick any $p \in \mathcal{P}$, $x \in E \cap \{p = 0\}$, and let X be a solution to (2.2) with $X_0 = x$. Then $p(X_t) = \int_0^t \mathcal{G} p(X_s) ds + \int_0^t \nabla p(X_s)^\top \sigma(X_s) dW_s$ and $p(X) \geq 0$, so (i) follows by Lemma A.1(ii). To prove (ii) for $q \in \mathcal{Q}$, simply apply the same argument to q and $-q$. □

The necessary condition $a \nabla p = 0$ states, roughly speaking, that at any boundary point of the state space, there can be no diffusive fluctuations orthogonally to the boundary. The necessary condition $\mathcal{G} p \geq 0$ can be interpreted as “inward-pointing adjusted drift” at the boundary. The following example shows that this cannot be replaced by a simple “inward-pointing drift” condition.

Example 5.2 Consider the bivariate process (U, V) with dynamics

$$\begin{aligned} dU_t &= dW_{1t}, & U_0 &\in \mathbb{R}, \\ dV_t &= \alpha dt + 2\sqrt{V_t} dW_{2t}, & V_0 &\in \mathbb{R}_+, \end{aligned}$$

where (W_1, W_2) is a Brownian motion and $\alpha > 0$. In other words, U is a Brownian motion and V is an independent squared Bessel process. The state space is $\mathbb{R} \times \mathbb{R}_+$. Now consider the process $(X, Y) = (U, V - U^2)$. Its dynamics is

$$\begin{aligned} dX_t &= dW_{1t}, \\ dY_t &= (\alpha - 1) dt - 2X_t dW_{1t} + 2\sqrt{X_t^2 + Y_t} dW_{2t}, \end{aligned}$$

and its state space is $E = \{(x, y) \in \mathbb{R}^2 : x^2 + y \geq 0\}$, the epigraph of the function $-x^2$. The drift of (X, Y) is $b(x, y) = (0, \alpha - 1)$, which points out of the state space at every boundary point, provided $\alpha < 1$. Nonetheless, with $p(x, y) = x^2 + y$, a calculation yields $\mathcal{G}p(x, y) = \alpha > 0$.

As a converse to Theorem 5.1, we now give sufficient conditions for the existence of an E -valued solution to (2.2). The proof of the following theorem is given in Appendix E.

Theorem 5.3 *Suppose E satisfies the following geometric and algebraic properties:*

- (G1) $\nabla r(x), r \in \mathcal{Q}$, are linearly independent for all $x \in M$;
- (G2) the ideals generated by $\mathcal{Q} \cup \{p\}$ and $M \cap \{p = 0\}$ are equal, i.e., we have $(\mathcal{Q} \cup \{p\}) = \mathcal{I}(M \cap \{p = 0\})$, for each $p \in \mathcal{P}$;

and the maps a and b satisfy

- (A0) $a \in \mathbb{S}_+^d$ on E ;
- (A1) $a \nabla p = 0$ on $M \cap \{p = 0\}$ and $\mathcal{G}p > 0$ on $E \cap \{p = 0\}$, for each $p \in \mathcal{P}$;
- (A2) $a \nabla q = 0$ and $\mathcal{G}q = 0$ on M , for each $q \in \mathcal{Q}$.

Then \mathcal{G} is polynomial on E , and there exists a continuous map $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ with $\sigma \sigma^\top = a$ on E and such that the stochastic differential equation (2.2) admits an E -valued solution X for any initial law of X_0 . This solution can be chosen so that it spends zero time in the sets $\{p = 0\}$, $p \in \mathcal{P}$. That is,

$$\int_0^t \mathbf{1}_{\{p(X_s)=0\}} ds = 0 \quad \text{for all } t \geq 0 \text{ and all } p \in \mathcal{P}. \tag{5.2}$$

Conditions (A1) and (A2) should be contrasted with the necessary conditions of Theorem 5.1. The latter are somewhat weaker, since they only make statements about a and b on E rather than M , and since the inequality in Theorem 5.1(i) is weak. Theorem 5.3 can be generalized to allow a weak inequality in (A1), at the cost of allowing absorption of the process at the boundary. We do not consider this generalization here.

Condition (G1) implies that M is an algebraic submanifold in \mathbb{R}^d of dimension $d - |\mathcal{Q}|$. The least obvious condition is arguably (G2). The crucial implication of (G2) is that any polynomial f that vanishes on $M \cap \{p = 0\}$ has a representation $f = hp$ on M for some polynomial h . In conjunction with (A1), this implies that $a(x)\nabla p(x)$ decays like $p(x)$ as $x \in E$ approaches the boundary set $E \cap \{p = 0\}$, for $p \in \mathcal{P}$. This allows one to prove that the local time of $p(X)$ at level zero vanishes, which makes Lemma A.1 applicable; see Appendix E for the details.

Condition (G2) is also the least straightforward to verify. We therefore present two sufficient conditions that are easier to check in concrete examples. The first condition is useful when $M = \mathbb{R}^d$, in which case each ideal appearing on the left-hand side in (G2) is generated by a single polynomial. This covers many interesting examples, yet yields conditions that are easy to verify in practice. A proof of the following result can be found in Bochnak et al. [6, Theorem 4.5.1].

Lemma 5.4 *Let $p \in \text{Pol}(\mathbb{R}^d)$ be an irreducible polynomial and $\mathcal{V}(p)$ its zero set. Then $(p) = \mathcal{I}(\mathcal{V}(p))$ if and only if p changes sign on \mathbb{R}^d , that is, $p(x)p(y) < 0$ for some $x, y \in \mathbb{R}^d$.*

The second condition applies when the ideals generated by the families $\mathcal{Q} \cup \{p\}$ with $p \in \mathcal{P}$ are prime and of full dimension.

Lemma 5.5 *For $p \in \mathcal{P}$, assume that the ideal $(\mathcal{Q} \cup \{p\})$ is prime with dimension $d - 1 - |\mathcal{Q}|$, and that there exists some $x \in M \cap \{p = 0\}$ such that the vectors $\nabla r(x)$, $r \in \mathcal{Q} \cup \{p\}$, are linearly independent. Then $(\mathcal{Q} \cup \{p\}) = \mathcal{I}(M \cap \{p = 0\})$.*

Proof This follows directly from Bochnak et al. [6, Proposition 3.3.16]. □

Remark 5.6 Stochastic invariance problems have been studied by a number of authors; see Da Prato and Frankowska [12], Filipović et al. [22], among many others. The approach in these papers is to impose an “inward-pointing Stratonovich drift” condition. This breaks down for polynomial diffusions. Indeed, consider the squared Bessel process

$$dX_t = \alpha dt + 2\sqrt{X_t} dW_t,$$

which is an \mathbb{R}_+ -valued affine process for $\alpha \geq 0$. The stochastic integral cannot always be written in Stratonovich form, since \sqrt{X} fails to be a semimartingale for $0 < \alpha < 1$. If nonetheless one formally computes the Stratonovich drift, one obtains $\alpha - 1$, suggesting that $\alpha \geq 1$ is needed for stochastic invariance of \mathbb{R}_+ . However, it is well known that $\alpha \geq 0$ is the correct condition. Our approach is rather in the spirit of Da Prato and Frankowska [13] who however focus on stochastic invariance of closed convex sets.

Apart from existence, Theorem 5.3 asserts that X spends zero time in the sets $\{p = 0\}$, $p \in \mathcal{P}$, which roughly speaking correspond to boundary segments of the state space. It does not, however, tell us whether these sets are actually hit. The purpose of the following theorem is to give necessary and sufficient conditions for this

to occur. The proof is given in Appendix F. The vector h of polynomials appearing in the theorem exists if (G2) and (A1) are satisfied.

Theorem 5.7 *Let X be an E -valued solution to (2.2) satisfying (5.2). Consider $p \in \mathcal{P}$ and let h be a vector of polynomials such that $a \nabla p = hp$ on M .*

(i) *Assume there exists a neighborhood U of $E \cap \{p = 0\}$ such that*

$$2\mathcal{G}p - h^\top \nabla p \geq 0 \quad \text{on } E \cap U.$$

Then $p(X_t) > 0$ for all $t > 0$.

(ii) *Assume (G2) holds and*

$$2\mathcal{G}p - h^\top \nabla p = 0 \quad \text{on } M \cap \{p = 0\}.$$

Then $p(X_t) > 0$ for all $t > 0$.

(iii) *Let $\bar{x} \in E \cap \{p = 0\}$ and assume*

$$\mathcal{G}p(\bar{x}) \geq 0 \quad \text{and} \quad 2\mathcal{G}p(\bar{x}) - h(\bar{x})^\top \nabla p(\bar{x}) < 0.$$

Then for any $T > 0$, there exists $\varepsilon > 0$ such that if $\|X_0 - \bar{x}\| < \varepsilon$ almost surely, then $p(X_t) = 0$ for some $t \leq T$ with positive probability.

As a simple example, we may apply Theorem 5.7 to the scalar square-root diffusion $dX_t = (b + \beta X_t)dt + \sigma \sqrt{X_t}dB_t$ with parameters $b, \sigma > 0$ and $\beta < 0$, and where B is a one-dimensional Brownian motion. In this case $E = \mathbb{R}_+$, and \mathcal{P} consists of the single polynomial $p(x) = x$. We have $a(x)p'(x) = \sigma^2 x = \sigma^2 p(x)$, so that $h(x) \equiv \sigma^2$, and thus

$$2\mathcal{G}p(x) - h(x)p'(x) = 2(b + \beta x) - \sigma^2.$$

It is well known that $X_t > 0$ for all $t > 0$ if and only if the Feller condition $2b \geq \sigma^2$ holds. Theorem 5.7(iii) gives the necessity of the Feller condition. Theorem 5.7(i) and (ii) together give the sufficiency of the Feller condition. Indeed, if $2b > \sigma^2$, then Theorem 5.7(i) applies, while the condition in Theorem 5.7(ii) is not satisfied. Theorem 5.7(ii) in turn applies when $2b = \sigma^2$, while Theorem 5.7(i) does not.

6 Examples of semialgebraic state spaces

We now discuss examples of semialgebraic state spaces of interest, where our results are applicable.

6.1 Some quadric sets

Let $Q \in \mathbb{S}^d$ be nonsingular, and consider the state space $E = \{x \in \mathbb{R}^d : x^\top Qx \leq 1\}$. Here \mathcal{P} consists of the single polynomial $p(x) = 1 - x^\top Qx$, and $M = \mathbb{R}^d$. After a linear change of coordinates, we may assume Q is diagonal with $Q_{ii} \in \{+1, -1\}$. We

also suppose $Q_{ii} = 1$ for at least some i , since otherwise $E = \mathbb{R}^d$. State spaces of this type include the closed unit ball, but also non-convex sets like $\{x \in \mathbb{R}^2 : x_1^2 - x_2^2 \leq 1\}$, whose boundary is a hyperbola. One can also consider complements of such sets; see Remark 6.3 below. One interesting aspect of the state spaces investigated here is that they do not admit non-deterministic affine diffusions; this follows directly from Proposition 6.1 below, which shows that a is either quadratic or identically zero. This is in contrast to the parabolic state spaces considered by Spreij and Veerman [44].

The following convex cone of polynomial maps plays a key role. Recall that a polynomial $r \in \text{Pol}(\mathbb{R}^d)$ is called *homogeneous of degree k* if $r(sx) = s^k r(x)$ for all $x \in \mathbb{R}^d$ and $s > 0$. We define

$$\mathcal{C}_+^Q = \left\{ c : \mathbb{R}^d \rightarrow \mathbb{S}_+^d : \begin{array}{l} c_{ij} \in \text{Pol}_2(\mathbb{R}^d) \text{ is homogeneous of degree 2 for all } i, j \\ \text{and } c(x)Qx = 0 \text{ for all } x \in \mathbb{R}^d \end{array} \right\}.$$

Note that the condition $c(x)Qx = 0$ is equivalent to $c(x)\nabla p(x) = 0$, meaning that all eigenvectors of $c(x)$ with nonzero eigenvalues are orthogonal to $\nabla p(x)$. The proof of the following proposition is given in Appendix G.

Proposition 6.1 *Conditions (G1) and (G2) hold for the above state space E . Moreover, the operator \mathcal{G} satisfies (A0)–(A2) if and only if*

$$\begin{aligned} a(x) &= (1 - x^\top Qx)\alpha + c(x), & (6.1) \\ b(x) &= \beta + Bx \end{aligned}$$

for some $\alpha \in \mathbb{S}_+^d$, $\beta \in \mathbb{R}^d$, $B \in \mathbb{R}^{d \times d}$ and $c \in \mathcal{C}_+^Q$ such that

$$\beta^\top Qx + x^\top B^\top Qx + \frac{1}{2} \text{Tr}(c(x)Q) < 0 \quad \text{for all } x \in \{p = 0\}. \quad (6.2)$$

Remark 6.2 If $c(x)$ satisfies the linear growth condition $\|c(x)\| \leq C(1 + \|x\|)$ for all $x \in E$, then $a(x)$ satisfies (3.4) and uniqueness in law for E -valued solutions to (2.2) holds by Theorem 4.2. In particular, this holds if Q is positive definite, i.e., $Q = \text{Id}$, so that E is the unit ball and hence compact.

Remark 6.3 The conditions of Proposition 6.1 can easily be modified to cover state spaces of the form $E = \{x \in \mathbb{R}^d : x^\top Qx \geq 1\}$. This amounts to replacing p by $-p$ above, and includes for example the complement of the open unit ball. With this modification, Proposition 6.1 is still true as stated, except that $-\alpha$ should lie in \mathbb{S}_+^d , and the inequality in (6.2) should be reversed.

A question that is not addressed by Proposition 6.1 is how to describe the set \mathcal{C}_+^Q in more explicit terms. We now provide a class of maps $c \in \mathcal{C}_+^Q$, which yields a large family of polynomial diffusions on E that we expect to be useful in applications.

Let $S_k, k = 1, \dots, d(d - 1)/2$, be a basis for the linear space of skew-symmetric $d \times d$ matrices. Using the skew-symmetry of the S_k together with the fact that

$Q^2 = \text{Id}$, it is easy to check that any map c of the form

$$c(x) = \sum_{k,\ell=1}^{d(d-1)/2} \gamma_{k\ell} Q S_k x x^\top S_\ell^\top Q, \tag{6.3}$$

where $\Gamma = (\gamma_{k\ell}) \in \mathbb{S}_+^{d(d-1)/2}$, lies in \mathcal{C}_+^Q . For any $c(x)$ of the form (6.3), condition (6.2) then becomes

$$\beta^\top Qx + x^\top \left(B^\top Q + \sum_{k,\ell} \gamma_{k\ell} S_k^\top Q S_\ell \right) x < 0 \quad \text{for all } x \in \{p = 0\}.$$

6.2 The product space $[0, 1]^m \times \mathbb{R}_+^n$

Consider the state space $E = [0, 1]^m \times \mathbb{R}_+^n$. Here $d = m + n$, and the generating family of polynomials can be taken to be

$$\mathcal{P} = \{x_i : i = 1, \dots, m + n; 1 - x_i : i = 1, \dots, m\}.$$

To simplify notation, introduce index sets $I = \{1, \dots, m\}$ and $J = \{m + 1, \dots, m + n\}$, and write x_I (resp. x_J) for the subvector of $x \in \mathbb{R}^d$ consisting of the components with indices in I (resp. J). Similarly, for a matrix $A \in \mathbb{R}^{d \times d}$, we write A_{II} , A_{IJ} , etc. for the submatrices with indicated row and column indices. The proof of the following proposition is given in Appendix H.

Proposition 6.4 *Conditions (G1) and (G2) hold for the above state space E . Moreover, the operator \mathcal{G} satisfies (A0)–(A2) if and only if*

(i) *the matrix a is given by*

$$\begin{aligned} a_{ii}(x) &= \gamma_i x_i (1 - x_i) && (i \in I), \\ a_{ij}(x) &= 0 && (i \in I, j \in I \cup J, i \neq j), \\ a_{jj}(x) &= \alpha_{jj} x_j^2 + x_j (\phi_j + \psi_{(j)}^\top x_I + \pi_{(j)}^\top x_J) && (j \in J), \\ a_{ij}(x) &= \alpha_{ij} x_i x_j && (i, j \in J, i \neq j) \end{aligned}$$

for some $\gamma \in \mathbb{R}_+^m$, some $\psi_{(j)} \in \mathbb{R}^m$, some $\pi_{(j)} \in \mathbb{R}_+^n$ with $\pi_{(j),j} = 0$, some $\phi \in \mathbb{R}^n$ with $\phi_j \geq (\psi_{(j)}^-)^\top \mathbf{1}$, and some $\alpha = (\alpha_{ij})_{i,j \in J} \in \mathbb{S}^n$ such that we have $\alpha + \text{Diag}(\Pi^\top x_J) \text{Diag}(x_J)^{-1} \in \mathbb{S}_+^n$ for all $x_J \in \mathbb{R}_{++}^n$, where $\Pi \in \mathbb{R}^{n \times n}$ is the matrix with columns $\pi_{(j)}$;

(ii) *the vector b is given by*

$$b(x) = \begin{pmatrix} \beta_I + B_{II} x_I \\ \beta_J + B_{JI} x_I + B_{JJ} x_J \end{pmatrix} \tag{6.4}$$

for some $\beta \in \mathbb{R}^d$ and $B \in \mathbb{R}^{d \times d}$ such that $(B_{i,I \setminus \{i\}}^-) \mathbf{1} < \beta_i < -B_{ii} - (B_{i,I \setminus \{i\}}^+) \mathbf{1}$ for all $i \in I$, $\beta_j > (B_{jI}^-) \mathbf{1}$ for all $j \in J$, and $B_{JJ} \in \mathbb{R}^{m \times m}$ has positive off-diagonal entries.

Remark 6.5 In the following two cases, we get uniqueness in law of E -valued solutions to (2.2); cf. Theorem 4.2. First, if $\alpha = 0$ and $\pi_{(j)} = 0$ for all j , then the linear growth condition (3.4) is satisfied and uniqueness follows by Theorem 4.2. Second, if $\psi_{(j)} = 0$ and $\pi_{(j)} = 0$ for all j and $\phi = 0$, then the submatrix $a_{JJ}(x)$ only depends on x_J and can be written $a_{JJ} = \sigma_{JJ}\sigma_{JJ}$, where $\sigma_{JJ}(x_J) = \text{Diag}(x_J)\alpha^{1/2}$ is Lipschitz-continuous. Since also X_I is an autonomous m -dimensional diffusion on $[0, 1]^m$, uniqueness follows from Theorem 4.4 in conjunction with Theorem 4.2. Note that X_I and X_J are coupled only through the drift in this case.

A natural next step is to consider the state space $[0, 1]^m \times \mathbb{R}_+^n \times \mathbb{R}^\ell$, $d = m + n + \ell$. In this case, one readily continues the above argument to deduce that the diffusion matrix is of the form

$$a(x) = \begin{pmatrix} a_{II}(x_I) & 0 & a_{IK}(x_I) \\ 0 & a_{JJ}(x_I, x_J) & a_{JK}(x_I, x_J) \\ a_{IK}(x_I)^\top & a_{JK}(x_I, x_J)^\top & a_{KK}(x_I, x_J, x_K) \end{pmatrix},$$

where $K = \{m + n + 1, \dots, d\}$, the matrices a_{II} and a_{JJ} are given by Proposition 6.4(i), we have $a_{IK}(x_I) = \text{Diag}(x_I)(\text{Id} - \text{Diag}(x_I))P$ for some $P \in \mathbb{R}^{m \times l}$ and $a_{JK}(x_I, x_J) = \text{Diag}(x_J)H(x_I, x_J)$ for some matrix H of polynomials in $\text{Pol}_1(E)$, and a_{KK} has component functions in $\text{Pol}_2(E)$. Regarding the drift vector $b = (b_I, b_J, b_K)$, the last part b_K is unrestricted within the class of affine functions of x , whereas (b_I, b_J) must satisfy Proposition 6.4(ii). With this structure, we have (A0)–(A2) if and only if $a \in \mathbb{S}_+^d$ on E . This of course imposes additional restrictions on P , H and a_{KK} . Stating these restrictions explicitly is cumbersome, and we refrain from doing so here.

6.3 The unit simplex

Let $d \geq 2$ and consider the unit simplex $E = \{x \in \mathbb{R}_+^d : x_1 + \dots + x_d = 1\}$. Here $\mathcal{P} = \{x_i : i = 1, \dots, d\}$ consists of the coordinate functions and \mathcal{Q} consists of the single polynomial $1 - \mathbf{1}^\top x$. The proof of the following proposition is given in Appendix I.

Proposition 6.6 *Conditions (G1) and (G2) hold for the above state space E . Moreover, the operator \mathcal{G} satisfies (A0)–(A2) if and only if*

(i) *the matrix a is given by*

$$a_{ii}(x) = \sum_{j \neq i} \alpha_{ij} x_i x_j,$$

$$a_{ij}(x) = -\alpha_{ij} x_i x_j \quad (i \neq j)$$

on E for some $\alpha_{ij} \in \mathbb{R}_+$ such that $\alpha_{ij} = \alpha_{ji}$ for all i, j ;

(ii) *the vector b is given by*

$$b(x) = \beta + Bx,$$

where $\beta \in \mathbb{R}^d$ and $B \in \mathbb{R}^{d \times d}$ satisfy $B^\top \mathbf{1} + (\beta^\top \mathbf{1})\mathbf{1} = 0$ and $\beta_i + B_{ji} > 0$ for all i and all $j \neq i$.

Remark 6.7 Since E is compact, Theorem 4.2 yields uniqueness in law for E -valued solutions to (2.2).

Remark 6.8 In the special case where $\alpha_{ij} = \sigma^2$ for some $\sigma > 0$ and all i, j , the diffusion matrix takes the form

$$\begin{aligned} a_{ii}(x) &= \sigma^2 x_i(1 - x_i), \\ a_{ij}(x) &= -\sigma^2 x_i x_j \quad (i \neq j). \end{aligned}$$

The resulting process is sometimes called a *multivariate Jacobi* process; see for instance Gouriéroux and Jasiak [27].

Remark 6.9 Alternatively, one can establish Proposition 6.6 by considering polynomial diffusions Y on the “solid” simplex $\{y \in \mathbb{R}_+^{d-1} : y_1 + \dots + y_{d-1} \leq 1\}$, and then set $X = (X_1, \dots, X_d) = (Y, 1 - Y_1 - \dots - Y_{d-1})$. In this case $\mathcal{Q} = \emptyset$, and it would be enough to invoke Lemma 5.4 rather than Lemma 5.5.

7 Polynomial diffusion models in finance

We now elaborate on various polynomial diffusion models in finance, following up on the introduction about (1.1). Let the state price density ζ be a positive semimartingale on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$. This induces an arbitrage-free financial market model on any finite time horizon T^* . Indeed, let S^1, \dots, S^m denote the price processes of m fundamental assets. According to (1.1), we have $\zeta_t S_t^i = \mathbb{E}[\zeta_{T^*} S_{T^*}^i | \mathcal{F}_t]$. Assuming that S^1 is positive, we choose it as numeraire. This implies an equivalent measure $\mathbb{Q}^1 \sim \mathbb{P}$ on \mathcal{F}_{T^*} by

$$\frac{d\mathbb{Q}^1}{d\mathbb{P}} = \frac{\zeta_{T^*} S_{T^*}^1}{\zeta_0 S_0^1}.$$

Discounted price processes $\frac{S_t^i}{S_t^1}$ are \mathbb{Q}^1 -martingales, because

$$\left. \frac{S_t^i}{S_t^1} \frac{d\mathbb{Q}^1}{d\mathbb{P}} \right|_{\mathcal{F}_t} = \frac{S_t^i}{S_t^1} \frac{\zeta_t S_t^1}{\zeta_0 S_0^1} = \frac{\zeta_t S_t^i}{\zeta_0 S_0^1}.$$

This implies that the market $\{S^1, \dots, S^m\}$ is arbitrage-free in the sense of no free lunch with vanishing risk; see Delbaen and Schachermayer [14].

Now let X be a polynomial diffusion on a state space $E \subseteq \mathbb{R}^d$. Fix $n \in \mathbb{N}$, and let $p \in \text{Pol}_n(E)$ be a positive polynomial on E with coordinate representation \vec{p} with respect to some basis $H(x) = (h_1(x), \dots, h_N(x))^\top$ for $\text{Pol}_n(E)$. The state price density is specified by $\zeta_t = e^{-\alpha t} p(X_t)$, where α is a real parameter. This setup yields an

arbitrage-free model for the term structure of interest rates. The time t price $P(t, T)$ of a zero coupon bond maturing at T , corresponding to $C_T = 1$ in (1.1), can now be computed explicitly, using Theorem 3.1, as

$$P(t, T) = e^{-\alpha(T-t)} \frac{\mathbb{E}[p(X_T) | \mathcal{F}_t]}{p(X_t)} = e^{-\alpha(T-t)} \frac{H(X_t)^\top e^{(T-t)G} \vec{p}}{H(X_t)^\top \vec{p}},$$

where $G \in \mathbb{R}^{N \times N}$ is the matrix representation of \mathcal{G} on $\text{Pol}_m(E)$. The short rate is obtained via the relation $r_t = -\partial_T \log P(t, T) |_{T=t}$, and is given by

$$r_t = \alpha - \frac{H(X_t)^\top G \vec{p}}{H(X_t)^\top \vec{p}}.$$

This expression clarifies the role of the parameter α adjusting the level of interest rates. Such models show great potential. The linear case with p of the form $p(x) = \phi + \psi^\top x$ has been studied in Filipović et al. [21], including an extensive empirical assessment. The parameter ψ is chosen such that E lies in the positive cone $\{x \in \mathbb{R}^d : \psi^\top x \geq 0\}$. A specific example is $E = \mathbb{R}_+^d$, as discussed in Sect. 6.2.

One attractive feature of the polynomial framework is that it yields efficient pricing formulae for options on coupon bearing bonds. This includes swaptions, which are among the most important interest rate options. The generic payoff of such an option at expiry date T is of the form

$$C_T = (c_0 + c_1 P(T, T_1) + \dots + c_m P(T, T_m))^+$$

for maturity dates $T < T_1 < \dots < T_m$ and deterministic coefficients c_0, \dots, c_m . Formula (1.1) for the time t price of this option boils down to computing the \mathcal{F}_t -conditional expectation of

$$\zeta_T C_T = \left(H(X_T)^\top \sum_{i=0}^m c_i e^{-\alpha T_i} e^{(T_i-T)G} \vec{p} \right)^+,$$

which is the positive part of a polynomial in X_T . Efficient methods involving the closed form \mathcal{F}_t -conditional moments of X_T are available; see Filipović et al. [20].

Polynomial diffusions can be employed in a similar way to build stochastic volatility models. We now interpret \mathbb{P} as risk-neutral measure, and specify the spot variance (squared volatility) of an underlying stock index by $v_t = p(X_t)$. The variance swap rate for period $[t, T]$ is then given in closed form by

$$\text{VS}(t, T) = \frac{1}{T-t} \mathbb{E} \left[\int_t^T v_s ds \mid \mathcal{F}_t \right] = \frac{1}{T-t} H(X_t)^\top \left(\int_t^T e^{(s-t)G} ds \right) \vec{p}.$$

Such models have been successfully employed in Filipović et al. [23] and Ackerer et al. [2]. Both papers consider the quadratic case, which falls into the setup of Sect. 6.1, with a quadric state space $E = \{x \in \mathbb{R}^d : x^\top Qx \leq 1\}$ and spot variance $v_t = p(X_t)$ for a polynomial p of the form $p(x) = \phi + 1 - x^\top Qx$, where $\phi \geq 0$ denotes the minimal spot variance. While Filipović et al. [23] study unbounded state spaces, Ackerer et al. [2] focus on the compact case, where Q is

positive definite. They derive analytic option pricing formulae in terms of Hermite polynomials for European call and put options on an asset with diffusive price process $dS_t = S_t r dt + S_t \sqrt{v_t} dW_t^*$, where r denotes the constant short rate and W^* is a Brownian motion, which is possibly correlated with W in (2.2).

An application of the unit simplex in Sect. 6.3 is obtained as follows. Consider a stock index, such as the S&P 500, whose price process is given by a semimartingale Z . As above, we interpret \mathbb{P} as risk-neutral measure and assume a constant short rate r such that $(e^{-rt} Z_t)$ is a martingale. Let d be the number of constituent stocks, and let X be a polynomial diffusion on $E = \{x \in \mathbb{R}_+^d : x_1 + \dots + x_d = 1\}$ which is independent of Z . We fix a finite time horizon T^* and define the E -valued martingale, for $t \leq T^*$,

$$Y_t = \mathbb{E}[X_{T^*} | \mathcal{F}_t].$$

Since X is polynomial, Y_t is a first degree polynomial in X_t whose coefficients can be determined by an application of Theorem 3.1. Specifically, with β and B being the drift parameters of X as given in Proposition 6.6, one finds

$$Y_t = \Phi(T^* - t) + \Psi(T^* - t)X_t \quad \text{with } \Phi(\tau) = \int_0^\tau e^{sB} \beta ds \text{ and } \Psi(\tau) = e^{\tau B}.$$

We now define the constituent stocks' price processes $S^i = Y^i Z, i = 1, \dots, d$, such that $S^1 + \dots + S^d = Z$. Assume that the price of the European call option on the index with maturity T and strike K is given in closed form, $C(T, K)$, for some analytic function C . The price of the call option on stock i with maturity T and strike K is then given by

$$C_i(T, K) = \mathbb{E}[Y_T^i C(T, K/Y_T^i)].$$

This price can be efficiently computed in three steps. First, compute $\xi C(T, K/\xi)$ for a finite set of grid points $\xi \in [0, 1]$. Second, apply some polynomial interpolation scheme, for example using Chebyshev polynomials, to obtain a polynomial approximation of degree n , say $q(T, K, \xi)$, of $\xi C(T, K/\xi)$ in $\xi \in [0, 1]$. Third, approximate the option price $C_i(T, K)$ by $H(X_0)^\top e^{TG} \tilde{p}_i(T, K)$, where $\tilde{p}_i(T, K)$ is the coordinate representation of the polynomial $p(x) = q(T, K, \Phi_i(T^* - T) + (\Psi(T^* - T)x)_i)$ in x with respect to some appropriately chosen basis of polynomials for $\text{Pol}_n(E)$. Extensions to basket and spread options on the stocks S^1, \dots, S^d are straightforward. This is work in progress.

An application of polynomial diffusions on a compact state space to credit risk is given in Ackerer and Filipović [1].

Acknowledgements The authors wish to thank Damien Ackerer, Peter Glynn, Kostas Kardaras, Guillermo Mantilla-Soler, Sergio Pulido, Mykhaylo Shkolnikov, Jordan Stoyanov and Josef Teichmann for useful comments and stimulating discussions. Thanks are also due to the referees, co-editor, and editor for their valuable remarks.

Appendix A: Nonnegative Itô processes

The following auxiliary result forms the basis of the proof of Theorem 5.3. It gives necessary and sufficient conditions for nonnegativity of certain Itô processes.

Lemma A.1 *Let Z be a continuous semimartingale of the form*

$$Z_t = Z_0 + \int_0^t \mu_s \, ds + \int_0^t \nu_s \, dB_s,$$

where $Z_0 \geq 0$, μ and ν are continuous processes, and B is a Brownian motion. Let L^0 be the local time of Z at level zero.

- (i) If $\mu > 0$ on $\{Z = 0\}$ and $L^0 = 0$, then $Z \geq 0$ and $\int_0^t \mathbf{1}_{\{Z_s=0\}} \, ds = 0$.
- (ii) If $Z \geq 0$, then on $\{Z = 0\}$, we have $\mu \geq 0$ and $\nu = 0$.

Proof After stopping we may assume that Z_t , $\int_0^t \mu_s \, ds$ and $\int_0^t \nu_s \, dB_s$ are uniformly bounded. This is done throughout the proof.

We first prove (i). By [41, Theorem VI.1.7] and using that $\mu > 0$ on $\{Z = 0\}$ and $L^0 = 0$, we obtain $0 = L_t^0 = L_t^{0-} + 2 \int_0^t \mathbf{1}_{\{Z_s=0\}} \mu_s \, ds \geq 0$. In particular, $\int_0^t \mathbf{1}_{\{Z_s=0\}} \, ds = 0$, as claimed. Furthermore, Tanaka’s formula [41, Theorem VI.1.2] yields

$$Z_t^- = - \int_0^t \mathbf{1}_{\{Z_s \leq 0\}} \, dZ_s - \frac{1}{2} L_t^0 = - \int_0^t \mathbf{1}_{\{Z_s \leq 0\}} \mu_s \, ds - \int_0^t \mathbf{1}_{\{Z_s \leq 0\}} \nu_s \, dB_s. \tag{A.1}$$

Define $\rho = \inf\{t \geq 0 : Z_t < 0\}$ and $\tau = \inf\{t \geq \rho : \mu_t = 0\} \wedge (\rho + 1)$. Using that $Z^- = 0$ on $\{\rho = \infty\}$ as well as dominated convergence, we obtain

$$\mathbb{E}[Z_{\tau \wedge n}^-] = \mathbb{E}[Z_{\tau \wedge n}^- \mathbf{1}_{\{\rho < \infty\}}] \longrightarrow \mathbb{E}[Z_\tau^- \mathbf{1}_{\{\rho < \infty\}}] \quad (n \rightarrow \infty).$$

Here Z_τ is well defined on $\{\rho < \infty\}$ since $\tau < \infty$ on this set. On the other hand, by (A.1), the fact that $\int_0^t \mathbf{1}_{\{Z_s \leq 0\}} \mu_s \, ds = \int_0^t \mathbf{1}_{\{Z_s=0\}} \mu_s \, ds = 0$ on $\{\rho = \infty\}$ and monotone convergence, we get

$$\begin{aligned} \mathbb{E}[Z_{\tau \wedge n}^-] &= \mathbb{E}\left[- \int_0^{\tau \wedge n} \mathbf{1}_{\{Z_s \leq 0\}} \mu_s \, ds\right] = \mathbb{E}\left[- \int_0^{\tau \wedge n} \mathbf{1}_{\{Z_s \leq 0\}} \mu_s \, ds \mathbf{1}_{\{\rho < \infty\}}\right] \\ &\longrightarrow \mathbb{E}\left[- \int_0^\tau \mathbf{1}_{\{Z_s \leq 0\}} \mu_s \, ds \mathbf{1}_{\{\rho < \infty\}}\right] \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Consequently,

$$\mathbb{E}[Z_\tau^- \mathbf{1}_{\{\rho < \infty\}}] = \mathbb{E}\left[- \int_0^\tau \mathbf{1}_{\{Z_s \leq 0\}} \mu_s \, ds \mathbf{1}_{\{\rho < \infty\}}\right]. \tag{A.2}$$

The following hold on $\{\rho < \infty\}$: $\tau > \rho$; $Z_t \geq 0$ on $[0, \rho]$; $\mu_t > 0$ on $[\rho, \tau)$; and $Z_t < 0$ on some nonempty open subset of (ρ, τ) . Therefore, the random variable inside the expectation on the right-hand side of (A.2) is strictly negative on $\{\rho < \infty\}$. The left-hand side, however, is nonnegative; so we deduce $\mathbb{P}[\rho < \infty] = 0$. Part (i) is proved.

The proof of Part (ii) involves the same ideas as used for instance in Spreij and Veerman [44, Proposition 3.1]. We first assume $Z_0 = 0$ and prove $\mu_0 \geq 0$ and $\nu_0 = 0$. Assume for contradiction that $\mathbb{P}[\mu_0 < 0] > 0$, and define $\tau = \inf\{t \geq 0 : \mu_t \geq 0\} \wedge 1$.

Then $0 \leq \mathbb{E}[Z_\tau] = \mathbb{E}[\int_0^\tau \mu_s ds] < 0$, a contradiction, whence $\mu_0 \geq 0$ as desired. Next, pick any $\phi \in \mathbb{R}$ and consider an equivalent measure $d\mathbb{Q} = \mathcal{E}(-\phi B)_1 d\mathbb{P}$. Then $B_t^{\mathbb{Q}} = B_t + \phi t$ is a \mathbb{Q} -Brownian motion on $[0, 1]$, and we have

$$Z_t = \int_0^t (\mu_s - \phi v_s) ds + \int_0^t v_s dB_s^{\mathbb{Q}}.$$

Pick any $\varepsilon > 0$ and define $\sigma = \inf\{t \geq 0 : |v_t| \leq \varepsilon\} \wedge 1$. The first part of the proof applied to the stopped process Z^σ under \mathbb{Q} yields $(\mu_0 - \phi v_0)\mathbf{1}_{\{\sigma > 0\}} \geq 0$ for all $\phi \in \mathbb{R}$. But this forces $\sigma = 0$ and hence $|v_0| \leq \varepsilon$. Since $\varepsilon > 0$ was arbitrary, we get $v_0 = 0$ as desired.

Now consider any stopping time ρ such that $Z_\rho = 0$ on $\{\rho < \infty\}$. Applying the result we have already proved to the process $(Z_{\rho+t}\mathbf{1}_{\{\rho < \infty\}})_{t \geq 0}$ with filtration $(\mathcal{F}_{\rho+t} \cap \{\rho < \infty\})_{t \geq 0}$ then yields $\mu_\rho \geq 0$ and $v_\rho = 0$ on $\{\rho < \infty\}$. Finally, let $\{\rho_n : n \in \mathbb{N}\}$ be a countable collection of such stopping times that are dense in $\{t : Z_t = 0\}$. Applying the above result to each ρ_n and using the continuity of μ and v , we obtain (ii). □

The following two examples show that the assumptions of Lemma A.1 are tight in the sense that the gap between (i) and (ii) cannot be closed.

Example A.2 The strict inequality appearing in Lemma A.1(i) cannot be relaxed to a weak inequality: just consider the deterministic process $Z_t = (1 - t)^3$.

Example A.3 The assumption of vanishing local time at zero in Lemma A.1(i) cannot be replaced by the zero volatility condition $v = 0$ on $\{Z = 0\}$, even if the strictly positive drift condition is retained. This is demonstrated by a construction that is closely related to the so-called Girsanov SDE; see Rogers and Williams [42, Sect. V.26]. Let Y be a one-dimensional Brownian motion, and define $\rho(y) = |y|^{-2\alpha} \vee 1$ for some $0 < \alpha < 1/4$. The occupation density formula implies that

$$\int_0^t \rho(Y_s)^2 ds = \int_{-\infty}^\infty (|y|^{-4\alpha} \vee 1) L_t^y(Y) dy < \infty$$

for all $t \geq 0$; so we may define a positive local martingale by

$$R_t = \exp\left(\int_0^t \rho(Y_s) dY_s - \frac{1}{2} \int_0^t \rho(Y_s)^2 ds\right).$$

Let τ be a strictly positive stopping time such that the stopped process R^τ is a uniformly integrable martingale. Then define the equivalent probability measure $d\mathbb{Q} = R_\tau d\mathbb{P}$, under which the process $B_t = Y_t - \int_0^{t \wedge \tau} \rho(Y_s) ds$ is a Brownian motion. We now change time via

$$\varphi_t = \int_0^t \rho(Y_s) ds, \quad A_u = \inf\{t \geq 0 : \varphi_t > u\},$$

and define $Z_u = Y_{A_u}$. This process satisfies $Z_u = B_{A_u} + u \wedge \sigma$, where $\sigma = \varphi_\tau$. Define then $\beta_u = \int_0^u \rho(Z_v)^{1/2} dB_{A_v}$, which is a Brownian motion because we have $\langle \beta, \beta \rangle_u = \int_0^u \rho(Z_v) dA_v = u$. This finally gives

$$Z_u = \int_0^u (|Z_v|^\alpha \wedge 1) d\beta_v + u \wedge \sigma.$$

This process starts at zero, has zero volatility whenever $Z_t = 0$, and strictly positive drift prior to the stopping time σ , which is strictly positive. Nonetheless, its sign changes infinitely often on any time interval $[0, t)$ since it is a time-changed Brownian motion viewed under an equivalent measure.

Appendix B: Proof of Theorem 3.1

We first establish a lemma.

Lemma B.1 *For any $k \in \mathbb{N}$ such that $\mathbb{E}[\|X_0\|^{2k}] < \infty$, there is a constant C such that*

$$\mathbb{E}[1 + \|X_t\|^{2k} \mid \mathcal{F}_0] \leq (1 + \|X_0\|^{2k})e^{Ct}, \quad t \geq 0.$$

Proof This is done as in the proof of Theorem 2.10 in Cuchiero et al. [10] via Gronwall’s inequality. Specifically, let $f \in \text{Pol}_{2k}(E)$ be given by $f(x) = 1 + \|x\|^{2k}$, and note that the polynomial property implies that there exists a constant C such that $|\mathcal{G}f(x)| \leq Cf(x)$ for all $x \in E$. For each m , let τ_m be the first exit time of X from the ball $\{x \in E : \|x\| < m\}$. We can always choose a continuous version of $t \mapsto \mathbb{E}[f(X_{t \wedge \tau_m}) \mid \mathcal{F}_0]$, so let us fix such a version. Then by Itô’s formula and the martingale property of $\int_0^{t \wedge \tau_m} \nabla f(X_s)^\top \sigma(X_s) dW_s$,

$$\begin{aligned} \mathbb{E}[f(X_{t \wedge \tau_m}) \mid \mathcal{F}_0] &= f(X_0) + \mathbb{E}\left[\int_0^{t \wedge \tau_m} \mathcal{G}f(X_s) ds \mid \mathcal{F}_0\right] \\ &\leq f(X_0) + C\mathbb{E}\left[\int_0^{t \wedge \tau_m} f(X_s) ds \mid \mathcal{F}_0\right] \\ &\leq f(X_0) + C \int_0^t \mathbb{E}[f(X_{s \wedge \tau_m}) \mid \mathcal{F}_0] ds. \end{aligned}$$

Gronwall’s inequality now yields $\mathbb{E}[f(X_{t \wedge \tau_m}) \mid \mathcal{F}_0] \leq f(X_0)e^{Ct}$. Sending m to infinity and applying Fatou’s lemma gives the result. \square

We can now prove Theorem 3.1. For any $p \in \text{Pol}_n(E)$, Itô’s formula yields

$$p(X_u) = p(X_t) + \int_t^u \mathcal{G}p(X_s) ds + \int_t^u \nabla p(X_s)^\top \sigma(X_s) dW_s.$$

The quadratic variation of the right-hand side satisfies

$$\int_0^T \nabla p^\top a \nabla p(X_s) ds \leq C \int_0^T (1 + \|X_s\|^{2n}) ds$$

for some constant C . This right-hand side has finite expectation by Lemma B.1, so the stochastic integral above is a martingale. Let $\vec{p} \in \mathbb{R}^N$ be the coordinate representation of p . Then (3.1) and (3.2) in conjunction with the linearity of the expectation and integration operators yield

$$\begin{aligned} \vec{p}^\top \mathbb{E}[H(X_u) | \mathcal{F}_t] &= \mathbb{E}[p(X_u) | \mathcal{F}_t] = p(X_t) + \mathbb{E}\left[\int_t^u \mathcal{G} p(X_s) ds \mid \mathcal{F}_t\right] \\ &= \vec{p}^\top H(X_t) + (G\vec{p})^\top \mathbb{E}\left[\int_t^u H(X_s) ds \mid \mathcal{F}_t\right]. \end{aligned}$$

Fubini’s theorem, justified by Lemma B.1, yields

$$\vec{p}^\top F(u) = \vec{p}^\top H(X_t) + \vec{p}^\top G^\top \int_t^u F(s) ds, \quad t \leq u \leq T,$$

where we define $F(u) = \mathbb{E}[H(X_u) | \mathcal{F}_t]$. By choosing unit vectors for \vec{p} , this gives a system of linear integral equations for $F(u)$, whose unique solution is given by $F(u) = e^{(u-t)G^\top} H(X_t)$. Hence

$$\mathbb{E}[p(X_T) | \mathcal{F}_t] = F(T)^\top \vec{p} = H(X_t)^\top e^{(T-t)G} \vec{p},$$

as claimed. This completes the proof of the theorem. □

Appendix C: Proof of Theorem 3.3

Theorem 3.3 is an immediate corollary of the following result.

Lemma C.1 *Consider the d -dimensional Itô process X with representation*

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

where σ satisfies a square-root growth condition

$$\|\sigma(X_t)\|^2 \leq C(1 + \|X_t\|) \quad \text{for all } t \geq 0 \tag{C.1}$$

for some constant C . If

$$\mathbb{E}[e^{\delta \|X_0\|}] < \infty \quad \text{for some } \delta > 0, \tag{C.2}$$

then for each $T \geq 0$, there exists $\varepsilon > 0$ with

$$\mathbb{E}[e^{\varepsilon \|X_T\|}] < \infty. \tag{C.3}$$

Proof Fix $T \geq 0$. Variation of constants lets us rewrite $X_t = A_t + e^{-\beta(T-t)}Y_t$ with

$$A_t = e^{\beta t}X_0 + \int_0^t e^{\beta(t-s)}b ds$$

and

$$Y_t = \int_0^t e^{\beta(T-s)}\sigma(X_s)dW_s = \int_0^t \sigma_s^Y dW_s,$$

where we write $\sigma_t^Y = e^{\beta(T-t)}\sigma(A_t + e^{-\beta(T-t)}Y_t)$. By (C.1), the dispersion process σ^Y satisfies

$$\|\sigma_t^Y\|^2 \leq C_Y(1 + \|Y_t\|) \tag{C.4}$$

for some constant C_Y .

Now let $f(y)$ be a real-valued and positive smooth function on \mathbb{R}^d satisfying $f(y) = \sqrt{1 + \|y\|}$ for $\|y\| > 1$. Some differential calculus gives, for $y \neq 0$,

$$\nabla\|y\| = \frac{y}{\|y\|} \quad \text{and} \quad \frac{\partial^2\|y\|}{\partial y_i \partial y_j} = \begin{cases} \frac{1}{\|y\|} - \frac{1}{2} \frac{y_i^2}{\|y\|^3}, & i = j, \\ -\frac{1}{2} \frac{y_i y_j}{\|y\|^3}, & i \neq j. \end{cases}$$

Hence

$$\nabla f(y) = \frac{1}{2\sqrt{1 + \|y\|}} \frac{y}{\|y\|}$$

and

$$\frac{\partial^2 f(y)}{\partial y_i \partial y_j} = -\frac{1}{4\sqrt{1 + \|y\|}^3} \frac{y_i}{\|y\|} \frac{y_j}{\|y\|} + \frac{1}{2\sqrt{1 + \|y\|}} \times \begin{cases} \frac{1}{\|y\|} - \frac{1}{2} \frac{y_i^2}{\|y\|^3}, & i = j \\ -\frac{1}{2} \frac{y_i y_j}{\|y\|^3}, & i \neq j \end{cases}$$

for $\|y\| > 1$, while the first and second order derivatives of $f(y)$ are uniformly bounded for $\|y\| \leq 1$. Itô’s formula for $Z_t = f(Y_t)$ gives

$$dZ_t = \mu_t^Z dt + \sigma_t^Z dW_t$$

with drift and dispersion processes

$$\mu_t^Z = \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 f(Y_t)}{\partial y_i \partial y_j} (\sigma_t^Y \sigma_t^{Y\top})_{ij}, \quad \sigma_t^Z = \nabla f(Y_t)^\top \sigma_t^Y.$$

In view of (C.4) and the above expressions for $\nabla f(y)$ and $\frac{\partial^2 f(y)}{\partial y_i \partial y_j}$, these are bounded,

$$\mu_t^Z \leq m \quad \text{and} \quad \|\sigma_t^Z\| \leq \rho,$$

for some constants m and ρ . Hajek [28, Theorem 1.3] now implies that

$$\mathbb{E}[\Phi(Z_T)] \leq \mathbb{E}[\Phi(V)]$$

for any nondecreasing convex function Φ on \mathbb{R} , where V is a Gaussian random variable with mean $f(0) + mT$ and variance $\rho^2 T$. Hence, for any $0 < \varepsilon' < 1/(2\rho^2 T)$, we have $\mathbb{E}[e^{\varepsilon' V^2}] < \infty$. We now let Φ be a nondecreasing convex function on \mathbb{R} with $\Phi(z) = e^{\varepsilon' z^2}$ for $z \geq 0$. Noting that Z_T is positive, we obtain $\mathbb{E}[e^{\varepsilon' Z_T^2}] < \infty$. As $f^2(y) = 1 + \|y\|$ for $\|y\| > 1$, this implies $\mathbb{E}[e^{\varepsilon' \|Y_T\|}] < \infty$. Combining this with the fact that $\|X_T\| \leq \|A_T\| + \|Y_T\|$ and (C.2), we obtain using Hölder’s inequality the existence of some $\varepsilon > 0$ with (C.3). \square

Appendix D: Proof of Theorem 4.4

We first provide a lemma.

Lemma D.1 *Assume uniqueness in law holds for E_Y -valued solutions to (4.1). Let Y^1, Y^2 be two E_Y -valued solutions to (4.1) with driving Brownian motions W^1, W^2 and with $Y_0^1 = Y_0^2 = y$ for some $y \in E_Y$. Then (Y^1, W^1) and (Y^2, W^2) have the same law.*

Proof Consider the equation

$$dY_t = \widehat{b}_Y(Y_t) dt + \widehat{\sigma}_Y(Y_t) dW_t,$$

where $\widehat{b}_Y(y) = b_Y(y)\mathbf{1}_{E_Y}(y)$ and $\widehat{\sigma}_Y(y) = \sigma_Y(y)\mathbf{1}_{E_Y}(y)$. Since E_Y is closed, any solution Y to this equation with $Y_0 \in E_Y$ must remain inside E_Y . To see this, let $\tau = \inf\{t : Y_t \notin E_Y\}$. Then there exists $\varepsilon > 0$, depending on ω , such that $Y_t \notin E_Y$ for all $\tau < t < \tau + \varepsilon$. However, since \widehat{b}_Y and $\widehat{\sigma}_Y$ vanish outside E_Y , Y_t is constant on $(\tau, \tau + \varepsilon)$. Since E_Y is closed this is only possible if $\tau = \infty$.

The hypothesis of the lemma now implies that uniqueness in law for \mathbb{R}^d -valued solutions holds for $dY_t = \widehat{b}_Y(Y_t) dt + \widehat{\sigma}_Y(Y_t) dW_t$. Since $(Y^i, W^i), i = 1, 2$, are two solutions with $Y_0^1 = Y_0^2 = y$, Cherny [8, Theorem 3.1] shows that (W^1, Y^1) and (W^2, Y^2) have the same law. \square

The proof of Theorem 4.4 follows along the lines of the proof of the Yamada–Watanabe theorem that pathwise uniqueness implies uniqueness in law; see Rogers and Williams [42, Theorem V.17.1]. Let $(W^i, Y^i, Z^i), i = 1, 2$, be E -valued weak solutions to (4.1), (4.2) starting from $(y_0, z_0) \in E \subseteq \mathbb{R}^m \times \mathbb{R}^n$. We need to show that (Y^1, Z^1) and (Y^2, Z^2) have the same law. Since uniqueness in law holds for E_Y -valued solutions to (4.1), Lemma D.1 implies that (W^1, Y^1) and (W^2, Y^2) have the same law, which we denote by $\pi(dw, dy)$. Let $Q^i(dz; w, y), i = 1, 2$, denote a regular conditional distribution of Z^i given (W^i, Y^i) . We equip the path space $C(\mathbb{R}_+, \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n)$ with the probability measure

$$\overline{\mathbb{P}}(dw, dy, dz, dz') = \pi(dw, dy)Q^1(dz; w, y)Q^2(dz'; w, y).$$

Let (W, Y, Z, Z') denote the coordinate process on $C(\mathbb{R}_+, \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n)$. Then the law under $\overline{\mathbb{P}}$ of (W, Y, Z) equals the law of (W^1, Y^1, Z^1) , and the law under $\overline{\mathbb{P}}$ of (W, Y, Z') equals the law of (W^2, Y^2, Z^2) . By well-known arguments, see for

instance Rogers and Williams [42, Lemma V.10.1 and Theorems V.10.4 and V.17.1], it follows that

$$\begin{aligned}
 Y_t &= y_0 + \int_0^t b_Y(Y_s) \, ds + \int_0^t \sigma_Y(Y_s) \, dW_s, \\
 Z_t &= z_0 + \int_0^t b_Z(Y_s, Z_s) \, ds + \int_0^t \sigma_Z(Y_s, Z_s) \, dW_s, \\
 Z'_t &= z_0 + \int_0^t b_Z(Y_s, Z'_s) \, ds + \int_0^t \sigma_Z(Y_s, Z'_s) \, dW_s.
 \end{aligned}$$

By localization, we may assume that b_Z and σ_Z are Lipschitz in z , uniformly in y . A standard argument based on the BDG inequalities and Jensen’s inequality (see Rogers and Williams [42, Corollary V.11.7]) together with Gronwall’s inequality yields $\mathbb{P}[Z' = Z] = 1$. Hence

$$\text{Law}(Y^1, Z^1) = \text{Law}(Y, Z) = \text{Law}(Y, Z') = \text{Law}(Y^2, Z^2),$$

as was to be shown. □

Remark D.2 Theorem 4.4 carries over, and its proof literally goes through, to the case where (Y, Z) is an arbitrary E -valued diffusion that solves (4.1), (4.2) and where uniqueness in law for E_Y -valued solutions to (4.1) holds, provided (4.3) is replaced by the assumption that both b_Z and σ_Z are locally Lipschitz in z , locally in y , on E . That is, for each compact subset $K \subseteq E$, there exists a constant κ such that for all $(y, z, y', z') \in K \times K$,

$$\|b_Z(y, z) - b_Z(y', z')\| + \|\sigma_Z(y, z) - \sigma_Z(y', z')\| \leq \kappa \|z - z'\|.$$

Appendix E: Proof of Theorem 5.3

The proof of Theorem 5.3 consists of two main parts. First, we construct coefficients $\widehat{a} = \widehat{\sigma}\widehat{\sigma}^\top$ and \widehat{b} that coincide with a and b on E , such that a local solution to (2.2), with b and σ replaced by \widehat{b} and $\widehat{\sigma}$, can be obtained with values in a neighborhood of E in M . This relies on (G1) and (A2), and occupies this section up to and including Lemma E.4. Second, we complete the proof by showing that this solution in fact stays inside E and spends zero time in the sets $\{p = 0\}$, $p \in \mathcal{P}$. This relies on (G2) and (A1).

Let $\pi : \mathbb{S}^d \rightarrow \mathbb{S}_+^d$ be the Euclidean metric projection onto the positive semidefinite cone. It has the following well-known property.

Lemma E.1 *For any symmetric matrix $A \in \mathbb{S}^d$ with the spectral decomposition $A = S\Lambda S^\top$, we have $\pi(A) = S\Lambda^+ S^\top$, where Λ^+ is the element-wise positive part of Λ .*

Proof This result follows from the fact that the map $\lambda : \mathbb{S}^d \rightarrow \mathbb{R}^d$ taking a symmetric matrix to its ordered eigenvalues is 1-Lipschitz; see Horn and Johnson [30, Theorem 7.4.51]. Indeed, for any $B \in \mathbb{S}_+^d$, we have

$$\|A - SA^+S^\top\| = \|\lambda(A) - \lambda(A)^+\| \leq \|\lambda(A) - \lambda(B)\| \leq \|A - B\|.$$

Here the first inequality uses that the projection of an ordered vector $x \in \mathbb{R}^d$ onto the set of ordered vectors with nonnegative entries is simply x^+ . \square

We use the projection π to modify the given coefficients a and b outside E in order to obtain candidate coefficients for the stochastic differential equation (2.2). The diffusion coefficients are defined by

$$\widehat{a}(x) = \pi \circ a(x), \quad \widehat{\sigma}(x) = \widehat{a}(x)^{1/2}.$$

In order to construct the drift coefficient \widehat{b} , we need the following lemma.

Lemma E.2 *There exists a continuous map $\widehat{b} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ with $\widehat{b} = b$ on E and such that the operator $\widehat{\mathcal{G}}$ given by*

$$\widehat{\mathcal{G}}f = \frac{1}{2} \text{Tr}(\widehat{a}\nabla^2 f) + \widehat{b}^\top \nabla f$$

satisfies $\widehat{\mathcal{G}}f = \mathcal{G}f$ on E and $\widehat{\mathcal{G}}q = 0$ on M for all $q \in \mathcal{Q}$.

Proof We first prove that there exists a continuous map $c : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that

$$c = 0 \text{ on } E \quad \text{and} \quad \nabla q^\top c = -\frac{1}{2} \text{Tr}((\widehat{a} - a)\nabla^2 q) \text{ on } M, \text{ for all } q \in \mathcal{Q}. \quad (\text{E.1})$$

Indeed, let $a = SA^\top S$ be the spectral decomposition of a , so that the columns S_i of S constitute an orthonormal basis of eigenvectors of a and the diagonal elements λ_i of A are the corresponding eigenvalues. These quantities depend on x in a possibly discontinuous way. For each $q \in \mathcal{Q}$,

$$\text{Tr}((\widehat{a} - a)\nabla^2 q) = \text{Tr}(SA^-S^\top \nabla^2 q) = \sum_{i=1}^d \lambda_i^- S_i^\top \nabla^2 q S_i. \quad (\text{E.2})$$

Consider now any fixed $x \in M$. For each i such that $\lambda_i(x)^- \neq 0$, $S_i(x)$ lies in the tangent space of M at x . Thus we may find a smooth path $\gamma_i : (-1, 1) \rightarrow M$ such that $\gamma_i(0) = x$ and $\gamma_i'(0) = S_i(x)$. For any $q \in \mathcal{Q}$, we have $q = 0$ on M by definition, whence

$$0 = \frac{d^2}{ds^2}(q \circ \gamma_i)(0) = \text{Tr}(\nabla^2 q(x)\gamma_i'(0)\gamma_i'(0)^\top) + \nabla q(x)^\top \gamma_i''(0),$$

or equivalently, $S_i(x)^\top \nabla^2 q(x) S_i(x) = -\nabla q(x)^\top \gamma_i''(0)$. In view of (E.2), this yields

$$\text{Tr}((\widehat{a}(x) - a(x))\nabla^2 q(x)) = -\nabla q(x)^\top \sum_{i=1}^d \lambda_i(x)^- \gamma_i''(0) \quad \text{for all } q \in \mathcal{Q}.$$

Let q_1, \dots, q_m be an enumeration of the elements of \mathcal{Q} , and write the above equation in vector form as

$$\begin{pmatrix} \text{Tr}((\widehat{a}(x) - a(x))\nabla^2 q_1(x)) \\ \vdots \\ \text{Tr}((\widehat{a}(x) - a(x))\nabla^2 q_m(x)) \end{pmatrix} = - \begin{pmatrix} \nabla q_1(x)^\top \\ \vdots \\ \nabla q_m(x)^\top \end{pmatrix} \sum_{i=1}^d \lambda_i(x)^{-1} \gamma'_i(0).$$

The left-hand side thus lies in the range of $[\nabla q_1(x) \cdots \nabla q_m(x)]^\top$ for each $x \in M$. Since linear independence is an open condition, (G1) implies that the latter matrix has full rank for all x in a whole neighborhood U of M . It thus has a Moore–Penrose inverse which is a continuous function of x ; see Penrose [39, page 408]. The desired map c is now obtained on U by

$$c(x) = -\frac{1}{2} \begin{pmatrix} \nabla q_1(x)^\top \\ \vdots \\ \nabla q_m(x)^\top \end{pmatrix}^{-1} \begin{pmatrix} \text{Tr}((\widehat{a}(x) - a(x))\nabla^2 q_1(x)) \\ \vdots \\ \text{Tr}((\widehat{a}(x) - a(x))\nabla^2 q_m(x)) \end{pmatrix},$$

where the Moore–Penrose inverse is understood. Finally, after shrinking U while maintaining $M \subseteq U$, c is continuous on the closure \overline{U} , and can then be extended to a continuous map on \mathbb{R}^d by the Tietze extension theorem; see Willard [47, Theorem 15.8]. This proves (E.1).

The extended drift coefficient is now defined by $\widehat{b} = b + c$, and the operator $\widehat{\mathcal{G}}$ by

$$\widehat{\mathcal{G}}f = \frac{1}{2} \text{Tr}(\widehat{a}\nabla^2 f) + \widehat{b}^\top \nabla f.$$

In view of (E.1), it satisfies $\widehat{\mathcal{G}}f = \mathcal{G}f$ on E and

$$\widehat{\mathcal{G}}q = \mathcal{G}q + \frac{1}{2} \text{Tr}((\widehat{a} - a)\nabla^2 q) + c^\top \nabla q = 0$$

on M for all $q \in \mathcal{Q}$, as desired. □

We now define the set

$$E_0 = M \cap \{\|\widehat{b} - b\| < 1\}.$$

Note that $E \subseteq E_0$ since $\widehat{b} = b$ on E . Furthermore, the linear growth condition

$$\|\widehat{a}(x)\|^{1/2} + \|\widehat{b}(x)\| \leq \|a(x)\|^{1/2} + \|b(x)\| + 1 \leq C(1 + \|x\|), \quad x \in E_0, \quad (\text{E.3})$$

is satisfied for some constant C . This uses that the component functions of a and b lie in $\text{Pol}_2(\mathbb{R}^d)$ and $\text{Pol}_1(\mathbb{R}^d)$, respectively.

An E_0 -valued local solution to (2.2), with b and σ replaced by \widehat{b} and $\widehat{\sigma}$, can now be constructed by solving the martingale problem for the operator $\widehat{\mathcal{G}}$ and state space E_0 . We first prove an auxiliary lemma.

Lemma E.3 *Let $f \in C^\infty(\mathbb{R}^d)$ and assume the support K of f satisfies $K \cap M \subseteq E_0$. Let x_0 be a maximizer of f over E_0 . Then $\widehat{\mathcal{G}}f(x_0) \leq 0$.*

Proof Let $\gamma : (-1, 1) \rightarrow M$ be any smooth curve in M with $\gamma(0) = x_0$. Optimality of x_0 and the chain rule yield

$$0 = \frac{d}{ds}(f \circ \gamma)(0) = \nabla f(x_0)^\top \gamma'(0),$$

from which it follows that $\nabla f(x_0)$ is orthogonal to the tangent space of M at x_0 . Thus

$$\nabla f(x_0) = \sum_{q \in \mathcal{Q}} c_q \nabla q(x_0) \tag{E.4}$$

for some coefficients c_q . Next, differentiating once more yields

$$0 \geq \frac{d^2}{ds^2}(f \circ \gamma)(0) = \text{Tr}(\nabla^2 f(x_0) \gamma'(0) \gamma'(0)^\top) + \nabla f(x_0)^\top \gamma''(0).$$

Similarly, for any $q \in \mathcal{Q}$,

$$0 = \frac{d^2}{ds^2}(q \circ \gamma)(0) = \text{Tr}(\nabla^2 q(x_0) \gamma'(0) \gamma'(0)^\top) + \nabla q(x_0)^\top \gamma''(0).$$

In view of (E.4), this implies

$$\text{Tr}\left(\left(\nabla^2 f(x_0) - \sum_{q \in \mathcal{Q}} c_q \nabla^2 q(x_0)\right) \gamma'(0) \gamma'(0)^\top\right) \leq 0. \tag{E.5}$$

Observe that Lemma E.1 implies that $\ker A \subseteq \ker \pi(A)$ for any symmetric matrix A . Thus $\widehat{a}(x_0) \nabla q(x_0) = 0$ for all $q \in \mathcal{Q}$ by (A2), which implies that $\widehat{a}(x_0) = \sum_i u_i u_i^\top$ for some vectors u_i in the tangent space of M at x_0 . Thus, choosing curves γ with $\gamma'(0) = u_i$, (E.5) yields

$$\text{Tr}\left(\left(\nabla^2 f(x_0) - \sum_{q \in \mathcal{Q}} c_q \nabla^2 q(x_0)\right) \widehat{a}(x_0)\right) \leq 0. \tag{E.6}$$

Combining (E.4), (E.6) and Lemma E.2, we obtain

$$\widehat{\mathcal{G}}f(x_0) = \frac{1}{2} \text{Tr}(\widehat{a}(x_0) \nabla^2 f(x_0)) + \widehat{b}(x_0)^\top \nabla f(x_0) \leq \sum_{q \in \mathcal{Q}} c_q \widehat{\mathcal{G}}q(x_0) = 0,$$

as desired. □

Let $C_0(E_0)$ denote the space of continuous functions on E_0 vanishing at infinity. Lemma E.3 implies that $\widehat{\mathcal{G}}$ is a well-defined linear operator on $C_0(E_0)$ with domain $C_c^\infty(E_0)$. It also implies that $\widehat{\mathcal{G}}$ satisfies the positive maximum principle as a linear operator on $C_0(E_0)$. Hence the following local existence result can be proved.

Lemma E.4 *Let μ be a probability measure on E . There exists an \mathbb{R}^d -valued càdlàg process X with initial distribution μ that satisfies*

$$X_t = X_0 + \int_0^t \widehat{b}(X_s) ds + \int_0^t \widehat{\sigma}(X_s) dW_s \tag{E.7}$$

for all $t < \tau$, where $\tau = \inf\{t \geq 0 : X_t \notin E_0\} > 0$, and some d -dimensional Brownian motion W .

Proof The conditions of Ethier and Kurtz [19, Theorem 4.5.4] are satisfied, so there exists an E_0^Δ -valued càdlàg process X such that $N_t^f = f(X_t) - f(X_0) - \int_0^t \widehat{\mathcal{G}}f(X_s) ds$ is a martingale for any $f \in C_c^\infty(E_0)$. Here E_0^Δ denotes the one-point compactification of E_0 with some $\Delta \notin E_0$, and we set $f(\Delta) = \widehat{\mathcal{G}}f(\Delta) = 0$. Bakry and Émery [4, Proposition 2] then yields that $f(X)$ and N^f are continuous.³ In particular, X cannot jump to Δ from any point in E_0 , whence τ is a strictly positive predictable time.

A localized version of the argument in Ethier and Kurtz [19, Theorem 5.3.3] now shows that on an extended probability space, X satisfies (E.7) for all $t < \tau$ and some Brownian motion W . It remains to show that X is non-explosive in the sense that $\sup_{t < \tau} \|X_t\| < \infty$ on $\{\tau < \infty\}$. Indeed, non-explosion implies that either $\tau = \infty$, or $\mathbb{R}^d \setminus E_0 \neq \emptyset$ in which case we can take $\Delta \in \mathbb{R}^d \setminus E_0$. In either case, X is \mathbb{R}^d -valued. To prove that X is non-explosive, let $Z_t = 1 + \|X_t\|^2$ for $t < \tau$, and observe that the linear growth condition (E.3) in conjunction with Itô’s formula yields $Z_t \leq Z_0 + C \int_0^t Z_s ds + N_t$ for all $t < \tau$, where $C > 0$ is a constant and N a local martingale on $[0, \tau)$. Let Y_t denote the right-hand side. Then

$$\begin{aligned} e^{-tC} Z_t &\leq e^{-tC} Y_t = Z_0 + C \int_0^t e^{-sC} (Z_s - Y_s) ds + \int_0^t e^{-sC} dN_s \\ &\leq Z_0 + \int_0^t e^{-sC} dN_s \end{aligned}$$

for all $t < \tau$. The right-hand side is a nonnegative supermartingale on $[0, \tau)$, and we deduce $\sup_{t < \tau} Z_t < \infty$ on $\{\tau < \infty\}$, as required. \square

Let X and τ be the process and stopping time provided by Lemma E.4. We now show that $\tau = \infty$ and that X_t remains in E for all $t \geq 0$ and spends zero time in each of the sets $\{p = 0\}$, $p \in \mathcal{P}$. This will complete the proof of Theorem 5.3, since \widehat{a} and \widehat{b} coincide with a and b on E .

We need to prove that $p(X_t) \geq 0$ for all $0 \leq t < \tau$ and all $p \in \mathcal{P}$. Fix $p \in \mathcal{P}$ and let L^y denote the local time of $p(X)$ at level y , where we choose a modification that is càdlàg in y ; see Revuz and Yor [41, Theorem VI.1.7]. Itô’s formula yields

$$p(X_t) = p(x) + \int_0^t \widehat{\mathcal{G}}p(X_s) ds + \int_0^t \nabla p(X_s)^\top \widehat{\sigma}(X_s)^{1/2} dW_s, \quad t < \tau.$$

³Note that unlike many other results in that paper, Proposition 2 in Bakry and Émery [4] does not require $\widehat{\mathcal{G}}$ to leave $C_c^\infty(E_0)$ invariant, and is thus applicable in our setting.

We first claim that $L_t^0 = 0$ for $t < \tau$. The occupation density formula [41, Corollary VI.1.6] yields

$$\int_{-\infty}^{\infty} \frac{1}{y} \mathbf{1}_{\{y>0\}} L_t^y dy = \int_0^t \frac{\nabla p^\top \widehat{a} \nabla p(X_s)}{p(X_s)} \mathbf{1}_{\{p(X_s)>0\}} ds.$$

By right-continuity of L_t^y in y , it suffices to show that the right-hand side is finite. For this, in turn, it is enough to prove that $(\nabla p^\top \widehat{a} \nabla p)/p$ is locally bounded on M . To this end, let $a = S\Lambda S^\top$ be the spectral decomposition of a , so that the columns S_i of S constitute an orthonormal basis of eigenvectors of a and the diagonal elements λ_i of Λ are the corresponding eigenvalues. Note that these quantities depend on x in general. Since $a \nabla p = 0$ on $M \cap \{p = 0\}$ by (A1), condition (G2) implies that there exists a vector $h = (h_1, \dots, h_d)^\top$ of polynomials such that

$$a \nabla p = hp \quad \text{on } M.$$

Thus $\lambda_i S_i^\top \nabla p = S_i^\top a \nabla p = S_i^\top hp$, and hence $\lambda_i (S_i^\top \nabla p)^2 = S_i^\top \nabla p S_i^\top hp$. In conjunction with Lemma E.1, this yields

$$\nabla p^\top \widehat{a} \nabla p = \nabla p^\top S \Lambda^+ S^\top \nabla p = \sum_i \lambda_i \mathbf{1}_{\{\lambda_i>0\}} (S_i^\top \nabla p)^2 = \sum_i \mathbf{1}_{\{\lambda_i>0\}} S_i^\top \nabla p S_i^\top hp.$$

Consequently,

$$\nabla p^\top \widehat{a} \nabla p \leq |p| \sum_i \|S_i\|^2 \|\nabla p\| \|h\|.$$

Since $\|S_i\| = 1$ and ∇p and h are locally bounded, we deduce that $(\nabla p^\top \widehat{a} \nabla p)/p$ is locally bounded, as required. Thus $L^0 = 0$ as claimed.

Next, since $\widehat{\mathcal{G}}p = \mathcal{G}p$ on E , the hypothesis (A1) implies that $\widehat{\mathcal{G}}p > 0$ on a neighborhood U_p of $E \cap \{p = 0\}$. Shrinking E_0 if necessary, we may assume that $E_0 \subseteq E \cup \bigcup_{p \in \mathcal{P}} U_p$ and thus

$$\widehat{\mathcal{G}}p > 0 \quad \text{on } E_0 \cap \{p = 0\}.$$

Since $L^0 = 0$ before τ , Lemma A.1 implies

$$p(X_t) \geq 0 \quad \text{for all } t < \tau.$$

Thus the stopping time $\tau_E = \inf\{t : X_t \notin E\} \leq \tau$ actually satisfies $\tau_E = \tau$. This implies $\tau = \infty$. Indeed, X has left limits on $\{\tau < \infty\}$ by Lemma E.4, and E_0 is a neighborhood in M of the closed set E . Thus $\tau_E < \tau$ on $\{\tau < \infty\}$, whence this set is empty. Finally, Lemma A.1 also gives $\int_0^t \mathbf{1}_{\{p(X_s)=0\}} ds = 0$. The proof of Theorem 5.3 is complete. \square

Appendix F: Proof of Theorem 5.7

The proof of Theorem 5.7 is divided into three parts.

Proof of Theorem 5.7(i) The following argument is a version of what is sometimes called “McKean’s argument”; see Mayerhofer et al. [37, Sect. 4.1] for an overview and further references. Suppose first $p(X_0) > 0$ almost surely. Itô’s formula and the identity $a\nabla h = hp$ on M yield

$$\begin{aligned} & \log p(X_t) - \log p(X_0) \\ &= \int_0^t \left(\frac{\mathcal{G} p(X_s)}{p(X_s)} - \frac{1}{2} \frac{\nabla p^\top a \nabla p(X_s)}{p(X_s)^2} \right) ds + \int_0^t \frac{\nabla p^\top \sigma(X_s)}{p(X_s)} dW_s \quad (\text{F.1}) \\ &= \int_0^t \frac{2\mathcal{G} p(X_s) - h^\top \nabla p(X_s)}{2p(X_s)} ds + \int_0^t \frac{\nabla p^\top \sigma(X_s)}{p(X_s)} dW_s \end{aligned}$$

for $t < \tau = \inf\{s \geq 0 : p(X_s) = 0\}$. We now modify $\log p(X)$ to turn it into a local submartingale. To this end, define

$$V_t = \int_0^t \mathbf{1}_{\{X_s \notin U\}} \frac{1}{p(X_s)} |2\mathcal{G} p(X_s) - h^\top \nabla p(X_s)| ds.$$

We claim that $V_t < \infty$ for all $t \geq 0$. To see this, note that the set $E \cap U^c \cap \{x : \|x\| \leq n\}$ is compact and disjoint from $\{p = 0\} \cap E$ for each n . Thus

$$\varepsilon_n = \min\{p(x) : x \in E \cap U^c, \|x\| \leq n\}$$

is strictly positive. Defining $\sigma_n = \inf\{t : \|X_t\| \geq n\}$, this yields

$$V_{t \wedge \sigma_n} \leq \frac{t}{2\varepsilon_n} \max_{\|x\| \leq n} |2\mathcal{G} p(x) - h^\top \nabla p(x)| < \infty.$$

Since $\sigma_n \rightarrow \infty$ due to the fact that X does not explode, we have $V_t < \infty$ for all $t \geq 0$ as claimed. It follows that the process

$$A_t = \int_0^t \mathbf{1}_{\{X_s \notin U\}} \frac{1}{p(X_s)} (2\mathcal{G} p(X_s) - h^\top \nabla p(X_s)) ds$$

is well defined and finite for all $t \geq 0$, with total variation process V .

Now define stopping times $\rho_n = \inf\{t \geq 0 : |A_t| + p(X_t) \geq n\}$ and note that $\rho_n \rightarrow \infty$ since neither A nor X explodes. Consider the process $Z = \log p(X) - A$, which satisfies

$$\begin{aligned} Z_t &= \log p(X_0) + \int_0^t \mathbf{1}_{\{X_s \in U\}} \frac{1}{2p(X_s)} (2\mathcal{G} p(X_s) - h^\top \nabla p(X_s)) ds \\ &\quad + \int_0^t \frac{\nabla p^\top \sigma(X_s)}{p(X_s)} dW_s. \end{aligned}$$

Then $-Z^{\rho_n}$ is a supermartingale on the stochastic interval $[0, \tau)$, bounded from below.⁴ Thus by the supermartingale convergence theorem, $\lim_{t \uparrow \tau} Z_{t \wedge \rho_n}$ exists in \mathbb{R} , which implies $\tau \geq \rho_n$. Since $\rho_n \rightarrow \infty$, we deduce $\tau = \infty$, as desired.

⁴Details regarding stochastic calculus on stochastic intervals are available in Maisonneuve [36]; see also Mayerhofer et al. [37], Carr et al. [7], Larsson and Ruf [34].

Finally, suppose $\mathbb{P}[p(X_0) = 0] > 0$. The above proof shows that $p(X)$ cannot return to zero once it becomes positive. But due to (5.2), we have $p(X_t) > 0$ for arbitrarily small $t > 0$, and this completes the proof. \square

Proof of Theorem 5.7(ii) As in the proof of (i), it is enough to consider the case where $p(X_0) > 0$. By (G2), we deduce $2\mathcal{L}p - h^\top \nabla p = \alpha p$ on M for some $\alpha \in \text{Pol}(\mathbb{R}^d)$. However, we have $\text{deg } \mathcal{L}p \leq \text{deg } p$ and $\text{deg } a \nabla p \leq 1 + \text{deg } p$, which yields $\text{deg } h \leq 1$. Consequently $\text{deg } \alpha p \leq \text{deg } p$, implying that α is constant. Inserting this into (F.1) yields

$$\log p(X_t) = \log p(X_0) + \frac{\alpha}{2}t + \int_0^t \frac{\nabla p^\top \sigma(X_s)}{p(X_s)} dW_s$$

for $t < \tau = \inf\{t : p(X_t) = 0\}$. The process $\log p(X_t) - \alpha t/2$ is thus locally a martingale bounded from above, and hence nonexplosive by the same ‘‘McKean’s argument’’ as in the proof of part (i). This proves the result. \square

Proof of Theorem 5.7(iii) The proof of relies on the following two lemmas.

Lemma F.1 *Let $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ be continuous functions with $\|b(x)\|^2 + \|\sigma(x)\|^2 \leq \kappa(1 + \|x\|^2)$ for some $\kappa > 0$, and fix $\rho > 0$. Let Y be a d -dimensional Itô process satisfying $Y_t = Y_0 + \int_0^t b(Y_s) ds + \int_0^t \sigma(Y_s) dW_s$. Then there exist constants $c_1, c_2 > 0$ that only depend on κ and ρ , but not on Y_0 , such that*

$$\mathbb{P}\left[\sup_{s \leq t} \|Y_s - Y_0\| < \rho\right] \geq 1 - tc_1(1 + \mathbb{E}[\|Y_0\|^2]), \quad t \leq c_2.$$

Proof By Markov’s inequality,

$$\mathbb{P}\left[\sup_{t \leq \varepsilon} \|Y_t - Y_0\| < \rho\right] \geq 1 - \rho^{-2} \mathbb{E}\left[\sup_{t \leq \varepsilon} \|Y_t - Y_0\|^2\right].$$

Let τ_n be the first time $\|Y_t\|$ reaches level n . A standard argument using the BDG inequality and Jensen’s inequality yields

$$\mathbb{E}\left[\sup_{s \leq t \wedge \tau_n} \|Y_s - Y_0\|^2\right] \leq 2c_2 \mathbb{E}\left[\int_0^{t \wedge \tau_n} (\|\sigma(Y_s)\|^2 + \|b(Y_s)\|^2) ds\right]$$

for $t \leq c_2$, where c_2 is the constant in the BDG inequality. The growth condition yields

$$\begin{aligned} \mathbb{E}\left[\sup_{s \leq t \wedge \tau_n} \|Y_s - Y_0\|^2\right] &\leq 2c_2 \kappa \mathbb{E}\left[\int_0^{t \wedge \tau_n} (1 + \|Y_s\|^2) ds\right] \\ &\leq 4c_2 \kappa (1 + \mathbb{E}[\|Y_0\|^2])t + 4c_2 \kappa \int_0^t \mathbb{E}\left[\sup_{u \leq s \wedge \tau_n} \|Y_u - Y_0\|^2\right] ds, \end{aligned}$$

for $t \leq c_2$, and Gronwall’s lemma then gives $\mathbb{E}[\sup_{s \leq t \wedge \tau_n} \|Y_s - Y_0\|^2] \leq c_3 t e^{4c_2 \kappa t}$, where $c_3 = 4c_2 \kappa (1 + \mathbb{E}[\|Y_0\|^2])$. Sending n to infinity and applying Fatou’s lemma concludes the proof, upon setting $c_1 = 4c_2 \kappa e^{4c_2^2 \kappa} \wedge c_2$. \square

Lemma F.2 *Let $0 < \alpha < 2$ and $z \geq 0$, and let Z be a BESQ(α) process starting from $z \geq 0$. Let \mathbb{P}_z denote its law. Let $\tau_0 = \inf\{t \geq 0 : Z_t = 0\}$ be the first time Z hits zero. Then for any $\varepsilon > 0$,*

$$\lim_{z \rightarrow 0} \mathbb{P}_z[\tau_0 > \varepsilon] = 0.$$

Proof By Göing-Jaeschke and Yor [26, Eq. (15)], we have

$$\mathbb{P}_z[\tau_0 > \varepsilon] = \int_\varepsilon^\infty \frac{1}{t \Gamma(\widehat{v})} \left(\frac{z}{2t}\right)^{\widehat{v}} e^{-z/(2t)} dt,$$

where $\Gamma(\cdot)$ is the Gamma function and $\widehat{v} = 1 - \alpha/2 \in (0, 1)$. Changing variables to $s = z/(2t)$ yields $\mathbb{P}_z[\tau_0 > \varepsilon] = \frac{1}{\Gamma(\widehat{v})} \int_0^{z/(2\varepsilon)} s^{\widehat{v}-1} e^{-s} ds$, which converges to zero as $z \rightarrow 0$ by dominated convergence. \square

We may now complete the proof of Theorem 5.7(iii). The hypotheses yield

$$0 \leq 2\mathcal{G}p(\bar{x}) < h(\bar{x})^\top \nabla p(\bar{x}).$$

Hence there exist some $\delta > 0$ such that $2\mathcal{G}p(\bar{x}) < (1 - 2\delta)h(\bar{x})^\top \nabla p(\bar{x})$ and an open ball U in \mathbb{R}^d of radius $\rho > 0$, centered at \bar{x} , such that

$$2\mathcal{G}p \leq (1 - \delta)h^\top \nabla p \quad \text{and} \quad h^\top \nabla p > 0 \quad \text{on } E \cap U.$$

Note that the radius ρ does not depend on the starting point X_0 .

For all $t < \tau(U) = \inf\{s \geq 0 : X_s \notin U\} \wedge T$, we have

$$\begin{aligned} p(X_t) - p(X_0) - \int_0^t \mathcal{G}p(X_s) ds &= \int_0^t \nabla p^\top \sigma(X_s) dW_s \\ &= \int_0^t \sqrt{\nabla p^\top a \nabla p(X_s)} dB_s \\ &= 2 \int_0^t \sqrt{p(X_s)} \frac{1}{2} \sqrt{h^\top \nabla p(X_s)} dB_s \end{aligned}$$

for some one-dimensional Brownian motion, possibly defined on an enlargement of the original probability space. Here the equality $a \nabla p = hp$ on E was used in the last step. Define an increasing process $A_t = \int_0^t \frac{1}{4} h^\top \nabla p(X_s) ds$. Since $h^\top \nabla p(X_t) > 0$ on $[0, \tau(U))$, the process A is strictly increasing there. It follows that the time-change $\gamma_u = \inf\{t \geq 0 : A_t > u\}$ is continuous and strictly increasing on $[0, A_\tau(U))$. The time-changed process $Y_u = p(X_{\gamma_u})$ thus satisfies

$$Y_u = p(X_0) + \int_0^u \frac{4\mathcal{G}p(X_{\gamma_v})}{h^\top \nabla p(X_{\gamma_v})} dv + 2 \int_0^u \sqrt{Y_v} d\beta_v, \quad u < A_\tau(U).$$

Consider now the BESQ(2 − 2δ) process Z defined as the unique strong solution to the equation

$$Z_u = p(X_0) + (2 - 2\delta)u + 2 \int_0^u \sqrt{Z_v} d\beta_v.$$

Since $4\mathcal{G}p(X_t)/h^\top \nabla p(X_t) \leq 2 - 2\delta$ for $t < \tau(U)$, a standard comparison theorem implies that $Y_u \leq Z_u$ for $u < A_{\tau(U)}$; see for instance Rogers and Williams [42, Theorem V.43.1]. It is well known that a BESQ(α) process hits zero if and only if $\alpha < 2$; see Revuz and Yor [41, page 442]. It thus remains to exhibit $\varepsilon > 0$ such that if $\|X_0 - \bar{x}\| < \varepsilon$ almost surely, there is a positive probability that Z_u hits zero before X_{γ_u} leaves U , or equivalently, that $Z_u = 0$ for some $u < A_{\tau(U)}$. To this end, set $C = \sup_{x \in U} h(x)^\top \nabla p(x)/4$, so that $A_{\tau(U)} \geq C\tau(U)$, and let $\eta > 0$ be a number to be determined later. We have

$$\begin{aligned} & \mathbb{P}\left[\eta < A_{\tau(U)} \text{ and } \inf_{u \leq \eta} Z_u = 0\right] \\ & \geq \mathbb{P}[\eta < A_{\tau(U)}] - \mathbb{P}\left[\inf_{u \leq \eta} Z_u > 0\right] \\ & \geq \mathbb{P}[\eta C^{-1} < \tau(U)] - \mathbb{P}\left[\inf_{u \leq \eta} Z_u > 0\right] \tag{F.2} \\ & = \mathbb{P}\left[\sup_{t \leq \eta C^{-1}} \|X_t - \bar{x}\| < \rho\right] - \mathbb{P}\left[\inf_{u \leq \eta} Z_u > 0\right] \\ & \geq \mathbb{P}\left[\sup_{t \leq \eta C^{-1}} \|X_t - X_0\| < \rho/2\right] - \mathbb{P}\left[\inf_{u \leq \eta} Z_u > 0\right], \end{aligned}$$

where we recall that ρ is the radius of the open ball U , and where the last inequality follows from the triangle inequality provided $\|X_0 - \bar{x}\| \leq \rho/2$. By Lemma F.1, we can choose $\eta > 0$ independently of X_0 so that $\mathbb{P}[\sup_{t \leq \eta C^{-1}} \|X_t - X_0\| < \rho/2] > 1/2$. Then by Lemma F.2, we have $\mathbb{P}[\inf_{u \leq \eta} Z_u > 0] < 1/3$ whenever $Z_0 = p(X_0)$ is sufficiently close to zero. This happens if X_0 is sufficiently close to \bar{x} , say within a distance $\rho' > 0$. Thus, setting $\varepsilon = \rho' \wedge (\rho/2)$, the condition $\|X_0 - \bar{x}\| < \rho' \wedge (\rho/2)$ implies that (F.2) is valid, with the right-hand side strictly positive. The theorem is proved. □

Appendix G: Proof of Proposition 6.1

Condition (G1) is vacuously true, so we prove (G2). If $d = 1$, then $\{p = 0\} = \{-1, 1\}$, and it is clear that any univariate polynomial vanishing on this set has $p(x) = 1 - x^2$ as a factor. Thus (G2) holds. If $d \geq 2$, then $p(x) = 1 - x^\top Qx$ is irreducible and changes sign, so (G2) follows from Lemma 5.4.

Next, it is straightforward to verify that (6.1), (6.2) imply (A0)–(A2), so we focus on the converse direction and assume (A0)–(A2) hold. We first prove that $a(x)$ has the stated form. Write $a(x) = \alpha + L(x) + A(x)$, where $\alpha = a(0) \in \mathbb{S}_+^d$, $L(x) \in \mathbb{S}^d$ is linear in x , and $A(x) \in \mathbb{S}^d$ is homogeneous of degree two in x .

Since $a(x)Qx = a(x)\nabla p(x)/2 = 0$ on $\{p = 0\}$, we have for any $x \in \{p = 0\}$ and $\epsilon \in \{-1, 1\}$ that

$$0 = \epsilon a(\epsilon x)Qx = \epsilon(\alpha Qx + A(x)Qx) + L(x)Qx.$$

This implies $L(x)Qx = 0$ for all $x \in \{p = 0\}$, and thus, by scaling, for all $x \in \mathbb{R}^d$. We now argue that this implies $L = 0$. To this end, consider the linear map $T : \mathcal{X} \rightarrow \mathcal{Y}$ where

$$\begin{aligned} \mathcal{X} &= \{\text{all linear maps } \mathbb{R}^d \rightarrow \mathbb{S}^d\}, \\ \mathcal{Y} &= \{\text{all second degree homogeneous maps } \mathbb{R}^d \rightarrow \mathbb{R}^d\}, \end{aligned}$$

and $TK \in \mathcal{Y}$ is given by $(TK)(x) = K(x)Qx$. One readily checks that we have $\dim \mathcal{X} = \dim \mathcal{Y} = d^2(d + 1)/2$. Thus if we can show that T is surjective, the rank-nullity theorem $\dim(\ker T) + \dim(\text{range } T) = \dim \mathcal{X}$ implies that $\ker T$ is trivial. But the identity $L(x)Qx \equiv 0$ precisely states that $L \in \ker T$, yielding $L = 0$ as desired. To see that T is surjective, note that \mathcal{Y} is spanned by elements of the form

$$(0, \dots, 0, x_i x_j, 0, \dots, 0)^\top$$

with the k th component being nonzero. But all these elements can be realized as $(TK)(x) = K(x)Qx$ as follows: If i, j, k are all distinct, one may take

$$\begin{pmatrix} K_{ii} & K_{ij} & K_{ik} \\ K_{ji} & K_{jj} & K_{jk} \\ K_{ki} & K_{kj} & K_{kk} \end{pmatrix} (x) = \frac{1}{2} \begin{pmatrix} 0 & -x_k & x_j \\ -x_k & 0 & x_i \\ x_j & x_i & 0 \end{pmatrix} \begin{pmatrix} Q_{ii} & 0 & 0 \\ 0 & Q_{jj} & 0 \\ 0 & 0 & Q_{kk} \end{pmatrix},$$

and all remaining entries of $K(x)$ equal to zero. If $i = k$, one takes $K_{ii}(x) = x_j$ and the remaining entries zero, and similarly if $j = k$. If $i = j \neq k$, one sets

$$\begin{pmatrix} K_{ii} & K_{ik} \\ K_{ki} & K_{kk} \end{pmatrix} (x) = \begin{pmatrix} -x_k & x_i \\ x_i & 0 \end{pmatrix} \begin{pmatrix} Q_{ii} & 0 \\ 0 & Q_{kk} \end{pmatrix},$$

and the remaining entries zero. This covers all possible cases, and shows that T is surjective. Thus $L = 0$ as claimed.

At this point, we have shown that $a(x) = \alpha + A(x)$ with A homogeneous of degree two. Next, since $a\nabla p = 0$ on $\{p = 0\}$, there exists a vector h of polynomials such that $a\nabla p/2 = hp$. By counting degrees, h is of the form $h(x) = f + Fx$ for some $f \in \mathbb{R}^d$, $F \in \mathbb{R}^{d \times d}$. For any $s > 0$ and $x \in \mathbb{R}^d$ such that $sx \in E$,

$$\alpha Qx + s^2 A(x)Qx = \frac{1}{2s} a(sx)\nabla p(sx) = (1 - s^2 x^\top Qx)(s^{-1} f + Fx).$$

By sending s to zero, we deduce $f = 0$ and $\alpha x = Fx$ for all x in some open set, hence $F = \alpha$. Thus $a(x)Qx = (1 - x^\top Qx)\alpha Qx$ for all $x \in E$. Defining $c(x) = a(x) - (1 - x^\top Qx)\alpha$, this shows that $c(x)Qx = 0$ for all $x \in \mathbb{R}^d$, that $c(0) = 0$, and that $c(x)$ has no linear part. In particular, c is homogeneous of degree

two. To prove that $c \in \mathcal{C}_+^Q$, it only remains to show that $c(x)$ is positive semidefinite for all x . For this we observe that for any $u \in \mathbb{R}^d$ and any $x \in \{p = 0\}$,

$$u^\top c(x)u = u^\top a(x)u \geq 0.$$

In view of the homogeneity property, positive semidefiniteness follows for any x . Thus $c \in \mathcal{C}_+^Q$ and hence this $a(x)$ has the stated form. Furthermore, the drift vector is always of the form $b(x) = \beta + Bx$, and a brief calculation using the expressions for $a(x)$ and $b(x)$ shows that the condition $\mathcal{G}p > 0$ on $\{p = 0\}$ is equivalent to (6.2). \square

Appendix H: Proof of Proposition 6.4

Condition (G1) is vacuously true, and it is not hard to check that (G2) holds.

Next, it is straightforward to verify that (i) and (ii) imply (A0)–(A2), so we focus on the converse direction and assume (A0)–(A2) hold.

We first deduce (i) from the condition $a \nabla p = 0$ on $\{p = 0\}$ for all $p \in \mathcal{P}$ together with the positive semidefinite requirement of $a(x)$. Taking $p(x) = x_i, i = 1, \dots, d$, we obtain $a(x) \nabla p(x) = a(x)e_i = 0$ on $\{x_i = 0\}$. Hence the i th column of $a(x)$ is a polynomial multiple of x_i . Similarly, with $p = 1 - x_i, i \in I$, it follows that $a(x)e_i$ is a polynomial multiple of $1 - x_i$ for $i \in I$. Hence, by symmetry of a , we get

$$\gamma_{ji}x_i(1 - x_i) = a_{ji}(x) = a_{ij}(x) = h_{ij}(x)x_j \quad (i \in I, j \in I \cup J)$$

for some constants γ_{ij} and polynomials $h_{ij} \in \text{Pol}_1(E)$ (using also that $\deg a_{ij} \leq 2$). For $i \neq j$, this is possible only if $a_{ij}(x) = 0$, and for $i = j \in I$ it implies that $a_{ii}(x) = \gamma_i x_i(1 - x_i)$ as desired. In order to maintain positive semidefiniteness, we necessarily have $\gamma_i \geq 0$.

Now consider $i, j \in J$. By the above, we have $a_{ij}(x) = h_{ij}(x)x_j$ for some $h_{ij} \in \text{Pol}_1(E)$. Similarly as before, symmetry of $a(x)$ yields

$$h_{ij}(x)x_j = a_{ij}(x) = a_{ji}(x) = h_{ji}(x)x_i,$$

so that for $i \neq j, h_{ij}$ has x_i as a factor. It follows that $a_{ij}(x) = \alpha_{ij}x_i x_j$ for some $\alpha_{ij} \in \mathbb{R}$. If $i = j$, we get $a_{jj}(x) = \alpha_{jj}x_j^2 + x_j(\phi_j + \psi_{(j)}^\top x_I + \pi_{(j)}^\top x_J)$ for some $\alpha_{jj} \in \mathbb{R}, \phi_j \in \mathbb{R}, \psi_{(j)} \in \mathbb{R}^m, \pi_{(j)} \in \mathbb{R}^n$ with $\pi_{(j),j} = 0$. Positive semidefiniteness requires $a_{jj}(x) \geq 0$ for all $x \in E$. This directly yields $\pi_{(j)} \in \mathbb{R}_+^n$. Further, by setting $x_i = 0$ for $i \in J \setminus \{j\}$ and making $x_j > 0$ sufficiently small, we see that $\phi_j + \psi_{(j)}^\top x_I \geq 0$ is required for all $x_I \in [0, 1]^m$, which forces $\phi_j \geq (\psi_{(j)}^-)^\top \mathbf{1}$. Finally, let $\alpha \in \mathbb{S}^n$ be the matrix with elements α_{ij} for $i, j \in J$, let $\Psi \in \mathbb{R}^{m \times n}$ have columns $\psi_{(j)}$, and $\Pi \in \mathbb{R}^{n \times n}$ columns $\pi_{(j)}$. We then have

$$s^{-2}a_{JJ}(x_I, sx_J) = \text{Diag}(x_J)\alpha \text{Diag}(x_J) + \text{Diag}(x_J) \text{Diag}(s^{-1}(\phi + \Psi^\top x_I) + \Pi^\top x_J),$$

so by sending s to infinity we see that $\alpha + \text{Diag}(\Pi^\top x_J) \text{Diag}(x_J)^{-1}$ must lie in \mathbb{S}_+^n for all $x_J \in \mathbb{R}_{++}^n$. This proves (i).

For (ii), note that $\mathcal{G}p(x) = b_i(x)$ for $p(x) = x_i$, and $\mathcal{G}p(x) = -b_i(x)$ for $p(x) = 1 - x_i$. In particular, if $i \in I$, then $b_i(x)$ cannot depend on x_J . This establishes (6.4). Next, for $i \in I$, we have $\beta_i + B_{iI}x_I > 0$ for all $x_I \in [0, 1]^m$ with $x_i = 0$, and this yields $\beta_i - (B_{i,I \setminus \{i\}}^-)\mathbf{1} > 0$. Similarly, $\beta_i + B_{iI}x_I < 0$ for all $x_I \in [0, 1]^m$ with $x_i = 1$, so that $\beta_i + (B_{i,I \setminus \{i\}}^+)\mathbf{1} + B_{ii} < 0$. For $j \in J$, we may set $x_J = 0$ to see that $\beta_j + B_{jI}x_I \in \mathbb{R}_{++}^n$ for all $x_I \in [0, 1]^m$. Hence $\beta_j > (B_{jI}^-)\mathbf{1}$ for all $j \in J$. Moreover, fixing $j \in J$, setting $x_j = 0$ and letting $x_i \rightarrow \infty$ for $i \neq j$ forces $B_{ji} > 0$. The proof of (ii) is complete. \square

Appendix I: Proof of Proposition 6.6

Since \mathcal{Q} consists of the single polynomial $q(x) = 1 - \mathbf{1}^\top x$, it is clear that (G1) holds. To prove (G2), it suffices by Lemma 5.5 to prove for each i that the ideal $(x_i, 1 - \mathbf{1}^\top x)$ is prime and has dimension $d - 2$. But an affine change of coordinates shows that this is equivalent to the same statement for (x_1, x_2) , which is well known to be true.

Next, the only nontrivial aspect of verifying that (i) and (ii) imply (A0)–(A2) is to check that $a(x)$ is positive semidefinite for each $x \in E$. To do this, fix any $x \in E$ and let A denote the diagonal matrix with $a_{ii}(x)$, $i = 1, \dots, d$, on the diagonal. Then for each $s \in [0, 1)$, the matrix $A(s) = (1 - s)(A + \text{Id}) + sa(x)$ is strictly diagonally dominant⁵ with positive diagonal elements. Hence by Horn and Johnson [30, Theorem 6.1.10], it is positive definite. But since \mathbb{S}_+^d is closed and $\lim_{s \rightarrow 1} A(s) = a(x)$, we get $a(x) \in \mathbb{S}_+^d$.

We now focus on the converse direction and assume (A0)–(A2) hold. We first prove (i). As the ideal $(x_i, 1 - \mathbf{1}^\top x)$ satisfies (G2) for each i , the condition $a(x)e_i = 0$ on $M \cap \{x_i = 0\}$ implies that

$$a_{ji}(x) = x_i h_{ji}(x) + (1 - \mathbf{1}^\top x)g_{ji}(x) \tag{I.1}$$

for some polynomials h_{ji} and g_{ji} in $\text{Pol}_1(\mathbb{R}^d)$. Suppose $j \neq i$. By symmetry of $a(x)$, we get

$$x_j h_{ij}(x) = x_i h_{ji}(x) + (1 - \mathbf{1}^\top x)(g_{ji}(x) - g_{ij}(x)).$$

Thus $h_{ij} = 0$ on $M \cap \{x_i = 0\} \cap \{x_j \neq 0\}$, and, by continuity, on $M \cap \{x_i = 0\}$. Another application of (G2) and counting degrees gives $h_{ij}(x) = -\alpha_{ij}x_i + (1 - \mathbf{1}^\top x)\gamma_{ij}$ for some constants α_{ij} and γ_{ij} . This proves $a_{ij}(x) = -\alpha_{ij}x_i x_j$ on E for $i \neq j$, as claimed. For $i = j$, note that (I.1) can be written as

$$a_{ii}(x) = -\alpha_{ii}x_i^2 + x_i(\phi_i + \psi_{(i)}^\top x) + (1 - \mathbf{1}^\top x)g_{ii}(x)$$

for some constants α_{ij} , ϕ_i and vectors $\psi_{(i)} \in \mathbb{R}^d$ with $\psi_{(i),i} = 0$. We need to identify ϕ_i and $\psi_{(i)}$. To this end, note that the condition $a(x)\mathbf{1} = 0$ on $\{1 - \mathbf{1}^\top x = 0\}$

⁵A matrix A is called *strictly diagonally dominant* if $|A_{ii}| > \sum_{j \neq i} |A_{ij}|$ for all i ; see Horn and Johnson [30, Definition 6.1.9].

yields $a(x)\mathbf{1} = (1 - \mathbf{1}^\top x)f(x)$ for all $x \in \mathbb{R}^d$, where f is some vector of polynomials $f_i \in \text{Pol}_1(\mathbb{R}^d)$. Writing the i th component of $a(x)\mathbf{1}$ in two ways then yields

$$\begin{aligned} x_i \left(- \sum_{j=1}^d \alpha_{ij} x_j + \phi_i + \psi_{(i)}^\top x \right) &= (1 - \mathbf{1}^\top x)(f_i(x) - g_{ii}(x)) \\ &= (1 - \mathbf{1}^\top x)(\eta_i + (\mathbf{H}x)_i) \end{aligned} \tag{I.2}$$

for all $x \in \mathbb{R}^d$ and some $\eta \in \mathbb{R}^d$, $\mathbf{H} \in \mathbb{R}^{d \times d}$. Replacing x by sx , dividing by s and sending s to zero gives $x_i \phi_i = \lim_{s \rightarrow 0} s^{-1} \eta_i + (\mathbf{H}x)_i$, which forces $\eta_i = 0$, $\mathbf{H}_{ij} = 0$ for $j \neq i$ and $\mathbf{H}_{ii} = \phi_i$. Substituting into (I.2) and rearranging yields

$$x_i \left(- \sum_{j=1}^d \alpha_{ij} x_j + \psi_{(i)}^\top x + \phi_i \mathbf{1}^\top x \right) = 0 \tag{I.3}$$

for all $x \in \mathbb{R}^d$. The coefficient in front of x_i^2 on the left-hand side is $-\alpha_{ii} + \phi_i$ (recall that $\psi_{(i),i} = 0$), which therefore is zero. That is, $\phi_i = \alpha_{ii}$. With this in mind, (I.3) becomes $x_i \sum_{j \neq i} (-\alpha_{ij} + \psi_{(i),j} + \alpha_{ii})x_j = 0$ for all $x \in \mathbb{R}^d$, which implies $\psi_{(i),j} = \alpha_{ij} - \alpha_{ii}$. At this point, we have proved

$$a_{ii}(x) = -\alpha_{ii}x_i^2 + x_i \left(\alpha_{ii} + \sum_{j \neq i} (\alpha_{ij} - \alpha_{ii})x_j \right) = \alpha_{ii}x_i(1 - \mathbf{1}^\top x) + \sum_{j \neq i} \alpha_{ij}x_i x_j$$

on E , which yields the stated form of $a_{ii}(x)$. It remains to show that $\alpha_{ij} \geq 0$ for all $i \neq j$. To see this, suppose for contradiction that $\alpha_{ik} < 0$ for some (i, k) . Pick $s \in (0, 1)$ and set $x_k = s$, $x_j = (1 - s)/(d - 1)$ for $j \neq k$. Then

$$a_{ii}(x) = x_i \sum_{j \neq i} \alpha_{ij} x_j = x_i \left(\alpha_{ik}s + \frac{1-s}{d-1} \sum_{j \neq i, k} \alpha_{ij} \right).$$

For s sufficiently close to 1, the right-hand side becomes negative, which contradicts positive semidefiniteness of a on E . This proves (i).

For (ii), first note that we always have $b(x) = \beta + Bx$ for some $\beta \in \mathbb{R}^d$ and $B \in \mathbb{R}^{d \times d}$. The condition $\mathcal{G}q = 0$ on M for $q(x) = 1 - \mathbf{1}^\top x$ yields $\beta^\top \mathbf{1} + x^\top B^\top \mathbf{1} = 0$ on M . Hence by Lemma 5.4, $\beta^\top \mathbf{1} + x^\top B^\top \mathbf{1} = \kappa(1 - \mathbf{1}^\top x)$ for all $x \in \mathbb{R}^d$ and some constant κ . This yields $\beta^\top \mathbf{1} = \kappa$ and then $B^\top \mathbf{1} = -\kappa \mathbf{1} = -(\beta^\top \mathbf{1})\mathbf{1}$. Next, the condition $\mathcal{G}p_i \geq 0$ on $M \cap \{p_i = 0\}$ for $p_i(x) = x_i$ can be written as

$$\min \left\{ \beta_i + \sum_{j=1}^d B_{ji}x_j : x \in \mathbb{R}_+^d, \mathbf{1}^\top x = \mathbf{1}, x_i = 0 \right\} \geq 0,$$

which in turn is equivalent to

$$\min \left\{ \beta_i + \sum_{j \neq i} B_{ji}x_j : x \in \mathbb{R}_+^d, \sum_{j \neq i} x_j = 1 \right\} \geq 0.$$

The feasible region of this optimization problem is the convex hull of $\{e_j : j \neq i\}$, and the linear objective function achieves its minimum at one of the extreme points. Thus we obtain $\beta_i + B_{ji} \geq 0$ for all $j \neq i$ and all i , as required. \square

Appendix J: Some notions from algebraic geometry

In this appendix, we briefly review some well-known concepts and results from algebra and algebraic geometry. The reader is referred to Dummit and Foote [16, Chaps. 7 and 15] and Bochnak et al. [6, Chap. 4] for more details.

An *ideal* I of $\text{Pol}(\mathbb{R}^d)$ is a subset of $\text{Pol}(\mathbb{R}^d)$ closed under addition and such that $f \in I$ and $g \in \text{Pol}(\mathbb{R}^d)$ implies $fg \in I$. Given a finite family $\mathcal{R} = \{r_1, \dots, r_m\}$ of polynomials, the *ideal generated by* \mathcal{R} , denoted by (\mathcal{R}) or (r_1, \dots, r_m) , is the ideal consisting of all polynomials of the form $f_1 r_1 + \dots + f_m r_m$, with $f_i \in \text{Pol}(\mathbb{R}^d)$. Given any set of polynomials S , its *zero set* is the set

$$\mathcal{V}(S) = \{x \in \mathbb{R}^d : f(x) = 0 \text{ for all } f \in S\}.$$

The zero set of the family \mathcal{R} coincides with the zero set of the ideal $I = (\mathcal{R})$, that is, $\mathcal{V}(\mathcal{R}) = \mathcal{V}(I)$. For example, the set M in (5.1) is the zero set of the ideal (\mathcal{Q}) . Given a set $V \subseteq \mathbb{R}^d$, the *ideal generated by* V , denoted by $\mathcal{I}(V)$, is the set of all polynomials that vanish on V . It follows from the definition that $S \subseteq \mathcal{I}(\mathcal{V}(S))$ for any set S of polynomials. A basic problem in algebraic geometry is to establish when an ideal I is equal to the ideal generated by the zero set of I ,

$$I = \mathcal{I}(\mathcal{V}(I)). \quad (\text{J.1})$$

If the ideal $I = (\mathcal{R})$ satisfies (J.1), then that means that any polynomial f that vanishes on the zero set $\mathcal{V}(I)$ has a representation $f = f_1 r_1 + \dots + f_m r_m$ for some polynomials f_1, \dots, f_m .

An ideal I of $\text{Pol}(\mathbb{R}^d)$ is said to be *prime* if it is not all of $\text{Pol}(\mathbb{R}^d)$ and if the conditions $f, g \in \text{Pol}(\mathbb{R}^d)$ and $fg \in I$ imply $f \in I$ or $g \in I$. The *dimension* of an ideal I of $\text{Pol}(\mathbb{R}^d)$ is the dimension of the quotient ring $\text{Pol}(\mathbb{R}^d)/I$; for a definition of the latter, see Dummit and Foote [16, Sect. 16.1].

References

1. Akerer, D., Filipović, D.: Linear credit risk models. Swiss Finance Institute Research Paper No. 16-34 (2016). Available online at <http://ssrn.com/abstract=2782455>
2. Akerer, D., Filipović, D., Pulido, S.: The Jacobi stochastic volatility model. Swiss Finance Institute Research Paper No. 16-35 (2016). Available online at <http://ssrn.com/abstract=2782486>
3. Akhiezer, N.I.: The Classical Moment Problem and Some Related Questions in Analysis. Oliver & Boyd, Edinburgh (1965)
4. Bakry, D., Émery, M.: Diffusions hypercontractives. In: Yor, M., Azéma, J. (eds.) Séminaire de Probabilités XIX. Lecture Notes in Mathematics, vol. 1123, pp. 177–206. Springer, Berlin (1985)
5. Berg, C., Christensen, J.P.R., Jensen, C.U.: A remark on the multidimensional moment problem. Math. Ann. **243**, 163–169 (1979)
6. Bochnak, J., Coste, M., Roy, M.-F.: Real Algebraic Geometry. Springer, Berlin (1998)

7. Carr, P., Fisher, T., Ruf, J.: On the hedging of options on exploding exchange rates. *Finance Stoch.* **18**, 115–144 (2014)
8. Cherny, A.: On the uniqueness in law and the pathwise uniqueness for stochastic differential equations. *Theory Probab. Appl.* **46**, 406–419 (2002)
9. Cuchiero, C.: Affine and polynomial processes. Ph.D. thesis, ETH Zurich (2011). Available online at <http://e-collection.library.ethz.ch/eserv/eth:4629/eth-4629-02.pdf>
10. Cuchiero, C., Keller-Ressel, M., Teichmann, J.: Polynomial processes and their applications to mathematical finance. *Finance Stoch.* **16**, 711–740 (2012)
11. Curtiss, J.H.: A note on the theory of moment generating functions. *Ann. Math. Stat.* **13**, 430–433 (1942)
12. Da Prato, G., Frankowska, H.: Invariance of stochastic control systems with deterministic arguments. *J. Differ. Equ.* **200**, 18–52 (2004)
13. Da Prato, G., Frankowska, H.: Stochastic viability of convex sets. *J. Math. Anal. Appl.* **333**, 151–163 (2007)
14. Delbaen, F., Schachermayer, W.: A general version of the fundamental theorem of asset pricing. *Math. Ann.* **300**, 463–520 (1994)
15. Delbaen, F., Shirakawa, H.: An interest rate model with upper and lower bounds. *Asia-Pac. Financ. Mark.* **9**, 191–209 (2002)
16. Dummit, D.S., Foote, R.M.: *Abstract Algebra*, 3rd edn. Wiley, Hoboken (2004)
17. Dunkl, C.F.: Hankel transforms associated to finite reflection groups. *Contemp. Math.* **138**, 123–138 (1992)
18. Ethier, S.N.: A class of degenerate diffusion processes occurring in population genetics. *Commun. Pure Appl. Math.* **29**, 483–493 (1976)
19. Ethier, S.N., Kurtz, T.G.: *Markov Processes: Characterization and Convergence*. Wiley, Hoboken (2005)
20. Filipović, D., Mayerhofer, E., Schneider, P.: Density approximations for multivariate affine jump-diffusion processes. *J. Econom.* **176**, 93–111 (2013)
21. Filipović, D., Larsson, M., Trolle, A.: Linear-rational term structure models. *J. Finance*. Forthcoming. Available at SSRN <http://ssrn.com/abstract=2397898>
22. Filipović, D., Tappe, S., Teichmann, J.: Invariant manifolds with boundary for jump-diffusions. *Electron. J. Probab.* **19**, 1–28 (2014)
23. Filipović, D., Gourier, E., Mancini, L.: Quadratic variance swap models. *J. Financ. Econ.* **119**, 44–68 (2016)
24. Forman, J.L., Sørensen, M.: The Pearson diffusions: a class of statistically tractable diffusion processes. *Scand. J. Stat.* **35**, 438–465 (2008)
25. Gallardo, L., Yor, M.: A chaotic representation property of the multidimensional Dunkl processes. *Ann. Probab.* **34**, 1530–1549 (2006)
26. Göing-Jaesckhe, A., Yor, M.: A survey and some generalizations of Bessel processes. *Bernoulli* **9**, 313–349 (2003)
27. Gouriéroux, C., Jasiak, J.: Multivariate Jacobi process with application to smooth transitions. *J. Econom.* **131**, 475–505 (2006)
28. Hajek, B.: Mean stochastic comparison of diffusions. *Z. Wahrscheinlichkeitstheor. Verw. Geb.* **68**, 315–329 (1985)
29. Heyde, C.C.: On a property of the lognormal distribution. *J. R. Stat. Soc., Ser. B, Stat. Methodol.* **25**, 392–393 (1963)
30. Horn, R.A., Johnson, C.A.: *Matrix Analysis*. Cambridge University Press, Cambridge (1985)
31. Ikeda, N., Watanabe, S.: *Stochastic Differential Equations and Diffusion Processes*. North-Holland, Amsterdam (1981)
32. Kleiber, C., Stoyanov, J.: Multivariate distributions and the moment problem. *J. Multivar. Anal.* **113**, 7–18 (2013)
33. Larsen, K.S., Sørensen, M.: Diffusion models for exchange rates in a target zone. *Math. Finance* **17**, 285–306 (2007)
34. Larsson, M., Ruf, J.: Convergence of local supermartingales and Novikov–Kazamaki type conditions for processes with jumps (2014). [arXiv:1411.6229](https://arxiv.org/abs/1411.6229)
35. Lord, R., Koekoek, R., van Dijk, D.: A comparison of biased simulation schemes for stochastic volatility models. *Quant. Finance* **10**, 177–194 (2012)
36. Maisonneuve, B.: Une mise au point sur les martingales locales continues définies sur un intervalle stochastique. In: Dellacherie, C., et al. (eds.) *Séminaire de Probabilités XI. Lecture Notes in Mathematics*, vol. 581, pp. 435–445. Springer, Berlin (1977)

37. Mayerhofer, E., Pfaffel, O., Stelzer, R.: On strong solutions for positive definite jump diffusions. *Stoch. Process. Appl.* **121**, 2072–2086 (2011)
38. Mazet, O.: Classification des semi-groupes de diffusion sur \mathbb{R} associés à une famille de polynômes orthogonaux. In: Azéma, J., et al. (eds.) *Séminaire de Probabilités XXXI. Lecture Notes in Mathematics*, vol. 1655, pp. 40–53. Springer, Berlin (1997)
39. Penrose, R.: A generalized inverse for matrices. *Math. Proc. Camb. Philos. Soc.* **51**, 406–413 (1955)
40. Petersen, L.C.: On the relation between the multidimensional moment problem and the one-dimensional moment problem. *Math. Scand.* **51**, 361–366 (1982)
41. Revuz, D., Yor, M.: *Continuous Martingales and Brownian Motion*, 3rd edn. Springer, Berlin (1999)
42. Rogers, L.C.G., Williams, D.: *Diffusions, Markov Processes and Martingales*. Cambridge University Press, Cambridge (1994)
43. Schmüdgen, K.: The K -moment problem for compact semi-algebraic sets. *Math. Ann.* **289**, 203–206 (1991)
44. Spreij, P., Veerman, E.: Affine diffusions with non-canonical state space. *Stoch. Anal. Appl.* **30**, 605–641 (2012)
45. Stieltjes, T.J.: Recherches sur les fractions continues. *Ann. Fac. Sci. Toulouse* **8**(4), 1–122 (1894)
46. Stoyanov, J.: Krein condition in probabilistic moment problems. *Bernoulli* **6**, 939–949 (2000)
47. Willard, S.: *General Topology*. Courier Corporation, North Chelmsford (2004)
48. Wong, E.: The construction of a class of stationary Markoff processes. In: Bellman, R. (ed.) *Stochastic Processes in Mathematical Physics and Engineering*, pp. 264–276. Am. Math. Soc., Providence (1964)
49. Zhou, H.: Itô conditional moment generator and the estimation of short-rate processes. *J. Financ. Econom.* **1**, 250–271 (2003)