Polynomial preserving diffusions on compact quadric sets

Martin Larsson, Sergio Pulido

Polynomial preserving diffusions on compact quadric sets*

Martin Larsson† Sergio Pulido‡

November 19, 2015

Abstract

Polynomial preserving processes are defined by the property that conditional expectations of polynomial functions of the process are again polynomials of the same or lower degree. Many fundamental stochastic processes, including affine processes, are polynomial preserving, and their tractable structure makes them important in applications. In this paper we study polynomial preserving diffusions whose state space is a compact quadric set. Necessary and sufficient conditions for existence, uniqueness, and boundary attainment are given. The existence of a convenient parameterization of the generator is shown to be closely related to the classical problem of expressing nonnegative polynomials—specifically, biquadratic forms vanishing on the diagonal—as a sum of squares. We prove that in dimension $d \leq 4$ every such biquadratic form is a sum of squares, while for $d \geq 6$ there are counterexamples. The case $d = 5$ remains open. An equivalent probabilistic description of the sum of squares property is provided, and we show how it can be used to obtain results on pathwise uniqueness and existence of smooth densities.

Keywords: Polynomial preserving processes; sums of squares; biquadratic forms; stochastic invariance; pathwise uniqueness; smooth densities.

MSC2010 subject classifications: 60J60, 60H10, 11E25.

1 Introduction

Many fundamental stochastic processes appearing in probability theory are *polynomial preserving*, meaning that conditional expectations of polynomial functions of the process again

*The authors would like to thank Guillermo Mantilla-Soler for valuable discussions and suggestions. Martin Larsson gratefully acknowledges support from SNF Grant 205121_163425. The research of Sergio Pulido also benefited from the support of the "Chair Markets in Transition" under the aegis of Louis Bachelier laboratory, a joint initiative of École Polytechnique, Université d'Évry Val d'Essonne and Fédération Bancaire Française. 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.

†ETH Zurich, Department of Mathematics, Rämistrasse 101, CH-8092, Zurich, Switzerland. Email: martin.larsson@math.ethz.ch

‡Laboratoire de Mathématiques et Modélisation d'Évry (LaMME), Université d'Évry-Val-d'Essonne, ENSIIE, UMR CNRS 8071, IBGBI 23 Boulevard de France, 91037 Évry Cedex, France, Email: sergio.pulidonino@ensiie.fr.

have polynomial form; see (1.4) below. Examples include Brownian motion, geometric Brownian motion, Ornstein-Uhlenbeck processes, squared Bessel processes, Jacobi processes, Lévy processes, as well as large classes of multidimensional generalizations, such as affine processes and Fisher-Wright diffusions. Polynomial preserving processes have been studied in various degrees of generality by several authors, for instance Wong (1964), Mazet (1997), Forman and Sørensen (2008), Cuchiero (2011), Cuchiero et al. (2012), Bakry (2014), Bakry et al. (2014), Filipović and Larsson (2015). Polynomial preserving processes are also important in applications. In mathematical finance, the polynomial preserving property is useful to build tractable and flexible pricing models in a variety of situations; see for instance Zhou (2003), Delbaen and Shirakawa (2002), Larsen and Sørensen (2007), Gourieroux and Jasiak (2006), Cuchiero et al. (2012), Filipović et al. (2015a,b), Ackerer et al. (2015), and Filipović and Larsson (2015). Every affine diffusion is polynomial preserving; however, non-deterministic affine diffusions do not admit compact state spaces, see Krühner and Larsson (2015), which may be a drawback in applications. In the present paper we consider polynomial preserving diffusions on (possibly solid) compact quadric sets. This is a natural class of simple state spaces that nonetheless exhibit a rich mathematical structure.

The state space E is defined as follows. Fix $d \in \mathbb{N}$ and let

$$E = \{x \in \mathbb{R}^d \colon p(x) \geq 0\} \qquad \text{or} \qquad E = \{x \in \mathbb{R}^d \colon p(x) = 0\}$$

for some $p \in \mathrm{Pol}_2$ such that E is compact and nondegenerate (nonempty and not a point). Here Pol_k denotes the vector space of polynomial functions on $\mathbb{R}^d$ of total degree less than or equal to k. After an affine change of coordinates, E is either the closed unit ball $\mathscr{B}^d$ or the unit sphere $\mathscr{S}^{d-1}$; see Section 2.

Next, consider continuous maps $a : \mathbb{R}^d \to \mathbb{S}^d$ and $b : \mathbb{R}^d \to \mathbb{R}^d$ such that

$$a_{ij} \in \mathrm{Pol}_2 \text{ and } b_i \in \mathrm{Pol}_1 \text{ for all } i, j \in \{1, \ldots, d\}, \tag{1.1}$$

where $\mathbb{S}^d$ is the set of $d \times d$ symmetric matrices. We are interested in E-valued weak solutions to stochastic differential equations of the form

$$dX_t = b(X_t)\, dt + \sigma(X_t)\, dW_t, \tag{1.2}$$

where $\sigma : \mathbb{R}^d \to \mathbb{R}^{d \times n}$ for $n \in \mathbb{N}$ is a continuous map with $\sigma\sigma^\top = a$ on E, and W is an n-dimensional Brownian motion. In particular, we are interested in how probabilistic properties of (1.2) are connected to algebraic properties of the coefficients a, b and the state space E.

Diffusions whose coefficients satisfy (1.1) are called polynomial preserving. The reason is that (1.1) holds if and only if the associated generator $\mathscr{G}$, given by

$$\mathscr{G}f(x) = \frac{1}{2} \operatorname{Tr}(a(x)\nabla^2 f(x)) + b(x)^\top \nabla f(x), \tag{1.3}$$

is polynomial preserving in the sense that $\mathscr{G}\mathrm{Pol}_k \subseteq \mathrm{Pol}_k$ holds for all $k \in \mathbb{N}$. If X is a polynomial preserving diffusion with generator $\mathscr{G}$, a simple argument based on Itô's formula shows that conditional expectations of polynomials have a particularly simple form: For any $q \in \mathrm{Pol}_k$, one has

$$\mathbb{E}[q(X_T) \mid X_t] = H(X_t)^\top e^{(T-t)G_k}\, \vec{q}, \tag{1.4}$$

where $H(x) = (h_1(x), \ldots, h_{N_k}(x))$ with $N_k = \dim \mathrm{Pol}_k$ is a basis for Pol_k, and $G_k \in \mathbb{R}^{N_k \times N_k}$ and $\vec{q} \in \mathbb{R}^{N_k}$ are the corresponding coordinate representations of $\mathscr{G}|_{\mathrm{Pol}_k}$ and q, respectively; see Filipović and Larsson (2015, Theorem 3.1) or Cuchiero et al. (2012, Theorem 2.7) for details.

Let us summarize the main results of the present paper. Conditions for existence and uniqueness in law of solutions to (1.2) can be extracted from known results in the literature, and we briefly review these issues in Section 2. A related question in the case $E = \mathscr{B}^d$ is whether the interior of the state space is stochastically invariant, to which we provide a complete answer; see Proposition 2.2. Next, in Section 3 we undertake a more detailed investigation of how to parameterize all weak solutions to (1.2). Theorem 3.3 gives a general representation, modulo the requirement that the diffusion matrix $a(x)$ be positive semidefinite on the state space. Characterizing positivity is a difficult problem, which translates into the algebraic question of describing when certain biquadratic forms $\mathrm{BQ}(x, y)$ are nonnegative. This is closely related to the representability of $\mathrm{BQ}(x, y)$ as a sum of squares of polynomials. Such a representation always exists if $d \leq 4$, but not if $d \geq 6$; see Theorem 3.5. The case $d = 5$ remains open. In Section 4 we study the probabilistic consequences of the existence of a sum of squares representation. We show that such a representation exists if and only if (1.2) can be replaced by a certain much more structured SDE; see Theorem 4.1. Solutions to this SDE can sometimes be shown to have the pathwise uniqueness property; see Corollary 4.4 and Theorem 4.6. Moreover, it becomes possible to make detailed assertions regarding the existence of smooth densities by relying on the classical Hörmander condition; see Theorem 4.9 and its corollaries. Finally, in Section 5, we collect the algebraic developments needed for the results presented in Section 3. Here an important role is played by the so-called Plücker relations from algebraic geometry.

The following notation will be used throughout this paper: Elements of $\mathbb{R}^d$ are viewed as column vectors. Hom_k denotes the subspace of Pol_k consisting of homogeneous polynomials of total degree exactly k. For two matrices A and B of compatible size, we write $\langle A, B \rangle = \mathrm{Tr}(A^\top B)$ for the trace product. The convex cone of positive semi-definite matrices in $\mathbb{S}^d$ is denoted $\mathbb{S}^d_+$. The space of skew-symmetric $d \times d$ matrices is written $\mathrm{Skew}(d)$, and we let $\mathrm{Skew}(2, d)$ denote the subset of rank-two elements. Let $e_i = (0, \ldots, 0, 1, 0, \ldots, 1)^\top$ be the ith canonical unit vector in $\mathbb{R}^d$. We denote by $D_1, \ldots, D_m$, with $m = \dim \mathrm{Skew}(d) = \binom{d}{2}$, the elementary skew-symmetric matrices $S_{ij} = e_i e_j^\top - e_j e_i^\top$ listed in lexicographic order. That is,

$$(D_1, \ldots, D_m) = (S_{12}, \ldots, S_{1d}, S_{23}, \ldots, S_{2d}, \ldots, S_{d-1,d}). \tag{1.5}$$

2 Existence, uniqueness, and boundary attainment

We consider polynomial preserving processes whose state space is either the unit ball $\mathscr{B}^d$ or the unit sphere $\mathscr{S}^{d-1}$. Up to an affine change of coordinates, this covers all nondegenerate compact quadric sets. Indeed, consider a compact set $E = \{x \in \mathbb{R}^d : p(x) \geq 0\}$ that is nonempty and not a point, where $p \in \mathrm{Pol}_2$. Then $\nabla^2 p$ is constant and negative definite, so by completing the square one sees that E is an affine transformation of $\mathscr{B}^d$. Similarly, the set $\{x \in \mathbb{R}^d \colon p(x) = 0\}$ is an affine transformation of $\mathscr{S}^{d-1}$. Note also that the condition (1.1) is invariant under affine transformations.

The following sets play an important role in the description of polynomial preserving diffusions on $\mathscr{B}^d$ and $\mathscr{S}^{d-1}$:

$$\begin{aligned}\mathscr{C} &= \{c : \mathbb{R}^d \to \mathbb{S}^d : c_{ij} \in \mathrm{Hom}_2 \text{ for all } i, j, \text{ and } c(x)x \equiv 0\},\\ \mathscr{C}_+ &= \{c \in \mathscr{C} : c(x) \in \mathbb{S}^d_+ \text{ for all } x\}.\end{aligned} \tag{2.1}$$

We have the following crude characterization theorems, most of which follows fairly directly from known results. We start with the unit ball.

Theorem 2.1. *Assume a and b satisfy* (1.1) *and the state space is $E = \mathscr{B}^d$. The following conditions are equivalent:*

(i) *There exists a continuous map $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ with $\sigma\sigma^\top = a$ on $\mathscr{B}^d$ such that* (1.2) *admits a $\mathscr{B}^d$-valued weak solution for any initial condition in $\mathscr{B}^d$.*

(ii) *The coefficients a and b are of the form*

$$\begin{aligned}a(x) &= (1 - \|x\|^2)\alpha + c(x),\\ b(x) &= b + Bx,\end{aligned} \tag{2.2}$$

for some $\alpha \in \mathbb{S}^d_+$, $c \in \mathscr{C}_+$, $b \in \mathbb{R}^d$, and $B \in \mathbb{R}^{d\times d}$ such that

$$b^\top x + x^\top Bx + \frac{1}{2}\operatorname{Tr}(c(x)) \le 0 \quad \textit{for all} \quad x \in \mathscr{S}^{d-1}. \tag{2.3}$$

Proof. It is proved in Filipović and Larsson (2015, Proposition 7.1) that (ii) is equivalent to the conditions

$$a(x) \succeq 0 \text{ on } E, \quad a(x)x = 0 \text{ on } \mathscr{S}^{d-1}, \quad \mathscr{G}p(x) \ge 0 \text{ on } \mathscr{S}^{d-1}, \tag{2.4}$$

where $p(x) = 1-\|x\|^2$ and $\mathscr{G}$ is given by (1.3). That (i) implies (2.4) follows from Theorem 6.1 in Filipović and Larsson (2015). The reverse implication follows from Theorem 6.3 and Proposition 7.1 in Filipović and Larsson (2015) when the inequality in (2.3) is strict. For the general case one can, for instance, invoke Example 2.5 and Theorem 2.2 in Da Prato and Frankowska (2007). Alternatively, one can use (2.4) to verify the positive maximum principle directly, and then apply Ethier and Kurtz (2005, Theorem 4.5.4). □

If $d = 1$, the polynomial preserving diffusions on $\mathscr{B}^1 = [-1, 1]$ are simply the Jacobi processes $dX_t = (b + BX_t)dt + \sigma\sqrt{1 - X_t^2}dW_t$. In this case $\mathscr{C}_+ = \mathscr{C} = \{0\}$. In contrast, for $d > 1$ the set $\mathscr{C}_+$ is non-trivial, and its elements c induce diffusive fluctuations tangentially to the sphere due to the property $c(x)x \equiv 0$. Theorem 2.1 provides no explicit description of $\mathscr{C}_+$, and a significant part of the present paper is devoted to studying this set in detail. This is the topic Section 3.

It is frequently of interest to know whether a solution X to (1.2) that starts in the interior of $\mathscr{B}^d$ remains in the interior. This question of boundary attainment is resolved by the following result, which is a refinement of Theorem 6.7 in Filipović and Larsson (2015). In its statement, we assume that $a(x)$ and $b(x)$ satisfy (1.1) along with the two equivalent conditions in Theorem 2.1. In particular, (2.3) is satisfied. Let $\mathbb{P}_x$ denote the law of the solution X to (1.2) with initial condition $x \in \mathscr{B}^d$. A measurable subset $D \subseteq \mathscr{B}^d$ is said to be *invariant for* X if, for every $x \in D$, one has $\mathbb{P}_x(X_t \in D \text{ for all } t \ge 0) = 1$.

Proposition 2.2. *The interior $\mathscr{B}^d \setminus \mathscr{S}^{d-1}$ is invariant for X if and only if*

$$b^\top x + x^\top (B + \alpha)x + \frac{1}{2}\operatorname{Tr}(c(x)) \leq 0 \quad \textit{for all} \quad x \in \mathscr{S}^{d-1}. \tag{2.5}$$

Proof. Let $p(x) = 1 - \|x\|^2$ and note that

$$a(x)\nabla p(x) = p(x)h(x) \quad \text{where} \quad h(x) = 2\alpha x.$$

Furthermore,

$$2\mathscr{G}p(x) - h(x)^\top \nabla p(x) = -\operatorname{Tr}(\alpha)p(x) - 2\left(b^\top x + x^\top(B+\alpha)x + \frac{1}{2}\operatorname{Tr}(c(x))\right). \tag{2.6}$$

If (2.5) fails, then the right-hand side is strictly positive for some $x \in \mathscr{S}^{d-1}$. Also, $\mathscr{G}p(x) \geq 0$ due to (2.3). Thus Filipović and Larsson (2015, Theorem 6.7(iii)) implies that $\mathscr{B}^d \setminus \mathscr{S}^{d-1}$ is not invariant for X (a look at the proof of that result shows that its assumption that $\{t : X_t \in \mathscr{S}^{d-1}\}$ has Lebesgue measure zero is not needed when $X_0 \notin \mathscr{S}^{d-1}$).

Suppose now (2.5) holds, and let $X_0 = x_0 \in \mathscr{B}^d \setminus \mathscr{S}^{d-1}$. Define $\tau = \inf\{t : X_t \in \mathscr{S}^{d-1}\}$. We must show that $\mathbb{P}_{x_0}(\tau < \infty) = 0$. For $t \in [0, \tau)$, Itô's formula yields

$$\log p(X_t) = \log p(x_0) + \int_0^t \frac{2\mathscr{G}p(X_s) - h(X_s)^\top \nabla p(X_s)}{2p(X_s)}\,ds + \int_0^t \frac{\nabla p(X_s)^\top \sigma(X_s)}{p(X_s)}\,dW_s.$$

Suppose we can find a constant $\kappa_1 > 0$ such that

$$2\mathscr{G}p(x) - h(x)^\top \nabla p(x) \geq -2\kappa_1 p(x) \quad \text{for all} \quad x \in \mathscr{B}^d. \tag{2.7}$$

Then $\log p(X_t) + \kappa_1 t$ is a local submartingale on $[0, \tau)$, bounded from above on bounded time intervals. The supermartingale convergence theorem then implies $\inf_{t<T\wedge\tau} \log p(X_t) > -\infty$ for all deterministic $T < \infty$, which gives $\tau = \infty$ as desired. This is the same "McKean's argument" as in the proof of Filipović and Larsson (2015, Theorem 6.7(i)–(ii)). A suitable version of the supermartingale convergence theorem can found in, for instance, Larsson and Ruf (2014, Lemma 4.14).

It remains to argue (2.7). Since $\operatorname{Tr}(c(x))$ is a quadratic form in x, we can find $C \in \mathbb{S}^d$ such that $\operatorname{Tr}(c(x)) = x^\top C x$. Define $\Sigma = -\alpha - \frac{1}{2}(B + B^\top + C) \in \mathbb{S}^d$. In view of (2.6), condition (2.7) then becomes

$$x^\top \Sigma x - b^\top x \geq -\kappa(1 - \|x\|^2) \quad \text{for all} \quad x \in \mathscr{B}^d, \tag{2.8}$$

where the constant κ is related to κ_1 by $\kappa = \kappa_1 - \operatorname{Tr}(\alpha)/2$. On the other hand, the assumption (2.5) written in terms of Σ becomes

$$x^\top \Sigma x - b^\top x \geq 0 \quad \text{for all} \quad x \in \mathscr{S}^{d-1}. \tag{2.9}$$

We claim that (2.9) in fact implies that (2.8) holds for $\kappa = \|b\|$. For $x = 0$ the statement is obvious. For $x \neq 0$, define $\overline{x} = x/\|x\|$ and apply (2.9) to get $\overline{x}^\top \Sigma \overline{x} \geq b^\top \overline{x}$. Multiplying both sides by $\|x\|^2$, rearranging, and applying the Cauchy-Schwartz inequality gives

$$x^\top \Sigma x - b^\top x \geq -b^\top x(1 - \|x\|) \geq -\|b\|\|x\|(1 - \|x\|) \geq -\|b\|(1 - \|x\|^2),$$

as required. □

The case of the sphere is simpler. In particular, the question of boundary attainment becomes vacuous.

Theorem 2.3. *Assume a and b satisfy* (1.1) *and the state space is $E = \mathscr{S}^{d-1}$. The following conditions are equivalent:*

(i) *There exists a continuous map $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ with $\sigma\sigma^\top = a$ on $\mathscr{S}^{d-1}$ such that* (1.2) *admits an $\mathscr{S}^{d-1}$-valued weak solution for any initial condition in $\mathscr{S}^{d-1}$.*

(ii) *The coefficients a and b satisfy $a = c$ on $\mathscr{S}^{d-1}$ for some $c \in \mathscr{C}_+$, and $b(x) = Bx$ for some $B \in \mathbb{R}^{d\times d}$, where*

$$2x^\top Bx + \operatorname{Tr}(c(x)) \equiv 0. \tag{2.10}$$

Proof. Filipović and Larsson (2015, Theorem 6.3) yields that (ii) implies (i). For the reverse implication, one argues as in the proof of Theorem 2.1 that (i) implies Theorem 2.1(ii), but with equality in (2.3). This in turn implies that $a(x) = c(x)$ for $x \in \mathscr{S}^{d-1}$ and that $b = 0$, whence (2.10) holds. □

Theorems 2.1 and 2.3 do not make any uniqueness statements. However, since the state space E is compact, the polynomial preserving property implies that uniqueness in law for (1.2) always holds. More precisely, we have the following result:

Lemma 2.4. *Assume a and b satisfy* (1.1)*, let $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times n}$, $n \in \mathbb{N}$, be a continuous map with $\sigma\sigma^\top = a$ on E, and let $x \in E$. Let X be an E-valued weak solution to* (1.2) *with initial condition $X_0 = x$. Then the law of X is uniquely determined by a, b, and x.*

Proof. See Filipović and Larsson (2015, Corollary 5.2). The argument relies on the fact that a, b, and x uniquely determine all joint moments of all finite-dimensional marginal distributions of X due to (1.4); see Filipović and Larsson (2015, Corollary 3.2) for details. Due to the compactness of E, this in turn determines the law of X. □

Remark 2.5. *So far no assertions have been made regarding pathwise uniqueness of solutions to* (1.2)*. Pathwise uniqueness depends on the choice of σ, and the existence of a suitable σ is a delicate question in general. See Spreij and Veerman (2012, Remark 2.3) for an instructive discussion in connection with affine diffusions on parabolic sets. Sometimes pathwise uniqueness is immediate: If $a(x)$ is positive definite on the interior $\mathscr{B}^d \setminus \mathscr{S}^{d-1}$ then $\sigma(x) = a(x)^{1/2}$ is locally Lipschitz there, which by standard arguments yields pathwise uniqueness up to the first hitting time of the boundary. In Section* (2.2) *we discuss other situations where pathwise uniqueness can be established.*

3 Tangential diffusion, biquadratic forms, and sums of squares

In this section we undertake a detailed analysis of the linear space $\mathscr{C}$ and the convex cone $\mathscr{C}_+$ appearing in Theorems 2.1 and 2.3, and whose role is to generate diffusive movements tangentially to $\mathscr{S}^{d-1}$. It tuns out that a complete description is difficult to obtain in general; see Theorem 3.5 below. Fortunately, a partial characterization suitable for applications is

within reach. We start with some general remarks, introducing some concepts and notation along the way.

Any map $c : \mathbb{R}^d \to \mathbb{S}^d$ whose components lie in Hom_2 induces a map $\mathrm{BQ} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ via the expression

$$\mathrm{BQ}(x, y) = y^\top c(x) y. \tag{3.1}$$

This map is a *biquadratic form*: a quadratic form in x for each fixed y, and a quadratic form in y for each fixed x. Conversely, any biquadratic form in $d+d$ variables induces a corresponding map c by the formula $c_{ij}(x) = \partial^2_{y_i y_j} \mathrm{BQ}(x, y)$ for $i \neq j$, and $c_{ii}(x) = \partial^2_{y_i y_i} \mathrm{BQ}(x, y)/2$. Furthermore, the positive semidefiniteness condition $c(x) \succeq 0$ for all x is equivalent to $\mathrm{BQ}(x, y) \geq 0$ for all (x, y). Finally, in the presence of this condition, one has $c(x)x \equiv 0$ if and only if the corresponding biquadratic form vanishes on the diagonal, $\mathrm{BQ}(x, x) \equiv 0$; indeed, $c(x)x = 0$ if and only if $x^\top c(x) x = \|c(x)^{1/2} x\|^2 = 0$. These observations can be summarized as follows:

Lemma 3.1. *The set $\mathscr{C}_+$ is in one-to-one correspondence with the set of all nonnegative biquadratic forms vanishing on the diagonal. The correspondence is given by* (3.1).

Remark 3.2. *Note that the set $\mathscr{C}$* does not *correspond one-to-one with the set of all (not necessarily nonnegative) biquadratic forms vanishing on the diagonal. The reason is that $x^\top c(x) x \equiv 0$ does not imply $c(x)x \equiv 0$ in the absence of positive semidefiniteness. An example is the map*

$$c(x) = \begin{pmatrix} -2x_1 x_2 & x_1^2 \\ x_1^2 & 0 \end{pmatrix},$$

which satisfies $x^\top c(x) x \equiv 0$ but not $c(x)x \equiv 0$. It will become apparent that the relevant stepping stone toward an understanding of $\mathscr{C}_+$ is the set $\mathscr{C}$, not the set of biquadratic forms vanishing on the diagonal.

Our first main result is a characterization of $\mathscr{C}$. Let $m = \binom{d}{2} = \dim \mathrm{Skew}(d)$. For any $H = (h_{pq}) \in \mathbb{S}^m$, define a map c_H by

$$c_H(x) = \sum_{p,q=1}^{m} h_{pq} \, D_p \, x x^\top D_q^\top, \tag{3.2}$$

where $D_1, \ldots, D_m$ are given by (1.5). It is clear that $c_H \in \mathscr{C}$, and it turns out that every element of $\mathscr{C}$ is in fact of this form. Naively one might then conjecture that the dimension of $\mathscr{C}$ equals $\dim \mathbb{S}^m = \binom{m+1}{2}$, the number of free parameters h_{pq} appearing in (3.2). However, this turns out to be incorrect. Indeed, there exist linear relations among the maps $D_p \, x x^\top D_q^\top$ in (3.2); to capture them, we consider the linear space

$$\mathscr{K} = \{H \in \mathbb{S}^m : c_H(x) \equiv 0\}. \tag{3.3}$$

The proof of the following theorem is provided in Section 5.1.

Theorem 3.3. *The set $\mathscr{C}$ and its dimension are given by*

$$\mathscr{C} = \{c_H : H \in \mathbb{S}^m\} \qquad \text{and} \qquad \dim \mathscr{C} = 2\binom{d}{4} + 3\binom{d}{3} + \binom{d}{2} = \frac{d^2(d^2-1)}{12}, \tag{3.4}$$

where $m = \binom{d}{2}$ and we set $\binom{d}{k} = 0$ for $k > d$. Moreover, $\dim \mathscr{K} = \dim \mathbb{S}^m - \dim \mathscr{C} = \binom{d}{4}$.

The linear space $\mathscr{K}$ has several interesting properties that are used in the proofs of our subsequent results. This is discussed in detail in Section 5.2. In a nutshell, the situation is the following: Each $H \in \mathbb{S}^m$ naturally corresponds to a quadratic form on the space of skew-symmetric matrices, which when evaluated at $A = xy^\top - yx^\top \in \mathrm{Skew}(2,d)$ yields precisely $\mathrm{BQ}(x,y)$; see (5.1). Now, it turns out that $\mathscr{K}$ admits a set of basis vectors that can be identified with the so-called Plücker polynomials from algebraic geometry. The zero set of these polynomials is isomorphic to $\mathrm{Skew}(2,d)$. As a consequence, the quotient $\mathbb{S}^m/\mathscr{K}$ can be identified with the set of quadratic forms acting on $\mathrm{Skew}(2,d)$. Thus, in view of Theorem 3.3, each element of $\mathscr{C}$ corresponds to a unique quadratic form on $\mathrm{Skew}(2,d)$. Finally, we mention that it is possible to explicitly write down a basis for $\mathscr{K}$; see Remark 5.4.

We now turn to the delicate question of positivity: $c(x) \succeq 0$ for all x, or equivalently $\mathrm{BQ}(x,y) \geq 0$ for all (x,y). A simple example where this holds is $c(x) = A\,xx^\top A^\top$, where $A \in \mathrm{Skew}(d)$. More generally, by taking conic combinations, elements of the form

$$c(x) = A_1\,xx^\top A_1^\top + \cdots + A_m\,xx^\top A_m^\top, \qquad A_1,\dots,A_m \in \mathrm{Skew}(d), \tag{3.5}$$

lie in $\mathscr{C}_+$. The corresponding biquadratic form is given by

$$\mathrm{BQ}(x,y) = (y^\top A_1 x)^2 + \cdots + (y^\top A_m x)^2,$$

which has the important property that it can be written as a sum of squares of polynomials (indeed, of quadratic forms). The converse is also true: whenever $\mathrm{BQ}(x,y)$ is a sum of squares, the corresponding $c(x)$ is of the form (3.5). Specifically, one has the following:

Lemma 3.4. *Let $c \in \mathscr{C}$, $m = \binom{d}{2}$. The following conditions are equivalent.*

(i) $\mathrm{BQ}(x,y) = y^\top c(x) y$ *is a sum of squares of polynomials.*

(ii) $c = c_H$ *for some* $H \in \mathbb{S}^m_+$.

(iii) $c(x) = \sum_{p=1}^m A_p\,xx^\top A_p^\top$ *for some* $A_1,\dots,A_m \in \mathrm{Skew}(d)$.

Proof. (iii) $\Longrightarrow$ (i): Obvious.

(i) $\Longrightarrow$ (ii): We first show that (i) implies that there exist $m' \in \mathbb{N}$ and $A_1,\dots,A_{m'} \in \mathrm{Skew}(d)$ such that $y^\top c(x) y = \sum_{p=1}^{m'} (y^\top A_p\,x)^2$. To this end, suppose $y^\top c(x) y = f_1(x,y)^2 + \cdots + f_{m'}(x,y)^2$ for some polynomials f_p, $p = 1,\dots,m'$, where $m' \in \mathbb{N}$. By setting $x = y = 0$, we see that $f_p(0,0) = 0$ for all p. Since also $f_p(sx,y)^2 \leq y^\top c(sx) y = s^2 y^\top c(x) y$, each f_p is linear in x. Similarly, each f_p is linear in y. It follows that $f_p(x,y) = y^\top A_p x$ for some matrix $A_p \in \mathbb{R}^{d\times d}$. Finally, $(x^\top A_p x)^2 \leq x^\top c(x) x \equiv 0$, whence A_p is skew-symmetric. This proves the claimed representation.

Now, since $D_1,\dots,D_m$ form a basis for $\mathrm{Skew}(d)$, for each p there exists $u_p = (u_p^1,\dots,u_p^m) \in \mathbb{R}^m$ such that $A_p = \sum_{i=1}^m u_p^i\,D_i$. Thus,

$$y^\top c(x) y = y^\top \Big(\sum_{p=1}^{m'} A_p\,xx^\top A_p^\top \Big) y = y^\top \Big(\sum_{i,j=1}^{m} h_{ij}\,D_i\,xx^\top D_j^\top \Big) y,$$

where $H = (h_{ij}) \in \mathbb{S}^m$ is given by $H = u_1 u_1^\top + \cdots + u_{m'} u_{m'}^\top$ and thus is positive semidefinite. We deduce (ii).

(ii) $\Longrightarrow$ (iii): Writing $H = u_1 u_1^\top + \cdots + u_m u_m^\top$ for some vectors u_p with components $(u_p^1, \ldots, u_p^m)$ and substituting into (3.2) yields

$$c_H(x) = \sum_{p=1}^{m} (u_p^1 D_1 + \cdots + u_p^m D_m)\, x x^\top (u_p^1 D_1 + \cdots + u_p^m D_m)^\top,$$

which is of the desired form with $A_p = \sum_{i=1}^m u_p^i D_i$. □

Lemma 3.4 shows that those $c \in \mathscr{C}_+$ whose associated biquadratic form is a sum of squares have a nice parameterization in terms of positive semidefinite matrices H. One is thus naturally led to ask whether every nonnegative biquadratic form vanishing on the diagonal can be written as a sum of squares. Our next main result addresses this question. Its proof is provided in Sections 5.3 and 5.4, relying on the material developed in Section 5.2. The proof also depends on a known result (Lemma 5.5) on existence of low-rank elements in the intersection of $\mathbb{S}^m_+$ with certain affine subspaces of $\mathbb{S}^m$.

Theorem 3.5. (i) *If $d \le 4$, then any nonnegative biquadratic form in $d + d$ variables vanishing on the diagonal is a sum of squares. Equivalently, any $c \in \mathscr{C}_+$ is of the form $c = c_H$ for some $H \in \mathbb{S}^m_+$.*

(ii) *If $d \ge 6$, then there exists a nonnegative biquadratic form in $d + d$ variables vanishing on the diagonal that is not a sum of squares. Equivalently, there exists $c \in \mathscr{C}_+$ that is not of the form $c = c_H$ for some $H \in \mathbb{S}^m_+$.*

Remark 3.6. *In László (2010) it was observed that the conclusion of Theorem 3.5*(i) *holds for $d \le 3$, while the case $d \ge 4$ was left open; see also László (2012). The case $d = 3$ can also be deduced from Quarez (2015). We have not been able to determine whether there exists a nonnegative biquadratic form in $5 + 5$ variables, vanishing on the diagonal, that is not a sum of squares. The counterexample that establishes Theorem 3.5*(ii) *comes from László (2010). For the benefit of the reader we recap this construction in Section 5.4 using the setting and notation of the present paper. The study of sum of squares representations of nonnegative biquadratic forms (not necessarily vanishing on the diagonal) goes back to Choi (1975).*

4 Consequences of the sum of squares property

Throughout this section, we assume that $a(x)$ and $b(x)$ satisfy (2.2)–(2.3) for some fixed $\alpha \in \mathbb{S}^d_+$, $c \in \mathscr{C}_+$, $b \in \mathbb{R}^d$, $B \in \mathbb{R}^{d \times d}$, and let $\mathrm{BQ}(x, y) = y^\top c(x) y$ be the associated nonnegative biquadratic form. Theorem 2.1 guarantees that a $\mathscr{B}^d$-valued weak solution to (1.2) exists for a suitable choice of σ and for any initial condition $x \in \mathscr{B}^d$. Denote its law by $\mathbb{P}_x$, and observe that it is uniquely determined by a, b, and x; see Lemma 2.4. In this section we discuss consequences of the existence of a sum of squares representation of $\mathrm{BQ}(x, y)$.

4.1 SDE representation

Our first result shows that $\mathbb{P}_x$ can be represented as the law of the solution to a certain structured SDE if and only if $\mathrm{BQ}(x,y)$ is a sum of squares. Before stating it, we give the following definition:

Definition 4.1. *A process Y is called a* $\mathrm{Skew}(d)$-valued correlated Brownian motion with drift *if it is of the form*

$$Y_t = A_0\, t + A_1 \widehat{W}_t^1 + \cdots + A_m \widehat{W}_t^m, \tag{4.1}$$

where $A_0, \ldots, A_m \in \mathrm{Skew}(d)$ and $\widehat{W} = (\widehat{W}^1, \ldots \widehat{W}^m)$ is an m-dimensional Brownian motion.

Theorem 4.2. *The following conditions are equivalent:*

(i) $\mathrm{BQ}(x,y) = y^\top c(x) y$ *is a sum of squares of polynomials.*

(ii) *For each $x \in \mathscr{B}^d$, $\mathbb{P}_x$ is the law of the unique $\mathscr{B}^d$-valued weak solution to the SDE*

$$dX_t = (b + \widehat{B} X_t)\, dt + \sqrt{1 - \|X_t\|^2}\, \alpha^{1/2}\, dW_t + (\circ\, dY_t) X_t \tag{4.2}$$

with initial condition $X_0 = x$, where $-\widehat{B} \in \mathbb{S}_+^d$, $W = (W^1, \ldots, W^d)$ is a d-dimensional Brownian motion, and Y is a $\mathrm{Skew}(d)$*-valued correlated Brownian motion with drift, independent of W. Here $\circ\, dY_t$ denotes Stratonovich differential.*

In this case, B is related to $\widehat{B}$ and $A_0, \ldots, A_m$ in (4.1) *through*

$$\frac{1}{2}(B - B^\top) = A_0 \qquad \text{and} \qquad \frac{1}{2}(B + B^\top) = \widehat{B} - \frac{1}{2}\sum_{p=1}^m A_p^\top A_p. \tag{4.3}$$

Proof. (i) $\Longrightarrow$ (ii): Lemma 3.4 implies that

$$c(x) = \sum_{p=1}^m A_p\, x x^\top A_p^\top \tag{4.4}$$

for some $A_1, \ldots, A_m \in \mathrm{Skew}(d)$. Define $A_0 \in \mathrm{Skew}(d)$ and $\widehat{B} \in \mathbb{S}^d$ via (4.3). We claim that $-\widehat{B} \in \mathbb{S}_+^d$. Indeed, writing the condition (2.3) in terms of $\widehat{B}$, and using that $x^\top A_0 x = 0$ by skew-symmetry, yields

$$b^\top x + x^\top \widehat{B} x \le 0 \quad \text{for all} \quad x \in \mathscr{S}^{d-1}. \tag{4.5}$$

Inserting x and $-x$ into this inequality and adding, one obtains $x^\top \widehat{B} x \le 0$ for all $x \in \mathscr{S}^{d-1}$, whence $-\widehat{B} \in \mathbb{S}_+^d$ as claimed. Next, writing the SDE (4.2) in Itô form yields

$$dX_t = (b + BX_t)\, dt + \widehat{\sigma}(X_t) \begin{pmatrix} dW_t \\ d\widehat{W}_t \end{pmatrix},$$

where

$$\widehat{\sigma}(x) = \begin{pmatrix} \sqrt{1 - \|x\|^2}\, \alpha^{1/2} & A_1 x & \cdots & A_m x \end{pmatrix} \in \mathbb{R}^{d \times (d+m)}. \tag{4.6}$$

In view of (4.4) we have $\widehat{\sigma}(x)\widehat{\sigma}(x)^\top = (1-\|x\|^2)\alpha + c(x) = a(x)$. Thus (4.2) has the same drift and diffusion coefficients as (1.2), which by Lemma 2.4 implies that the law of any solution to (4.2) with initial condition x is indeed $\mathbb{P}_x$, as desired. Finally, the existence of a $\mathscr{B}^d$-valued solution to (4.2) follows, as in the proof of Theorem 2.1, from Example 2.5 and Theorem 2.2 in Da Prato and Frankowska (2007).

(ii) $\Longrightarrow$ (i): Let $\widehat{\sigma}(x)$ be given by (4.6). Since $\mathbb{P}_x$ is the uniquely determined law of the solution to (4.2), one obtains

$$a(x) = \widehat{\sigma}(x)\widehat{\sigma}(x)^\top = (1 - \|x\|^2)\alpha + \sum_{p=1}^{m} A_p\, xx^\top A_p^\top.$$

Since $a(x) = (1 - \|x\|^2)\alpha + c(x)$, it follows that (4.4) holds. Lemma 3.4 then implies (i). $\square$

The terms on the right-hand side of (4.2) have natural interpretations. Since $\widehat{B}$ is symmetric with real non-positive eigenvalues, the drift has a mean-reverting component $\widehat{B}X_t$ in addition to the constant part b. The term involving W_t induces diffusive motion along the eigenvectors of α, with quadratic variation that is scaled by the squared distance $1 - \|X_t\|^2$ to the boundary of $\mathscr{B}^d$. Finally, the term $(\circ\, dY_t)X_t$ can be viewed as an infinitesimal random rotation of X_t; see also Price and Williams (1983) and Van den Berg and Lewis (1985). Heuristically, the Stratonovich increment $\circ\, dY_t$ is a small random Skew-symmetric matrix that maps X_t to an infinitesimal tangent vector of the ball with radius $\|X_t\|$.

Remark 4.3. *In the setting of Theorem 4.2,* (4.3) *implies that the necessary and sufficient condition* (2.5) *for boundary attainment reduces to*

$$b^\top x + x^\top(\widehat{B} + \alpha)x \leq 0 \quad \textit{for all} \quad x \in \mathscr{S}^{d-1}.$$

If X takes values in $\mathscr{S}^{d-1}$, the sum of squares property of $\mathrm{BQ}(x, y)$ has stronger implications. In particular, X can now be constructed as the *unique strong* solution to a suitable SDE.

Corollary 4.4. *Assume the coefficients a and b satisfy Theorem 2.3*(ii)*. The following conditions are equivalent:*

(i) $\mathrm{BQ}(x, y) = y^\top c(x) y$ *is a sum of squares of polynomials.*

(ii) *For each $x \in \mathscr{S}^{d-1}$, $\mathbb{P}_x$ is the law of the $\mathscr{S}^{d-1}$-valued unique strong solution to the SDE*

$$dX_t = (\circ\, dY_t)X_t \tag{4.7}$$

with initial condition $X_0 = x$, where Y is a $\mathrm{Skew}(d)$*-valued correlated Brownian motion with drift.*

(iii) *The operator $\mathscr{G} = \frac{1}{2}\,\mathrm{Tr}(a\,\nabla^2) + b^\top\nabla$ corresponding to the SDE* (1.2) *can be expressed in Hörmander form as*

$$\mathscr{G} = V_0 + \frac{1}{2}\sum_{p=1}^{m} V_p^2, \tag{4.8}$$

where for $p = 0, \ldots, m$, V_p is the linear vector field given by $V_p(x) = A_p x$ for some $A_p \in \mathrm{Skew}(d)$.

Proof. (i) $\iff$ (ii) is deduced from the corresponding equivalence in Theorem 4.2. Indeed, the forward implication follows since $b = 0$, $\alpha = 0$, and $\widehat{B} = 0$. Here the latter is due to (2.10) and (4.3), which give

$$x^\top \widehat{B} x = x^\top B x + \frac{1}{2} \sum_{p=1}^{m} x^\top A_p A_p^\top x = x^\top B x + \frac{1}{2} \operatorname{Tr}(c(x)) = 0, \qquad x \in \mathbb{R}^d.$$

Note that the form of $c(x)$ follows from Lemma 3.4. For the converse, simply note that (ii) implies that Theorem 4.2(ii) is satisfied with $b = 0$, $\alpha = 0$, and $\widehat{B} = 0$. This yields (i). The implication (ii) $\implies$ (iii) is well-known, and follows from a brief calculation using the identity $V_p^2 f(x) = \operatorname{Tr}(\nabla^2 f(x) A_p \, x x^\top A_p^\top) - (A_p^\top A_p \, x)^\top \nabla f(x)$. Finally, the implication (iii) $\implies$ (i) is again a calculation showing that $c(x) = A_1 \, x x^\top A_1^\top + \cdots + A_m \, x x^\top A_m^\top$, which gives (i) via Lemma 3.4. $\square$

Example 4.5. *The classical Brownian motion on the sphere is covered by Corollary 4.4. In particular, it is a polynomial preserving diffusion. Indeed, by taking $A_p = D_p$ for $p = 1, \ldots, m$, one obtains*

$$c(x) = \sum_{i<j} (e_i e_j^\top - e_j e_i^\top) x x^\top (e_j e_i^\top - e_i e_j^\top) = (x^\top x) \mathrm{Id} - x x^\top,$$

so that $a(x) = c(x) = \mathrm{Id} - x x^\top$ on $\mathscr{S}^{d-1}$. This is the orthogonal projection onto the tangent space of $\mathscr{S}^{d-1}$ at x. Setting $A_0 = 0$, we thus recover the following well-known quadratic SDE for Brownian motion on the sphere:

$$dX_t = (\mathrm{Id} - X_t X_t^\top) \circ dW_t.$$

The linear SDE in Corollary 4.4 is different, and was originally obtained by Price and Williams (1983); see also Van den Berg and Lewis (1985).

4.2 Pathwise uniqueness

It is natural to ask whether polynomial preserving diffusions on the unit ball can be realized as solutions to SDEs for which pathwise uniqueness holds. As mentioned in Remark 2.5, if $a(x)$ is positive definite on the interior of $\mathscr{B}^d$, one obtains pathwise uniqueness up to the first hitting time of $\mathscr{S}^{d-1}$ by taking $\sigma(x) = a(x)^{1/2}$.

If $\mathrm{BQ}(x, y)$ is a sum of squares, more can be said. Indeed, in this case Theorem 4.2 yields pathwise uniqueness up to the first hitting time of $\mathscr{S}^{d-1}$, regardless of whether or not $a(x)$ is nonsingular. Moreover, if the state space is $\mathscr{S}^{d-1}$, Corollary 4.4 shows that pathwise uniqueness always holds (still under the sum of squares assumption, of course).

Looking at the SDE (4.2) in Theorem 4.2, there may be some hope that pathwise uniqueness holds globally, not just up the first hitting time of $\mathscr{S}^{d-1}$. We have not been able to prove this in general. Instead, we present a partial result relying on the method developed by DeBlassie (2004) for proving pathwise uniqueness for certain SDEs on the unit ball; see also Swart (2002). Specifically, we consider the following special case of (4.2):

$$dX_t = -\kappa X_t \, dt + \nu \sqrt{1 - \|X_t\|^2} \, dW_t + A_0 X_t \, dt + A_1 X_t \circ d\widehat{W}_t^1 + \cdots + A_m X_t \circ d\widehat{W}_t^m, \quad (4.9)$$

where $A_0, \dots, A_m \in \mathrm{Skew}(d)$, $W = (W^1, \dots, W^d)$ and $\widehat{W}^1, \dots, \widehat{W}^m$ are independent Brownian motions, and, crucially, κ and ν are positive scalar constants. Within the polynomial preserving class, this is a generalization of the equation considered by Swart (2002) and DeBlassie (2004), who take the tangential components $A_0, \dots, A_m$ to be zero. Note that even in this case, it is not known whether pathwise uniqueness holds for (4.9) for arbitrary values of κ and ν.

The reason why DeBlassie (2004) method can be applied in the present situation is contained in the following calculation. Define $Y_t = 1 - \|X_t\|^2$. Then, an application of Itô's formula and the change-of-variable formula for the Stratonovich integral yield

$$dY_t = 2\kappa\|X_t\|^2\,dt - 2\nu\sqrt{Y_t}X_t^\top dW_t - d\,\nu^2 Y_t\,dt - 2X_t^\top A_0 X_t\,dt - 2\sum_{p=1}^m X_t^\top A_p X_t\,d\widehat{W}_t^p.$$

The skew-symmetry of the A_p implies that the quadratic terms $X_t^\top A_p X_t$ all vanish, so that

$$dY_t = \bigl(2\kappa\|X_t\|^2 - d\,\nu^2 Y_t\bigr)\,dt - 2\nu\sqrt{Y_t}X_t^\top dW_t. \tag{4.10}$$

This no longer involves the matrices A_p. Note that this crucially relies on the linearity in X_t of the last $m+1$ terms of (4.9). This property is specific for the sum of squares case (4.2), and is in general not present in (1.2).

Theorem 4.6. *Assume $\kappa/\nu^2 > \sqrt{2} - 1$. Then pathwise uniqueness holds for* (4.9).

Proof. Let X and $\widetilde{X}$ be two solutions to (4.9), driven by the same Brownian motions and starting from the same starting point $x_0 \in \mathscr{B}^d$. Pathwise uniqueness clearly holds up to the first time the boundary is attained, so it suffices to take x_0 on the boundary, and prove $X_t = \widetilde{X}_t$ for all $t \le \tau$, where $\tau = \inf\{t \ge 0\colon \|X_t\| \wedge \|\widetilde{X}_t\| \le 1 - \varepsilon\}$ for some arbitrary $\varepsilon > 0$. Define $Y_t = 1 - \|X_t\|^2$ and $\widetilde{Y}_t = 1 - \|\widetilde{X}_t\|^2$. In order to improve readability we often omit time subscripts below. We also omit the details of several cumbersome but straightforward calculations.

We apply the method by DeBlassie (2004). Define the function

$$F(y, \widetilde{y}) = (y^p - \widetilde{y}^p)^2$$

where $p \in (1/2, 1)$ is required to satisfy $\kappa > \nu^2(1-p)$. It then follows from Lemma 2.1 in DeBlassie (2004) together with (4.10) that $F(Y, \widetilde{Y})$ satisfies Itô's formula on the time interval $[0, \tau]$, provided $\varepsilon > 0$ is small enough depending on κ, ν, and p. Writing $F_1 = \partial_y F(Y, \widetilde{Y})$, $F_2 = \partial_{\widetilde{y}} F(Y, \widetilde{Y})$, etc., this means that the following calculations are rigorous. First,

$$dF(Y, \widetilde{Y}) = F_1\,dY + F_2\,d\widetilde{Y} + \frac{1}{2}F_{11}\,d\langle Y, Y\rangle + F_{12}\,d\langle Y, \widetilde{Y}\rangle + \frac{1}{2}F_{22}\,d\langle \widetilde{Y}, \widetilde{Y}\rangle.$$

Next, in view of (4.10), a calculation yields

$$\begin{aligned} dF(Y,\widetilde{Y}) &= (\text{local martingale}) \\ &\quad + \Big(\kappa(F_1+F_2) + \nu^2(F_{11}Y + 2F_{12}Y^{1/2}\widetilde{Y}^{1/2} + F_{22}\widetilde{Y})\Big)(\|X\|^2 + \|\widetilde{X}\|^2)\,dt \\ &\quad + \Big(\kappa(F_1-F_2) + \nu^2(F_{11}Y - F_{22}\widetilde{Y})\Big)(\|X\|^2 - \|\widetilde{X}\|^2)\,dt \\ &\quad - 2\nu^2 F_{12}Y^{1/2}\widetilde{Y}^{1/2}\|X-\widetilde{X}\|^2\,dt - d\,\nu^2(F_1Y + F_2\widetilde{Y})\,dt \\ &= (\text{local martingale}) + \mathbf{I}_1(\|X\|^2 + \|\widetilde{X}\|^2)\,dt + \mathbf{I}_2\,dt + \mathbf{I}_3\,dt + \mathbf{I}_4\,dt. \end{aligned}$$

We now insert the explicit expression for F and its derivatives. Define $Z = (Y^{p-1} - \widetilde{Y}^{p-1})(\widetilde{Y}^p - Y^p)$ and note that Lemma A.1 in DeBlassie (2004) yields

$$(Y^{p-1/2} - \widetilde{Y}^{p-1/2})^2 \le \frac{(2p-1)^2}{4p(1-p)}Z$$

whenever $p \in (\frac{1}{2}, 1)$. Making use of this inequality, one obtains

$$\mathbf{I}_1 = -2p\kappa Z + 2p\nu^2(1-p)Z + 2p^2\nu^2(Y^{p-1/2} - \widetilde{Y}^{p-1/2})^2 \le -2p\nu^2\left(\frac{\kappa}{\nu^2} - 1 + p - \frac{(2p-1)^2}{4(1-p)}\right)Z.$$

The expression in parentheses on the right-hand side is maximized by $p = 1 - \sqrt{2}/4 \in (\frac{1}{2}, 1)$, which makes its value equal to $\kappa/\nu^2 + 1 - \sqrt{2}$. Note that the hypothesis on κ and ν ensures that $\kappa > \nu^2(1-p)$ holds for this choice of p, as was required above. Thus, on $[0, \tau]$,

$$\mathbf{I}_1(\|X\|^2 + \|\widetilde{X}\|^2) \le -4(1-\varepsilon)p\nu^2\left(\frac{\kappa}{\nu^2} + 1 - \sqrt{2}\right)Z.$$

Next, we have

$$\mathbf{I}_2 = -2p\nu^2\left(\frac{\kappa}{\nu^2} - 1 + p\right)(Y^{p-1} + \widetilde{Y}^{p-1})(Y^p - \widetilde{Y}^p)(Y - \widetilde{Y}) - 2p^2\nu^2(Y^{2p-1} - \widetilde{Y}^{2p-1})(Y - \widetilde{Y}).$$

Since $\kappa/\nu^2 \ge 1 - p$, one obtains $\mathbf{I}_2 \le 0$. Furthermore, since

$$-F_{12}Y^{1/2}\widetilde{Y}^{1/2} = 2p^2 Y^{p-1/2}\widetilde{Y}^{p-1/2} \le 2p^2\varepsilon^{2p-1}$$

for $t \le \tau$, we get $\mathbf{I}_3 \le 4p^2\varepsilon^{2p-1}\nu^2\|X - \widetilde{X}\|^2$ on $[0, \tau]$. Finally, $\mathbf{I}_4 = -2pd\,\nu^2 F \le 0$. Putting together these estimates yields, for $t \le \tau$,

$$\begin{aligned} F(Y_t, \widetilde{Y}_t) \le (\text{local martingale}) &- 4(1-\varepsilon)p\nu^2\left(\frac{\kappa}{\nu^2} + 1 - \sqrt{2}\right)\int_0^t Z_s\,ds \\ &+ 4p^2\varepsilon^{2p-1}\nu^2\int_0^t \|X_s - \widetilde{X}_s\|^2\,ds. \end{aligned} \tag{4.11}$$

We need the dynamics of $\|X - \widetilde{X}\|^2$. Similar calculations as those leading to (4.10) yield

$$d\|X-\widetilde{X}\|^2 = \Big(-2\kappa\|X-\widetilde{X}\|^2 + d\,\nu^2(Y^{1/2} - \widetilde{Y}^{1/2})^2\Big)\,dt + 2\nu(Y^{1/2} - \widetilde{Y}^{1/2})(X-\widetilde{X})^\top dW.$$

Moreover, one has the inequality $(Y^{1/2}-\widetilde{Y}^{1/2})^2 \le c\varepsilon^{2-2p}Z$ on $[0,\tau]$, where c is a constant that only depends on p; see the proof of (DeBlassie, 2004, Lemma 3.6). In conjunction with (4.11) one then obtains, for $t \le \tau$,

$$\begin{aligned} F(Y_t,\widetilde{Y}_t) + \|X_t - \widetilde{X}_t\|^2 \le {} & \text{(local martingale)} \\ & + \Big(d\,\nu^2 c\varepsilon^{2-2p} - 4(1-\varepsilon)p\nu^2 \left(\frac{\kappa}{\nu^2} + 1 - \sqrt{2}\right)\Big) \int_0^t Z_s\,ds \\ & + (4p^2\varepsilon^{2p-1}\nu^2 - 2\kappa) \int_0^t \|X_s - \widetilde{X}_s\|^2\,ds. \end{aligned}$$

Since $\kappa/\nu^2 + 1 - \sqrt{2} > 0$ by assumption, we may choose ε sufficiently small to make the finite variation terms nonpositive. Thus, on $[0,\tau]$, $F(Y_t,\widetilde{Y}_t) + \|X_t - \widetilde{X}_t\|^2$ is bounded from above by a nonnegative local martingale M_t with $M_0 = 0$. Thus $M_t = 0$ on $[0,\tau]$, implying that the same holds for the left-hand side. This yields $X = \widetilde{X}$ on $[0,\tau]$, as desired. The proof is complete. □

Remark 4.7. *While the restriction on κ and ν excludes some specifications, there is certainly an overlap with the set of parameters for which the boundary is attained. Indeed, by Proposition 2.2 the boundary may be attained if and only if $\kappa/\nu^2 < 1$.*

4.3 Existence of smooth densities

If $\mathrm{BQ}(x,y)$ is a sum of squares and the state space is the unit sphere, Corollary 4.4(iii) and Hörmander's theorem lead to simple conditions under which the solution to (1.2) possesses a smooth density. To state the precise result we introduce some notation. Let $A_0,\dots,A_m$ be elements of $\mathrm{Skew}(d)$ and fix $x_0 \in \mathscr{S}^{d-1}$. Define

$$\begin{aligned} \mathfrak{g}_0 &= \{A_1,\dots,A_m\} \\ \mathfrak{g}_k &= \mathfrak{g}_{k-1} \cup \{[B,A_p]\colon B \in \mathfrak{g}_{k-1},\ p = 0,\dots,m\} \qquad (k \ge 1) \\ \mathfrak{g} &= \operatorname{span} \bigcup_{k\ge 0} \mathfrak{g}_k \\ \mathfrak{h} &= \text{Lie algebra generated by } A_0,\dots,A_m. \end{aligned} \tag{4.12}$$

Here $[A,B] = AB - BA$ is the usual matrix commutator. Note that $\mathfrak{g}$ is a Lie sub-algebra of $\mathfrak{h}$, and that $\mathfrak{h} = \mathfrak{g} + \mathbb{R}A_0$. In fact, $\mathfrak{g}$ is even an ideal in $\mathfrak{h}$, meaning that $[\mathfrak{g},\mathfrak{h}] \subseteq \mathfrak{g}$. Let $\boldsymbol{G}$ denote the connected Lie subgroup of $SO(d)$ generated by e^B, $B \in \mathfrak{g}$, and $\boldsymbol{H}$ the connected Lie group similarly generated by $\mathfrak{h}$. Then $\mathfrak{g}$ (respectively $\mathfrak{h}$) is the Lie algebra of $\boldsymbol{G}$ (respectively $\boldsymbol{H}$). Finally, define $\boldsymbol{G}x_0 = \{Qx_0\colon Q \in \boldsymbol{G}\}$, the orbit of x_0 under $\boldsymbol{G}$, and similarly $\boldsymbol{H}x_0$. These are smooth submanifolds of $\mathscr{S}^{d-1}$, and their tangent spaces are given by

$$T_x(\boldsymbol{G}x_0) = \mathfrak{g}x \qquad \text{and} \qquad T_x(\boldsymbol{H}x_0) = \mathfrak{h}x.$$

In view of the identities $Q^{-1}\mathfrak{g}Q = \mathfrak{g}$ and $Q^{-1}\mathfrak{h}Q = \mathfrak{h}$ for all $Q \in \boldsymbol{G}$, it follows that

$$T_{Qx}(\boldsymbol{G}x_0) = Q\,T_x(\boldsymbol{G}x_0) \quad \text{and} \quad T_{Qx}(\boldsymbol{H}x_0) = Q\,T_x(\boldsymbol{H}x_0), \qquad Q \in \boldsymbol{G}. \tag{4.13}$$

All these properties are well-known, and simply reflect the fact that $\boldsymbol{G}x_0$ and $\boldsymbol{H}x_0$ are homogeneous spaces for $\boldsymbol{G}$ and $\boldsymbol{H}$, respectively.

The natural state space for solutions to (4.7) starting from $x_0 \in \mathscr{S}^{d-1}$ is $\boldsymbol{H}x_0$. However, by slightly adjusting such a solution one obtains a process which remains in $\boldsymbol{G}x_0$. Specifically, one has the following lemma.

Lemma 4.8. *Let X be the solution to (4.7) with $X_0 = x_0$, and define a process Z by*

$$Z_t = e^{-tA_0} X_t.$$

Then $Z_t \in \boldsymbol{G}x_0$ for all $t \geq 0$.

Proof. Since $\mathfrak{g}$ is an ideal in $\mathfrak{h}$, we have for each $p \in \{1, \ldots, m\}$,

$$e^{-tA_0} A_p e^{tA_0} z \in \mathfrak{g}z = T_z(\boldsymbol{G}x_0) \quad \text{for all} \quad z \in \boldsymbol{G}x_0, \quad t \geq 0. \tag{4.14}$$

To see this, write $f(t) = e^{-tA_0} A_p e^{tA_0}$. Then $f^{(k+1)}(0) = [f^{(k)}(0), A_0]$ for all $k \in \mathbb{N}$. Since $\mathfrak{g}$ is an ideal in $\mathfrak{h}$ one has $f'(0) = [A_p, A_0] \in \mathfrak{g}$, whence $f^{(k)}(0) \in \mathfrak{g}$ for all $k \in \mathbb{N}$ by induction. Thus $e^{-tA_0} A_p e^{tA_0} = f(t) \in \mathfrak{g}$ since f is analytic with infinite radius of convergence. Hence (4.14) follows. Now, a brief calculation shows that Z satisfies the time-inhomogeneous SDE

$$dZ_t = e^{-tA_0} A_1 e^{tA_0} Z_t \circ dW_t^1 + \cdots + e^{-tA_0} A_m e^{tA_0} Z_t \circ dW_t^m,$$

which admits a unique strong solution. In view of (4.14), the vector fields on the right-hand side are tangent to $\boldsymbol{G}x_0$, which by Hsu (2002, Theorem 1.2.9) implies that Z takes values there. □

One now has the following fairly precise result regarding existence of smooth densities in the case where $\mathrm{BQ}(x, y)$ admits a sum of squares representation.

Theorem 4.9. *Let X be the solution to (4.7) with $X_0 = x_0 \in \mathscr{S}^{d-1}$; in particular, $\mathrm{BQ}(x, y)$ is a sum of squares of polynomials. Then X takes values in $\boldsymbol{H}x_0$. Moreover, the following conditions are equivalent:*

(i) *X_t has a smooth density with respect to surface area measure on $\boldsymbol{H}x_0$ for all $t > 0$.*

(ii) *$\boldsymbol{H}x_0 = \boldsymbol{G}x_0$.*

(iii) *$A_0 x_0 \in \mathfrak{g}x_0$.*

Proof. It follows from Lemma 4.8 and the definition of $\boldsymbol{H}$ that X takes values in $\boldsymbol{H}x_0$. We now prove that the stated conditions are equivalent. To this end, first note that since $\boldsymbol{G}x_0 \subseteq \boldsymbol{H}x_0$, and $\boldsymbol{H}x_0$ is connected due to the connectedness of $\boldsymbol{H}$, one has $\boldsymbol{G}x_0 = \boldsymbol{H}x_0$ if and only if $T_x(\boldsymbol{G}x_0) = T_x(\boldsymbol{H}x_0)$ for all $x \in \boldsymbol{G}x_0$. Thanks to (4.13) this yields

$$\begin{aligned} \boldsymbol{G}x_0 = \boldsymbol{H}x_0 \quad &\Longleftrightarrow \quad T_x(\boldsymbol{G}x_0) = T_x(\boldsymbol{H}x_0) \text{ for some } x \in \boldsymbol{G}x_0. \\ &\Longleftrightarrow \quad T_{x_0}(\boldsymbol{G}x_0) = T_{x_0}(\boldsymbol{H}x_0). \end{aligned} \tag{4.15}$$

The latter condition is equivalent to $\mathfrak{g}x_0 = \mathfrak{h}x_0$. Since $\mathfrak{h} = \mathfrak{g} + \mathbb{R}A_0$, this holds if and only if $A_0x_0 \in \mathfrak{g}x_0$. This proves (ii) $\iff$ (iii).

To prove (i) $\implies$ (ii), let σ denote surface area measure on $\boldsymbol{H}x_0$ and note that for any measurable subset $U \subseteq \boldsymbol{H}x_0$, one has $\sigma(U) > 0$ if and only if $\sigma(e^{tA_0}U) > 0$, where $t \in \mathbb{R}$ is arbitrary. Fix some $t > 0$. Since Lemma 4.8 yields $X_t \in e^{tA_0}\boldsymbol{G}x_0$, it follows that $\sigma(e^{tA_0}\boldsymbol{G}x_0) > 0$ and hence $\sigma(\boldsymbol{G}x_0) > 0$. This in turn implies that there exists some $x \in \boldsymbol{G}x_0$ for which the tangent space $T_x(\boldsymbol{G}x_0)$ cannot be a proper subspace of $T_x(\boldsymbol{H}x_0)$; that is, $T_x(\boldsymbol{G}x_0) = T_x(\boldsymbol{H}x_0)$. Thus (ii) follows by (4.15)

It remains to prove (ii) $\implies$ (i). Consider the vector fields V_p given by $V_p(x) = A_px$, appearing in the expression (4.8) for the generator $\mathscr{G}$. One has $[V_p, V_q](x) = [A_p, A_q]x$, where $[A_p, A_q] = A_pA_q - A_qA_p$. Hörmander's parabolic condition (see Hörmander (1967), or Theorem 1.3 in Hairer (2011) for a formulation suited to our current setting) is therefore equivalent to requiring $T_x(\boldsymbol{H}x_0) = \mathfrak{g}x$ for all $x \in \boldsymbol{H}x_0$. This holds thanks to the assumption $\boldsymbol{H}x_0 = \boldsymbol{G}x_0$. Applying Hörmander's theorem in each smooth coordinate chart of $\boldsymbol{H}x_0$ now gives the existence of a smooth density. □

Corollary 4.10. *Let X be the solution to (4.7) with $X_0 = x_0 \in \mathscr{S}^{d-1}$. Then X_t has a smooth density with respect to surface area measure on $\mathscr{S}^{d-1}$ for all $t > 0$ if and only if $\mathfrak{g} = \mathrm{Skew}(d)$.*

Example 4.11. *As a simple illustration of the distinction between $\boldsymbol{G}$ and $\boldsymbol{H}$, let $d = 4$ and consider the skew-symmetric matrices*

$$A_0 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad A_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}.$$

Then we have $A_0A_1 = A_1A_0 = 0$, so that $\mathfrak{g} = \mathrm{span}\{A_1\}$ and $\mathfrak{h} = \mathrm{span}\{A_0, A_1\}$. Moreover, $e^{-tA_0}A_1e^{tA_0} = A_1$, whence the process Z in Lemma 4.8 satisfies $dZ_t = A_1Z_t \circ dW_t^1$. It follows that $(Z_t^3)^2 + (Z_t^4)^2$ is constant, showing that Z is confined to a circle embedded in $\mathbb{R}^4$, namely $\boldsymbol{G}x_0 = e^{\mathbb{R}A_1}x_0$. On the other hand, the process X given by $X_t = e^{tA_0}Z_t$ is not restricted to any fixed subset of $\boldsymbol{H}x_0$, although for a fixed time $t \geq 0$ it necessarily lies in the transformed circle $e^{tA_0}\boldsymbol{G}x_0$ (we suppose that $A_0x_0 \neq 0$ to avoid a trivial situation). This is a nullset in $\boldsymbol{H}x_0$, so a density cannot exist.

By means of a projection argument, Corollary 4.10 can be used to derive conditions that guarantee the existence of smooth densities for a particular class of polynomial preserving diffusions taking values in the ball $\mathscr{B}^d$.

Corollary 4.12. *Let X be the solution to (4.2) with $X_0 = x_0 \in \mathscr{S}^{d-1}$ and $\widehat{B} = -\frac{1}{2}\alpha$. Write $\alpha = a_1a_1^\top + \cdots + a_da_d^\top$ for some vectors $a_1, \ldots, a_d \in \mathbb{R}^d$, and define matrices $\widetilde{A}_p \in \mathrm{Skew}(d+1)$ by*

$$\widetilde{A}_p = \begin{pmatrix} A_p & 0 \\ 0 & 0 \end{pmatrix} \text{ for } p = 0, \ldots, m, \quad \text{and} \quad \widetilde{A}_p = \begin{pmatrix} 0 & a_i \\ -a_i^\top & 0 \end{pmatrix} \text{ for } p = m + i,\ i = 1, \ldots, d.$$

Let $\widetilde{\mathfrak{g}}$ be defined as in (4.12) *with A_p replaced by $\widetilde{A}_p$. If $\widetilde{\mathfrak{g}} = \mathrm{Skew}(d+1)$, then for all $t > 0$, X_t has a density that is smooth on the interior of $\mathscr{B}^d$.*

Proof. Let $z_0 = (x_0, \sqrt{1 - \|x_0\|^2}) \in \mathscr{S}^d$, and define the $\mathscr{S}^d$-valued process Z as the unique solution to the SDE

$$dZ_t = \Big(\widetilde{A}_0 dt + (\circ\, d\widetilde{W}_t^1)\,\widetilde{A}_1 + \cdots (\circ\, d\widetilde{W}_t^{m+d})\,\widetilde{A}_{m+d}\Big)\, Z_t, \qquad Z_0 = z_0,$$

where $\widetilde{W} = (\widetilde{W}^1, \dots, \widetilde{W}^{m+d})$ is an $(m+d)$-dimensional Brownian motion. Let

$$\pi : \mathscr{S}^d \to \mathbb{R}^d, \quad (z_1, \dots, z_d, z_{d+1}) \mapsto (z_1, \dots, z_d)$$

be the canonical projection of $\mathscr{S}^d$ onto $\mathbb{R}^d$, and define $X' = \pi(Z)$. Then X' satisfies

$$dX'_t = A_0 X'_t\, dt + Z_{d+1,t} \sum_{i=1}^{d} a_i \circ d\widetilde{W}_t^{m+i} + \Big(\sum_{p=1}^{m} A_p \circ d\widetilde{W}_t^p\Big) X'_t.$$

Moreover, since

$$dZ_{d+1,t} = -\sum_{i=1}^{d} a_i^\top X'_t \circ d\widetilde{W}_t^{m+i},$$

one has $\langle Z_{d+1}, \widetilde{W}^{m+i}\rangle_t = -a_i^\top X'_t dt$. Thus, writing the middle term in the above expression for dX'_t in Itô form, one obtains

$$dX'_t = \Big(A_0 - \frac{1}{2}\alpha\Big) X'_t\, dt + Z_{d+1,t} \sum_{i=1}^{d} a_i\, d\widetilde{W}_t^{m+i} + \Big(\sum_{p=1}^{m} A_p \circ d\widetilde{W}_t^p\Big) X'_t.$$

Note that $Z_{d+1,t} = \sqrt{1 - \|X'_t\|^2}\,\mathrm{sign}(Z_{d+1,t})$. Thus X' satisfies an SDE of the form (1.2), and its generator coincides with that of X. Thus X and X' have the same law by Lemma 2.4.

Next, let σ be surface area measure on $\mathscr{S}^d$. By Corollary 4.10, Z_t has a smooth density $p_Z(t; z)$ with respect to σ for all $t > 0$. We claim that this implies that X'_t has a density, smooth on the interior of $\mathscr{B}^d$, for every $t > 0$. Since X and X' have the same law, this will complete the proof of the corollary.

Let $\mathscr{S}^d_+ = \{z \in \mathscr{S}^d \colon z_{d+1} > 0\}$ denote the upper hemisphere of $\mathscr{S}^d$. Let $\mu = \sigma|_{\mathscr{S}^d_+} \circ \pi^{-1}$ be the pushforward under π of the restriction of σ to $\mathscr{S}^d_+$, and recall that one has

$$\mu(dx) = (1 - \|x\|^2)^{-1/2}\, dx, \qquad x \in \mathscr{B}^d.$$

For any measurable subset $A \subseteq \mathscr{B}^d$ we compute:

$$\begin{aligned}
\mathbb{P}(X'_t \in A) = \mathbb{P}(Z_t \in \pi^{-1}(A)) &= \int_{\pi^{-1}(A)} p_Z(t; z)\sigma(dz) \\
&= \int_{\mathscr{S}^d_+ \cap \pi^{-1}(A)} \big(p_Z(t; z_1, \dots, z_d, z_{d+1}) + p_Z(t; z_1, \dots, z_d, -z_{d+1})\big)\sigma(dz) \\
&= \int_A \Big(p_Z(t; x, \sqrt{1 - \|x\|^2}) + p_Z(t; x, -\sqrt{1 - \|x\|^2})\Big)\, \mu(dx).
\end{aligned}$$

Thus X'_t, hence X_t, has a density $p_X(t;x)$ given by

$$p_X(t;x) = \Big(p_Z(t;x,\sqrt{1-\|x\|^2}) + p_Z(t;x,-\sqrt{1-\|x\|^2})\Big)\,(1-\|x\|^2)^{-1/2},$$

which is smooth on the interior of $\mathscr{B}^d$. □

5 Proofs for Section 3

In this section we collect the proofs of the main results in Section 3. In particular, the proof of Theorem 3.5 relies on results regarding the space $\mathscr{K}$ defined in (3.3); we develop these results in Section 5.2.

5.1 Proof of Theorem 3.3

Recall that we set $\binom{d}{k} = 0$ for $k > d$. Consider the linear spaces

$$\mathscr{D} = \Big\{c : \mathbb{R}^d \to \mathbb{S}^d : c_{ij} \in \mathrm{Hom}_2 \text{ for all } i,j\Big\}$$
$$\mathscr{E} = \Big\{f : \mathbb{R}^d \to \mathbb{R}^d : f_i \in \mathrm{Hom}_3 \text{ for all } i\Big\}$$

as well as the linear map $T : \mathscr{D} \to \mathscr{E}$ defined by $(Tc)(x) = c(x)x$. Then $\ker T = \mathscr{C}$, so we need to prove that $\ker T = \{c_H : H \in \mathbb{S}^d\}$. We make the following observations.

(i) $\dim \mathscr{D} = \binom{d+1}{2}^2$, since each $c \in \mathscr{D}$ is specified by $\binom{d+1}{2}$ independently chosen elements of Hom_2, each of which is specified by $\dim \mathrm{Hom}_2 = \binom{d+1}{2}$ independent parameters.

(ii) $\dim \mathscr{E} = d\binom{d+2}{3}$, since each $f \in \mathscr{E}$ is specified by d independently chosen elements of Hom_3, each of which is specified by $\dim \mathrm{Hom}_3 = \binom{d+2}{3}$ independent parameters.

(iii) $\dim \mathscr{D} - \dim \mathscr{E} = 2\binom{d}{4} + 3\binom{d}{3} + \binom{d}{2}$. This follows from a direct calculation.

(iv) T is surjective. To see this, fix any $i,j,k,l \in \{1,\dots,d\}$. We need to find $c \in \mathscr{D}$ such that $c(x)x = x_i x_j x_k e_l$. One such c is given by $c(x) = x_i(x_j E_{kl} + x_k E_{jl} - x_l E_{jk})$, where we define $E_{st} = e_s e_t^\top + e_t e_s^\top$.

Since the collection $\{D_p\, xx^\top D_q^\top + D_q\, xx^\top D_p^\top : 1 \le p < q \le m\}$ is a subset of $\ker T$, the rank-nullity theorem together with (iii) and (iv) implies (3.4) once we prove that this collection contains $2\binom{d}{4} + 3\binom{d}{3} + \binom{d}{2}$ linearly independent elements. To this end, recall the elementary skew-symmetric matrices

$$S_{ij} = e_i e_j^\top - e_j e_i^\top$$

appearing in (1.5), and consider the sub-collection

$$\begin{aligned}\mathscr{B} \;&= \{S_{ij}\,xx^\top S_{kl}^\top + S_{kl}\,xx^\top S_{ij}^\top : 1 \le i < j < k < l \le d\}\\ &\cup \{S_{ik}\,xx^\top S_{jl}^\top + S_{jl}\,xx^\top S_{ik}^\top : 1 \le i < j < k < l \le d\}\\ &\cup \{S_{ij}\,xx^\top S_{ik}^\top + S_{ik}\,xx^\top S_{ij}^\top : 1 \le i < j < k \le d\}\\ &\cup \{S_{ij}\,xx^\top S_{jk}^\top + S_{jk}\,xx^\top S_{ij}^\top : 1 \le i < j < k \le d\}\\ &\cup \{S_{ik}\,xx^\top S_{jk}^\top + S_{jk}\,xx^\top S_{ik}^\top : 1 \le i < j < k \le d\}\\ &\cup \{2S_{ij}\,xx^\top S_{ij}^\top : 1 \le i < j \le d\}.\end{aligned}$$

It contains $2\binom{d}{4} + 3\binom{d}{3} + \binom{d}{2}$ elements, which as we now show are linearly independent. For any $i < j$, the matrix $2S_{ij}\,xx^\top S_{ij}^\top$ is the only matrix in $\mathscr{B}$ with x_i^2 appearing in position (j,j). Moreover, for any $i < j < k$, $S_{ij}\,xx^\top S_{jk}^\top + S_{jk}\,xx^\top S_{ij}^\top$ is the only matrix in $\mathscr{B}$ with x_j^2 in position (i,k); $S_{ik}\,xx^\top S_{jk}^\top + S_{jk}\,xx^\top S_{ik}^\top$ is the only matrix with x_k^2 in position (i,j); and $S_{ij}\,xx^\top S_{ik}^\top + S_{ik}\,xx^\top S_{ij}^\top$ is the only matrix with x_i^2 in position (j,k). Finally, for any $i < j < k < l$, $S_{ij}\,xx^\top S_{kl}^\top + S_{kl}\,xx^\top S_{ij}^\top$ is the only matrix in $\mathscr{B}$ with x_ix_l in position (j,k); and $S_{ik}\,xx^\top S_{jl}^\top + S_{jl}\,xx^\top S_{ik}^\top$ is the only matrix with x_kx_l in position (i,j). These observations imply that $\mathscr{B}$ is a linearly independent set, as claimed. Hence (3.4) is proved.

It remains to show that $\dim \mathscr{K} = \binom{d}{4}$. To see this, observe that $\mathscr{K}$ is the kernel of the linear map

$$\mathbb{S}^m \to \mathscr{C}, \qquad H \mapsto c_H.$$

By (3.4), this map is surjective, and the dimension of its kernel is

$$\dim \mathbb{S}^m - \dim \mathscr{C} = \binom{m+1}{2} - 2\binom{d}{4} - 3\binom{d}{3} - \binom{d}{2} = \binom{d}{4},$$

where the last equality follows from a direct calculation.

5.2 The space $\mathscr{K}$ and the Plücker relations

The goal is now to develop a better understanding of the space $\mathscr{K}$ in (3.3). The main outcome is Lemma 5.3 below, which is crucial for the proof of Theorem 3.5. The lemma shows in particular that the set $\mathscr{K}$ can be identified with the Plücker relations from algebraic geometry.

Let $m = \binom{d}{2} = \dim \mathrm{Skew}(d)$. Each matrix $H = (h_{pq}) \in \mathbb{S}^m$ can be viewed as a symmetric linear map on $\mathrm{Skew}(d)$, acting on matrices $A = (a_{ij}) \in \mathrm{Skew}(d)$ by the formula

$$H[A]_{ij} = \sum_{k<l} h_{\pi(i,j),\pi(k,l)} a_{kl}, \qquad i < j,$$

with the remaining components of $H[A]$ given by skew-symmetry. Here the map π lists strictly upper-triangular matrix elements in lexicographic order,

(i,j)	$(1,2)$	$(1,3)$	$(1,4)$	$\cdots$	$(1,d)$	$(2,3)$	$\cdots$	$(2,d)$	$\cdots$	$(d-1,d)$
$\pi(i,j)$	1	2	3	$\cdots$	$d-1$	d	$\cdots$	$2d-3$	$\cdots$	$m = \binom{d}{2}$

The symmetry property $\langle A, H[B]\rangle = \langle H[A], B\rangle$ is immediate from the symmetry of H as a matrix. Using this notation, the biquadratic form corresponding to an element $c_H \in \mathscr{C}$, where $H \in \mathbb{S}^m$, can be expressed as

$$y^\top c_H(x) y = \frac{1}{2}\langle A, H[A]\rangle \qquad \text{where} \qquad A = xy^\top - yx^\top. \tag{5.1}$$

This follows from (3.2) using that $x^\top D_{\pi(i,j)} y = x_i y_j - x_j y_i$. Together with Lemma 5.1 below this immediately yields

$$\mathscr{K} = \{K \in \mathbb{S}^m \colon \langle A, K[A]\rangle = 0 \text{ for all } A \in \operatorname{Skew}(2,d)\}. \tag{5.2}$$

In particular, this shows that $\mathscr{C} \cong \mathbb{S}^m/\mathscr{K}$ can be identified with the set of restrictions to $\operatorname{Skew}(2,d)$ of quadratic forms on $\operatorname{Skew}(d)$.

Lemma 5.1. *Let $A \in \operatorname{Skew}(d)$. Then $\operatorname{rank} A = 2$ if and only if $A = xy^\top - yx^\top$ for some linearly independent $x, y \in \mathbb{R}^d$.*

Proof. Only the necessity needs proof, so pick $A \in \operatorname{Skew}(2,d)$. Any rank-two matrix A can be written $A = xx^\top A + yy^\top A$, where $\{x, y\}$ is an orthonormal basis for the range of A. By skew-symmetry of A, we have $A^\top x = \gamma y$ and $A^\top y = -\gamma x$, where $\gamma = x^\top A y$. Thus $A = (\gamma x)y^\top - y(\gamma x)^\top$, which is of the stated form. □

For each ordered 4-tuple $i < j < k < l$ of indices from $\{1, \dots, d\}$, define a polynomial P_{ijkl} in the entries of $A \in \operatorname{Skew}(d)$ by

$$P_{ijkl}(A) = a_{ij}a_{kl} - a_{ik}a_{jl} + a_{il}a_{jk}.$$

These polynomials generate the variety of rank-two skew-symmetric matrices: all P_{ijkl} vanish at a nonzero $A \in \operatorname{Skew}(d)$ if and only if $\operatorname{rank} A = 2$. Moreover, they constitute a basis for the homogeneous ideal generated by this variety: any homogeneous polynomial vanishing on $\operatorname{Skew}(2,d)$ is a polynomial linear combination of the P_{ijkl}. The classical use of this fact is to give the Grassmannian manifold of two-dimensional subspaces of $\mathbb{R}^d$ the structure of a projective variety. This is achieved via the *Plücker embedding*, which maps a subspace $V = \operatorname{span}\{x, y\}$ of $\mathbb{R}^d$ to $A = xy^\top - yx^\top \in \operatorname{Skew}(d)$. Note that, up to scaling, A only depends on V, not on the choice of spanning vectors x and y. Conversely, any $A \in \operatorname{Skew}(2,d)$ corresponds to a unique two-dimensional subspace $V \subset \mathbb{R}^d$ due to Lemma 5.1. The entries a_{ij} of A are called the *Plücker (or Grassmann) coordinates* of V, and the quadratic equations $P_{ijkl}(A) = 0$ that they satisfy are called the *Plücker relations*. For us, the relevant result is the following:

Lemma 5.2. *The polynomials P_{ijkl} have the following properties:*

(i) $\operatorname{Skew}(2,d) = \{A \in \operatorname{Skew}(d) \setminus \{0\} \colon P_{ijkl}(A) = 0 \text{ for all } i < j < k < l\}$.

(ii) *Let $P(A)$ be a homogeneous polynomial in the entries of $A \in \operatorname{Skew}(d)$. Then $P(A) = 0$ for all $A \in \operatorname{Skew}(2,d)$ if and only if it is of the form*

$$P(A) = \sum_{i<j<k<l} Q_{ijkl}(A) P_{ijkl}(A)$$

for some polynomials Q_{ijkl}.

Proof. Keeping in mind the identification of $\mathrm{Skew}(2,d)$ with the set of two-dimensional subspaces as above, part (i) follows from Theorem II on page 312 in Hodge et al. (1994), and part (ii) follows from Theorem I on page 315 in the same reference. □

The following lemma is the primary reason for studying $\mathscr{K}$ in such detail; it is a crucial ingredient in the proof of Theorem 3.5. In particular, it shows that the set $\mathscr{K}$ can be identified with the Plücker relations.

Lemma 5.3. *The set $\mathscr{K}$ can be identified with the linear span of the polynomials P_{ijkl}, $i<j<k<l$. Moreover, one has*

$$\mathrm{Skew}(2,d) = \{A \in \mathrm{Skew}(d) \setminus \{0\} \colon \langle A, K[A]\rangle = 0 \text{ for all } K \in \mathscr{K}\}\,. \tag{5.3}$$

Proof. Lemma 5.2(ii) together with (5.2) implies that $K \in \mathscr{K}$ holds if and only if the polynomial $P(A) = \langle A, K[A]\rangle$ is a linear combination of the $P_{ijkl}(A)$. The latter polynomials can therefore be identified with a basis for $\mathscr{K}$. It follows that the right-hand side of (5.3) is equal to $\{A \in \mathrm{Skew}(d) \setminus \{0\} \colon P_{ijkl}(A) = 0 \text{ for all } i<j<k<l\}$. This, in turn, is equal to $\mathrm{Skew}(2,d)$ by Lemma 5.2(i). □

Remark 5.4. *The identification of $\mathscr{K}$ with the linear span of the polynomials P_{ijkl} leads to an explicit description of a basis for $\mathscr{K}$. Indeed, one has*

$$P_{ijkl}(A) = \frac{1}{4}\langle A, K_{(i,j,k,l)}[A]\rangle,$$

where $K_{(i,j,k,l)}$ is the linear map on $\mathrm{Skew}(d)$ *characterized by*

$$\begin{aligned}
K_{(i,j,k,l)}[A]_{ij} &= a_{kl}\\
K_{(i,j,k,l)}[A]_{ik} &= -a_{jl}\\
K_{(i,j,k,l)}[A]_{il} &= a_{jk}\\
K_{(i,j,k,l)}[A]_{jk} &= a_{il}\\
K_{(i,j,k,l)}[A]_{jl} &= -a_{ik}\\
K_{(i,j,k,l)}[A]_{kl} &= a_{ij}\\
K_{(i,j,k,l)}[A]_{rs} &= 0 \text{ if } \{r,s\} \not\subseteq \{i,j,k,l\}.
\end{aligned}$$

The collection of all such $K_{(i,j,k,l)}$ for $i<j<k<l$ thus constitutes a basis for $\mathscr{K}$. Note that there are $\binom{d}{4}$ ways of choosing the indices i,j,k,l, which is in agreement with the dimension of $\mathscr{K}$ as given in Theorem 3.3.

5.3 Proof of Theorem 3.5(i)

A preliminary result is required for the proof of Theorem 3.5(i). Recall that an *affine subspace* of a vector space is a subset $\mathscr{A}$ such that $x,y \in \mathscr{A}$, $\lambda \in \mathbb{R}$ implies $\lambda x + (1-\lambda)y \in \mathscr{A}$. If $\mathscr{A}$ is an affine subspace and $x \in \mathscr{A}$, then $\mathscr{A} - x$ is a linear subspace that does not depend on x. The *dimension* of $\mathscr{A}$ is the dimension of $\mathscr{A} - x$. The following result is Theorem 1.1 in Barvinok (2001); see Theorem 1.3 and its reformulation (2.2) in Barvinok (1995) for a proof.

Lemma 5.5. *Let $\mathscr{A} \subset \mathbb{S}^m$ be an affine subspace such that the intersection $\mathscr{A} \cap \mathbb{S}^m_+$ is nonempty and* $\operatorname{codim} \mathscr{A} \leq \binom{r+2}{2} - 1$ *for some $r \in \mathbb{N}$. Then there exists some $B \in \mathscr{A} \cap \mathbb{S}^m_+$ with* $\operatorname{rank} B \leq r$.

We now proceed with the proof of Theorem 3.5(i). We need to show that, for any $H \in \mathbb{S}^m$,

$$y^\top c_H(x) y \geq 0 \text{ for all } x, y \in \mathbb{R}^d \qquad \text{implies} \qquad (H + \mathscr{K}) \cap \mathbb{S}^m_+ \neq \emptyset. \tag{5.4}$$

Indeed, then any $c_H \in \mathscr{C}_+$ is equal to c_{H+K} for some $K \in \mathscr{K}$ such that $H + K \in \mathbb{S}^m_+$.

We prove the contrapositive. Thus, consider any $H \in \mathbb{S}^m$ such that $(H + \mathscr{K}) \cap \mathbb{S}^m_+ = \emptyset$. We need to prove that $y^\top c_H(x) y < 0$ for some $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$. The separating hyperplane theorem gives some $B \in \mathbb{S}^m$ such that $\langle C, B \rangle \geq 0$ for all $C \in \mathbb{S}^m_+$ and $\langle H + K, B \rangle < 0$ for all $K \in \mathscr{K}$. The former inequality yields $B \in \mathbb{S}^m_+$ by self-duality of $\mathbb{S}^m_+$. The latter inequality yields $\langle K, B \rangle = 0$ for all $K \in \mathscr{K}$, since otherwise $\langle H + tK, B \rangle$ would become positive for some $K \in \mathscr{K}$ and some sufficiently large t. Thus, after scaling B if necessary, we have $B \in \mathscr{K}^\perp \cap \mathbb{S}^m_+$ and $\langle H, B \rangle = -1$. Hence $\mathscr{A} \cap \mathbb{S}^m_+ \neq \emptyset$ for the affine subspace

$$\mathscr{A} = \left\{ C \in \mathbb{S}^m \colon C \in \mathscr{K}^\perp \text{ and } \langle C, H \rangle = -1 \right\}.$$

Note that $\mathscr{A} - B = \{C \in \mathbb{S}^m \colon C \in \mathscr{K}^\perp \text{ and } \langle C, H \rangle = 0\}$. Our initial assumption excludes $H \in \mathscr{K}$, since otherwise we would have $0 \in (H + \mathscr{K}) \cap \mathbb{S}^m_+$. Thus $\operatorname{codim} \mathscr{A} = 1 + \dim \mathscr{K}$, so that $\operatorname{codim} \mathscr{A} = 2$ for $d = 4$, and $\operatorname{codim} \mathscr{A} = 1$ for $d \in \{2, 3\}$. In either case, $\operatorname{codim} \mathscr{A} \leq \binom{3}{2} - 1$. Applying Lemma 5.5 with $r = 1$ then yields a rank-one matrix $zz^\top$, $z \in \mathbb{R}^m \setminus \{0\}$, in the intersection $\mathscr{A} \cap \mathbb{S}^m_+$. In particular, $z^\top K z = 0$ for all $K \in \mathscr{K}$. Identifying z with the skew-symmetric matrix A whose entries are given by $a_{ij} = z_{\pi(i,j)}$ for $i < j$, this says that $\langle A, K[A] \rangle = 0$ for all $K \in \mathscr{K}$. Thus $\operatorname{rank} A = 2$ by Lemma 5.3, whence $A = xy^\top - yx^\top$ for some $x, y \in \mathbb{R}^d$ by Lemma 5.1. Consequently, (5.1) yields $y^\top c_H(x) y = \langle A, H[A] \rangle / 2 = z^\top H z = \langle zz^\top, H \rangle = -1$, as required. Theorem 3.5(i) is proved.

5.4 Proof of Theorem 3.5(ii)

We will use the construction given in László (2010) to obtain an example for $d = 6$. In order to make the presentation self-contained, we summarize the construction here. Note that it suffices to consider the case $d = 6$. Indeed, suppose $f(x_1, \ldots, x_6, y_1, \ldots, y_6)$ is a nonnegative biquadratic form in $6 + 6$ variables vanishing on the diagonal that is not a sum of squares. For $d > 6$, define

$$g(x, y) = f(x_1, \ldots, x_6, y_1, \ldots, y_6).$$

The left-hand side is then a nonnegative biquadratic form in $d + d$ variables vanishing on the diagonal that is not a sum of squares.

From now on we thus take $d = 6$ and hence $m = 15$. In view of Lemma 3.4 we must find $H \in \mathbb{S}^{15}$ such that $y^\top c_H(x) y \geq 0$ for all x, y, but at the same time

$$(H + \mathscr{K}) \cap \mathbb{S}^{15}_+ = \emptyset; \tag{5.5}$$

in other words, we must find a counterexample to (5.4). A candidate H is obtained from the inequality

$$2(\|X\|^2 \|Y\|^2 - \operatorname{Tr}(X^\top Y)^2) - \|XY - YX\|^2 \geq 0, \tag{5.6}$$

which is valid for any $d \times d$ matrices X and Y; see for instance Böttcher and Wenzel (2008, Theorem 2.2). Here $\|X\| = \sqrt{\mathrm{Tr}(X^\top X)}$ is the Frobenius norm. Note also that the left-hand side defines a biquadratic form vanishing on the diagonal. Consider now the special case where for $x, y \in \mathbb{R}^6$,

$$X = \begin{pmatrix} x_1 & x_2 & x_3 \\ 0 & x_6 & x_4 \\ 0 & 0 & x_5 \end{pmatrix}, \qquad Y = \begin{pmatrix} y_1 & y_2 & y_3 \\ 0 & y_6 & y_4 \\ 0 & 0 & y_5 \end{pmatrix}.$$

As in Section 5.2 we use the notation

$$A = xy^\top - yx^\top.$$

One then easily verifies the identities $2(\|X\|^2\|Y\|^2 - \mathrm{Tr}^2(X^\top Y)) = \|A\|^2$ and $\|XY - YX\|^2 = (a_{12} + a_{26})^2 + (a_{45} - a_{46})^2 + (a_{13} + a_{24} + a_{35})^2$. Furthermore, Lemma 5.2(i) yields $P_{1234}(A) = P_{2345}(A) = 0$. In view of (5.6) it follows that the function

$$f(x, y) = \|A\|^2 - (a_{12} + a_{26})^2 - (a_{45} - a_{46})^2 - (a_{13} + a_{24} + a_{35})^2 - P_{1234}(A) - P_{2345}(A)$$

is a nonnegative biquadratic form vanishing on the diagonal. The correspondence established in (5.1) then yields

$$f(x, y) = y^\top c_H(x) y,$$

where the matrix $H \in \mathbb{S}^{15}$ can be found to be

$$H = \frac{1}{2} \begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \end{pmatrix}.$$

The characteristic polynomial of H is given by

$$\det(s\,\mathrm{Id} - H) = 2^{-5}(s-2)^7\,(2s^2 - 2s - 1)\,(4s^3 - 16s^2 + 14s + 1)^2.$$

Here $2s^2 - 2s - 1$ has one single negative root $\lambda = (1 - \sqrt{3})/2$. Moreover, the polynomial $\phi(s) = 4s^3 - 16s^2 + 14s + 1$ has one single negative root μ, since $\phi(0) = 1 > 0$, $\phi(2) = -3 < 0$,

$\lim_{s\to\infty}\phi(s)=\infty$, and $\lim_{s\to-\infty}\phi(s)=-\infty$. We now define the three vectors

$$\begin{aligned}
v_1 &= \frac{1}{2}e_2 - \lambda\, e_7 + \frac{1}{2}e_{11}\\
v_2 &= \frac{\mu}{2}e_3 + \mu(2-\mu)\, e_6 + \frac{1}{2}(\mu-1)\, e_{13} + \frac{1}{2}e_{14}\\
v_3 &= \frac{1}{2}(1-\mu)\, e_1 - \frac{\mu}{2}\, e_8 + \frac{1}{2}e_9 + \mu(\mu-2)\, e_{10},
\end{aligned}$$

where e_i denotes the ith canonical basis vector in $\mathbb{R}^{15}$. A calculation shows that v_1 is an eigenvector of H corresponding to the eigenvalue λ, and v_2, v_3 are linearly independent eigenvectors of H corresponding to the eigenvalue μ. We now define the positive semidefinite matrix

$$B = \delta v_1 v_1^\top + v_2 v_2^\top + v_3 v_3^\top,$$

where $\delta = \mu(\mu-2)(2\mu-1)/\lambda > 0$. This matrix satisfies

$$\langle H, B\rangle = \operatorname{Tr}(HB) = \lambda\delta\|v_1\|^2 + \mu\|v_2\|^2 + \mu\|v_3\|^2 < 0. \tag{5.7}$$

In addition, by considering each of the $\binom{6}{4} = 15$ basis elements $K_{(i,j,k,l)}$ of $\mathscr{K}$ described in Remark 5.4, one can check by direct computation that

$$\langle K, B\rangle = 0 \quad \text{for all} \quad K \in \mathscr{K}. \tag{5.8}$$

Since B is nonzero and positive semidefinite, (5.7) and (5.8) imply (5.5). This completes the proof of Theorem 3.5(ii).

Remark 5.6. *The condition* (5.5) *and its relation to sum of squares representations can also be expressed via the following primal-dual pair of semidefinite programs:*

$$v(P) = \min_{B\in\mathbb{S}^m} \{\langle H, B\rangle \colon B \succeq 0,\ \langle \mathrm{Id}, B\rangle = 1,\ \langle K, B\rangle = 0 \text{ for all } K \in \mathscr{K}\} \tag{5.9}$$

$$v(D) = \max_{\substack{\lambda\in\mathbb{R}\\ K\in\mathscr{K}}} \{\lambda \colon H - \lambda\mathrm{Id} + K \succeq 0\}. \tag{5.10}$$

Note that $v(D) \geq 0$ if and only if $(H+\mathscr{K}) \cap \mathbb{S}^{15}_+ \neq \emptyset$. Moreover, weak duality $v(D) \leq v(P)$ always holds. Indeed, for any primal feasible solution B and dual feasible solution (λ, K) one has $0 \leq \langle B, H - \lambda\mathrm{Id} + K\rangle = \langle H, B\rangle - \lambda$. Thus, for a given candidate $c = c_H$, a primal feasible solution B with $\langle H, B\rangle < 0$ provides a certificate that no sum of squares representation can exist. Furthermore, (5.7) *and* (5.8) *imply that the matrix $B/\operatorname{Tr}(B)$ is thus a feasible solution to the primal problem* (5.9) *with negative value, and hence certifies that no sum of squares representation can exist.*

Remark 5.7. *One can write down the component functions of the resulting map $c = c_H$:*

$$
\begin{aligned}
c_{11}(x) &= x_2^2 + x_3^2 + 2x_4^2 + 2x_5^2 + 2x_6^2 \\
c_{12}(x) &= -x_1x_2 - x_2x_6 - x_3x_4 \\
c_{13}(x) &= -x_1x_3 - x_3x_5 \\
c_{14}(x) &= x_2x_3 - 2x_1x_4 \\
c_{15}(x) &= -2x_1x_5 + x_3^2 \\
c_{16}(x) &= -2x_1x_6 + x_2^2 \\
c_{22}(x) &= x_1^2 + 2x_1x_6 + 2x_3^2 + x_4^2 + 2x_5^2 + x_6^2 \\
c_{23}(x) &= x_1x_4 - 2x_2x_3 - x_4x_5 \\
c_{24}(x) &= -x_2x_4 \\
c_{25}(x) &= x_3x_4 - 2x_2x_5 \\
c_{26}(x) &= -x_1x_2 - x_2x_6 \\
c_{33}(x) &= x_1^2 + 2x_1x_5 + 2x_2^2 + 2x_4^2 + x_5^2 + 2x_6^2 \\
c_{34}(x) &= -x_1x_2 + x_2x_5 - 2x_3x_4 \\
c_{35}(x) &= -x_1x_3 - x_3x_5 \\
c_{36}(x) &= -2x_3x_6 \\
c_{44}(x) &= 2x_1^2 + x_2^2 + 2x_3^2 + x_5^2 + 2x_5x_6 + x_6^2 \\
c_{45}(x) &= -x_2x_3 - x_4x_5 - x_4x_6 \\
c_{46}(x) &= -x_4x_5 - x_4x_6 \\
c_{55}(x) &= 2x_1^2 + 2x_2^2 + x_3^2 + x_4^2 + 2x_6^2 \\
c_{56}(x) &= x_4^2 - 2x_5x_6 \\
c_{66}(x) &= 2x_1^2 + x_2^2 + 2x_3^2 + x_4^2 + 2x_5^2
\end{aligned}
$$

References

D. Ackerer, D. Filipović, and S. Pulido. The Jacobi stochastic volatility model. *Working paper*, 2015.

D. Bakry, S. Orevkov, and M. Zani. Orthogonal polynomials and diffusion operators. *arXiv:1309.5632v2*, 2014.

Dominique Bakry. Symmetric diffusions with polynomial eigenvectors. In *Stochastic Analysis and Applications 2014*, volume 100 of *Springer Proceedings in Mathematics & Statistics*, pages 25–49. Springer International Publishing, 2014.

Alexander Barvinok. A remark on the rank of positive semidefinite matrices subject to affine constraints. *Discrete & Computational Geometry*, 25(1):23–31, 2001.

Alexander I. Barvinok. Problems of distance geometry and convex properties of quadratic maps. *Discrete & Computational Geometry*, 13(1):189–202, 1995.

Albrecht Böttcher and David Wenzel. The Frobenius norm and the commutator. *Linear Algebra and Its Applications*, 8(429):1864–1885, 2008.

Man-Duen Choi. Positive semidefinite biquadratic forms. *Linear Algebra and its Applications*, 12(2):95–100, 1975.

C. Cuchiero. *Affine and Polynomial Processes*. PhD thesis, ETH Zurich, 2011.

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

G. Da Prato and H. Frankowska. Stochastic viability of convex sets. *J. Math. Anal. Appl.*, 333:151–163, 2007.

Dante DeBlassie. Uniqueness for diffusions degenerating at the boundary of a smooth bounded set. *The Annals of Probability*, 32(4):3167–3190, 2004.

F. Delbaen and H. Shirakawa. An interest rate model with upper and lower bounds. *Asia-Pacific Financial Markets*, 9:191–209, 2002.

S. N. Ethier and T. G. Kurtz. *Markov processes: characterization and convergence*. Wiley Interscience, Hoboken, 2005.

D. Filipović, E. Gourier, and L. Mancini. Quadratic variance swap models. *Journal of Financial Economics*, Forthcoming, 2015a.

D. Filipović, M. Larsson, and A. Trolle. Linear-rational term structure models. Technical report, Swiss Finance Institute Research Paper No. 14-15, 2015b. Available at SSRN: http://ssrn.com/abstract=2397898.

Damir Filipović and Martin Larsson. Polynomial preserving diffusions and applications in finance. *arXiv:1404.0989*, 2015.

J. L. Forman and M. Sørensen. The Pearson diffusions: a class of statistically tractable diffusion processes. *Scand. J. Statist.*, 35(3):438–465, 2008.

C. Gourieroux and J. Jasiak. Multivariate Jacobi process with application to smooth transitions. *Journal of Econometrics*, 131:475–505, 2006.

Martin Hairer. On Malliavins proof of Hörmanders theorem. *Bulletin des sciences mathematiques*, 6(135):650–666, 2011.

William Vallance Douglas Hodge, WVD Hodge, and Daniel Pedoe. *Methods of algebraic geometry*, volume 2. Cambridge University Press, 1994.

Lars Hörmander. Hypoelliptic second order differential equations. *Acta Mathematica*, 119 (1):147–171, 1967.

Elton P Hsu. *Stochastic analysis on manifolds*, volume 38. American Mathematical Soc., 2002.

P. Krühner and M. Larsson. Classification of affine processes with compact state space. *Working paper*, 2015.

K.S. Larsen and M. Sørensen. Diffusion models for exchange rates in a target zone. *Mathematical Finance*, 17(2):285–306, 2007.

M. Larsson and J. Ruf. Convergence of local supermartingales and Novikov-Kazamaki type conditions for processes with jumps. arXiv:1411.6229, 2014.

L László. On sum of squares decomposition for a biquadratic matrix function. *Ann. Univ. Sci. Budapest. Sect. Comput*, 33:273–284, 2010.

Lajos László. Sum of squares representation for the Böttcher-Wenzel biquadratic form. *Acta Univ. Sapientiae*, 4(1):17–32, 2012.

O. Mazet. Classification des semi-groupes de diffusion sur $\mathbb{R}$ associé à une famille de polynômes orthogonaux. *Séminaire de probabilités (Strasbourg)*, 31:40–53, 1997.

Gareth C Price and David Williams. Rolling with "slipping": I. *Séminaire de probabilités de Strasbourg*, 17:194–197, 1983.

Ronan Quarez. On the real zeros of positive semidefinite biquadratic forms. *Communications in Algebra*, 43(3):1317–1353, 2015.

P. Spreij and E. Veerman. Affine diffusions with non-canonical state space. *Stochastic Analysis and Applications*, 30:605–641, 2012.

JM Swart. Pathwise uniqueness for a sde with non-lipschitz coefficients. *Stochastic processes and their applications*, 98(1):131–149, 2002.

M Van den Berg and JT Lewis. Brownian motion on a hypersurface. *Bulletin of the London Mathematical Society*, 17(2):144–150, 1985.

E. Wong. The construction of a class of stationary Markoff processes. In R. Bellman, editor, *Stochastic Processes in Mathematical Physics and Engineering*, pages 264–276. Providence, R.I.: American Mathematical Society, 1964.

H. Zhou. Itô conditional moment generator and the estimation of short-rate processes. *Journal of Financial Econometrics*, 1:250–271, 2003.