



**HAL**  
open science

# Stochastic invariance of closed sets with non-Lipschitz coefficients

Eduardo Abi Jaber, Bruno Bouchard, Camille Illand

► **To cite this version:**

Eduardo Abi Jaber, Bruno Bouchard, Camille Illand. Stochastic invariance of closed sets with non-Lipschitz coefficients. 2016. hal-01349639v1

**HAL Id: hal-01349639**

**<https://hal.science/hal-01349639v1>**

Preprint submitted on 28 Jul 2016 (v1), last revised 19 Jun 2018 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Stochastic invariance of closed sets with non-Lipschitz coefficients

Eduardo Abi Jaber <sup>\*1,2</sup>, Bruno Bouchard<sup>†1</sup> and Camille Illand <sup>‡2</sup>

<sup>1</sup>Université Paris-Dauphine, PSL Research University, CNRS, UMR [7534],  
CEREMADE, 75016 Paris, France.

<sup>2</sup>AXA Investment Managers, Multi Asset Client Solutions, Quantitative Research,  
6 place de la Pyramide, 92908 Paris - La Défense, France.

July 28, 2016

## Abstract

This paper provides a new characterization of the stochastic invariance of a closed subset of  $\mathbb{R}^d$  with respect to a diffusion. We extend the well-known inward pointing Stratonovich drift condition to the case where the diffusion matrix can fail to be differentiable: we only assume that the covariance matrix is. In particular, our result can be directly applied to construct affine diffusions and polynomial preserving diffusions on any arbitrary closed set.

*Keywords:* Stochastic differential equation, stochastic invariance, affine diffusions, polynomial preserving diffusions.

*MSC Classification:* 93E03, 60H10.

## 1 Introduction

Let  $b : \mathbb{R}^d \mapsto \mathbb{R}^d$  and  $\sigma : \mathbb{R}^d \mapsto \mathbb{M}^d$  be continuous functions, where  $\mathbb{M}^d$  denotes the space of  $d \times d$  matrices. We assume that  $b$  and  $\sigma$  satisfy the following linear growth conditions: there exists  $L > 0$  such that

$$\|b(x)\| + \|\sigma\sigma^\top(x)\|^{\frac{1}{2}} \leq L(1 + \|x\|), \quad \forall x \in \mathbb{R}^d, \quad (H_1)$$

and we consider a weak solution of the stochastic differential equation

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x, \quad (1.1)$$

i.e. a  $d$ -dimensional Brownian motion  $W$  and an adapted process  $X$  such that the above equation holds.

The aim of this paper is to provide a characterization of the *stochastic invariance* of a closed set  $\mathcal{D} \subset \mathbb{R}^d$ , i.e. find necessary and sufficient conditions on the instantaneous drift  $b$  and the instantaneous covariance matrix  $\sigma\sigma^\top$  under which there exists a weak solution of (1.1) that remains in  $\mathcal{D}$  for all  $t \geq 0$ , almost surely, given that  $x \in \mathcal{D}$ . (See Definition 2.2 below for a

---

\*abijaber@ceremade.dauphine.fr, eduardo.abijaber@axa-im.com

†bouchard@ceremade.dauphine.fr. I would like to thank Christa Cuchiero and Joseph Teichmann for very fruitful discussions.

‡camille.illand@axa-im.com

precise formulation.)

The first stochastic invariance results can be found in Friedman [21] and Doss [16]. Since then, many extensions were considered in the literature. For an arbitrary closed set, the stochastic invariance was characterized through the second order normal cone in Bardi and Goatin [3] and Bardi and Jensen [4]. Aubin and Doss [1] used the notion of curvature, while Da Prato and Frankowska [13] provided a characterization in terms of the Stratonovich drift. For a closed convex set, the distance function was used in Da Prato and Frankowska [14], and the invariance was characterized for affine jump-diffusions in Tappe [37].

Although these approaches differ, they have at least one thing in common: the tradeoff one has to make between the assumptions on the topology/smoothness of the domain and the regularity of the coefficients  $b$  and  $\sigma$ . This makes all of these existing results difficult to apply in practice. Let us start by highlighting this difficulty through the two main contributions to the literature:

- (i) In Bardi and Jensen [4], the stochastic invariance is characterized by using Nagumo-type geometric conditions on the second order normal cone, for Lipschitz continuous coefficients  $b$  and  $\sigma$ . Their main result states that the closed set  $\mathcal{D}$  is stochastically invariant if and only if

$$u^\top b(x) + \frac{1}{2} \text{Tr}(vC(x)) \leq 0, \quad \forall x \in \mathcal{D} \text{ and } (u, v) \in \mathcal{N}_{\mathcal{D}}^2(x),$$

in which  $C := \sigma\sigma^\top$  on  $\mathcal{D}$  and  $\mathcal{N}_{\mathcal{D}}^2(x)$  is the second order normal cone at the point  $x$ :

$$\mathcal{N}_{\mathcal{D}}^2(x) := \left\{ (u, v) \in \mathbb{R}^d \times \mathbb{S}^d : \langle u, y - x \rangle + \frac{1}{2} \langle y - x, v(y - x) \rangle \leq o(\|y - x\|^2), \forall y \in \mathcal{D} \right\}. \quad (1.2)$$

Here,  $\mathbb{S}^d$  stands for the cone of symmetric  $d \times d$  matrices. In practice, we face two restrictions. First, prior to deriving the conditions on  $b$  and  $\sigma$ , we have to determine the second order normal cone at all points of a given set. When the boundary is smooth, the computation of the second order normal cone is an easy task, see e.g. [4, Example 1]. However, it is much more challenging in general, by lack of efficient techniques. This renders the result of [4] difficult to use in practice. Second, the Lipschitz continuity condition imposed on the coefficients is an important restriction.

- (ii) Building on Doss [16], Da Prato and Frankowska [13] give necessary and sufficient conditions for the stochastic invariance in terms of the Stratonovich drift and the first order normal cone:

$$\sigma(x)^\top u = 0 \text{ and } \langle u, b(x) - \frac{1}{2} \sum_{j=1}^d D\sigma^j(x)\sigma^j(x) \rangle \leq 0, \quad \forall x \in \mathcal{D} \text{ and } u \in \mathcal{N}_{\mathcal{D}}^1(x), \quad (1.3)$$

where  $\sigma^j(x)$  denotes the  $j$ -th column of the matrix  $\sigma(x)$ ,  $D\sigma^j$  is the Jacobian of  $\sigma^j$ , and the first order normal cone  $\mathcal{N}_{\mathcal{D}}^1(x)$  at  $x$  (sometimes simply called *normal cone*) is defined as

$$\mathcal{N}_{\mathcal{D}}^1(x) := \left\{ u \in \mathbb{R}^d : \langle u, y - x \rangle \leq o(\|y - x\|), \forall y \in \mathcal{D} \right\}. \quad (1.4)$$

In practice, the first order normal cone is much simpler to compute than the second order cone used in [4], see [2] and [34]. However, the price to pay is to impose a strong regularity condition on the diffusion matrix  $\sigma$ , which is assumed to be bounded and differentiable on  $\mathbb{R}^d$ , with a bounded Lipschitz derivative. Again, this constitutes a sticking point for applications, it cannot be applied to simple models (think about square-root processes for instance, see below).

The aim of the present paper is to extend the characterization (1.3), given in terms of the *easy-to-compute* first order normal cone, under weaker regularity conditions on the diffusion matrix  $\sigma$ . We make the following trivial observation:  $C$  might be differentiable at a point  $x$  while  $\sigma$  is not. It is the case for the square-root process mentioned above, at the boundary point  $x = 0$ . Moreover, the terms  $D\sigma^j(x)\sigma^j(x)$  can be rewritten in terms of the Jacobian of  $C$  whenever both quantities are well defined, see Proposition 2.4 for a precise formulation. This suggests to reformulate (1.3) with the Jacobian matrices of the columns of  $C$  instead of  $\sigma$ .

Our main result, Theorem 2.3 below, states that the stochastic invariance is equivalent to the following conditions:

$$C(x)u = 0 \text{ and } \langle u, b(x) - \frac{1}{2} \sum_{j=1}^d DC^j(x)(CC^+)^j(x) \rangle \leq 0, \forall x \in \mathcal{D} \text{ and } u \in \mathcal{N}_{\mathcal{D}}^1(x). \quad (1.5)$$

Here,  $(CC^+)^j(x)$  is the  $j$ -th column of  $(CC^+)(x)$  with  $C(x)^+$  defined as the Moore-Penrose pseudoinverse of  $C(x)$ , see Definition A.1 in the Appendix. We only assume that

$$C \text{ can be extended to a } \mathcal{C}_{loc}^{1,1}(\mathbb{R}^d, \mathbb{S}^d) \text{ function that coincides with } \sigma\sigma^\top \text{ on } \mathcal{D}, \quad (H_2)$$

in which  $\mathcal{C}_{loc}^{1,1}$  means  $\mathcal{C}^1$  with a locally Lipschitz derivative. Note that we do not impose the extension of  $C$  to be positive semi-definite outside  $\mathcal{D}$ , so that  $\sigma$  might only match with its square-root on  $\mathcal{D}$ . Also, it should be clear that the extension needs only to be local around  $\mathcal{D}$ .

The term  $CC^+$  in (1.5) plays the role of the projection on the image of  $C$ , see Proposition A.3 in the Appendix and the discussion in Remark 2.5 below. This projection term cannot be removed. To see this, let us consider the square-root process with  $C(x) = x$  and  $\mathcal{D} = \mathbb{R}_+$ , so that  $\mathcal{N}_{\mathcal{D}}^1(0) = \mathbb{R}_-$ . Then,

$$C(0)(-1) = 0 \quad \text{and} \quad \langle -1, b(0) - \frac{1}{2}DC(0) \rangle \leq 0$$

leads to  $b(0) \geq 1/2$  while the correct condition for invariance is  $b(0) \geq 0$ , which is recovered from (1.5) by using the fact that  $(CC^+)(0) = 0$ .

Therefore, our contribution to the theory of stochastic invariance is to extend the characterization of Da Prato and Frankowska [13] to weaker conditions by allowing the volatility matrix of the diffusion to degenerate and not be  $\mathcal{C}^1$  on the whole domain.

This is not simply a theoretical contribution. In fact, many models used in practice, in mathematical finance for instance, do not have  $\mathcal{C}^1$  volatility maps but satisfy our conditions. This is in particular the case of affine diffusions (see Filipović and Mayerhofer [20]), or of polynomial preserving diffusions that are characterized by a quadratic covariance matrix (see Filipović and Larsson [19]), etc. When applied to such processes, stochastic invariance results have been so far tweaked in order to fit in the previous set up, or have been proved under limiting conditions. For instance, in their construction of affine processes on the cone of symmetric semi-definite matrices, Cuchiero *et al.* [12] start by regularizing the martingale problem before applying the stochastic invariance characterization of [13] and then pass to the limit. In Spreij and Veerman [35], some stochastic invariance results are also derived for affine diffusions but only on convex sets with smooth boundary. More recently, the mathematical foundation for polynomial preserving diffusions is given in Filipović and Larsson [19]. Necessary conditions for the stochastic invariance are derived for basic closed semialgebraic sets. However, these conditions are not sharp, their sufficient conditions differ from their necessary conditions. All the above cases can now be treated by using our characterization. See Section 5 for a simple example.

Our proof of the necessary condition is in the spirit of Buckdahn *et al.* [8]. They use a second order stochastic Taylor expansion together with small time behavior results for double stochastic integrals. However, in our case, the stochastic Taylor expansion cannot be applied directly since  $\sigma$  is not differentiable and  $\sigma(X)$  fails to be a semi-martingale whenever an eigenvalue vanishes (see [31, Example 1.2]). If  $\sigma$  is diagonal, vanishing eigenvalues can be eliminated by taking the conditional expectation with respect to the path of the Brownian motion acting on the non-vanishing ones. This corresponds to the projection term  $CC^+$  in (1.5). Then, the proof is essentially the same as in [8] except that our double stochastic integrals contain conditional expectations. If  $\sigma$  is not diagonal, we can reduce to the former case by considering its spectral decomposition and a suitable change of Brownian motion (based on the corresponding basis change), see Lemma 3.2 below. However, it requires a smooth spectral decomposition which is not guaranteed when repeated eigenvalues are present. To avoid this, we need an additional transformation of the state space, see Proposition 3.5.

Conversely, we show that the infinitesimal generator of our diffusion satisfies the *positive maximum principle* whenever (1.5) holds, see Section 4 below. Applying [18, Theorem 4.5.4] shows that this condition is indeed sufficient. (Note that the approach based on the comparison principle for viscosity solutions used in [4, 8] cannot be applied to our case since  $\sigma$  is not Lipschitz.)

The rest of the paper is organized as follows. Our main result is stated in Section 2. The proofs are collected in Sections 3 and 4. In Section 5, we exemplify our characterization by deriving the stochastic invariance conditions for affine diffusions living in a parabolic state space. Finally, Section 6 provides a complementary tractable sufficient condition ensuring the stochastic invariance of the interior of a domain. For the convenience of the reader, we collect some standard results of matrix calculus and differentiation in the Appendix.

From now on, all identities involving random variables have to be considered in the a.s. sense, the probability space and the probability measure being given by the context. Elements of  $\mathbb{R}^d$  are viewed as column vectors. The vector  $e_i \in \mathbb{R}^d$  is the  $i$ -th element of the canonical basis, and we use the standard notation  $I_d$  to denote the  $d \times d$  identity matrix. We denote by  $\mathbb{M}^d$  the collection of  $d \times d$  matrices. We say that  $A \in \mathbb{S}^d$  (resp.  $\mathbb{S}_+^d$ ) if it is a symmetric (resp. and positive semi-definite) element of  $\mathbb{M}^d$ . Given  $x = (x^1, \dots, x^d) \in \mathbb{R}^d$ ,  $\text{diag}[x]$  denotes the diagonal matrix whose  $i$ -th diagonal component is  $x^i$ . If  $A$  is a symmetric positive semi-definite matrix, then  $A^{\frac{1}{2}}$  stands for its symmetric square-root.

## 2 Main result

In this section, we state our main result, Theorem 2.3, that extends Theorem 4.1 in Da Prato and Frankowska [13] to weaker regularity assumptions.

Since we are dealing with general coefficients  $b$  and  $\sigma$ , i.e. not necessarily Lipschitz coefficients, solutions to the stochastic differential equation (1.1) should be considered in the weak sense rather than in the strong sense. Existence is guaranteed by our condition  $(H_1)$ , together with our standing assumption of continuity of  $b$  and  $\sigma$ : there exist a filtered probability space  $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$  satisfying the usual conditions, a  $d$ -dimensional  $\mathbb{F}$ -Brownian motion  $W$  and a  $\mathbb{F}$ -adapted process  $X$  with continuous sample paths such that (1.1) holds  $\mathbb{P}$ -a.s. See e.g. [23, Theorems IV.2.3 and IV.2.4].

For later use, note that  $(H_1)$  implies that, for any positive integer  $p$ , there exists  $K_{p,x} > 0$  such that

$$\mathbb{E} [\|X_t - X_s\|^p] \leq K_{p,x} |t - s|^{\frac{p}{2}} \quad (2.1)$$

for all  $0 \leq s, t \leq 1$ . Hence, Kolmogorov's continuity criterion ensures that the sample paths of  $X$  are (locally)  $\eta$ -Hölder continuous for any  $\eta \in (0, \frac{1}{2})$  (up to considering a suitable modification).

**Remark 2.1.** *The collection  $\mathcal{Q}$  of possible distributions of  $X$  is entirely determined by the infinitesimal generator  $\mathcal{L}$  defined on the space of smooth functions  $\phi$  by  $\mathcal{L}\phi := D\phi b + \frac{1}{2}\text{Tr}[\sigma\sigma^\top D^2\phi]$ . Therefore,  $\mathcal{Q}$  is the same if  $\sigma$  is replaced by  $\tilde{\sigma}$  such that  $\tilde{\sigma}\tilde{\sigma}^\top = \sigma\sigma^\top$ , see e.g. [36, Remark 5.1.7]. Hence, we can reduce to the case where  $\sigma$  is the symmetric square-root of  $C$  on  $\mathcal{D}$ , which we will assume from now on.*

Before stating our main result, let us make precise the definition of stochastic invariance.

**Definition 2.2** (Stochastic invariance). *A closed subset  $\mathcal{D} \subset \mathbb{R}^d$  is said to be stochastically invariant with respect to the diffusion (1.1) if, for all  $x \in \mathcal{D}$ , there exists a weak solution  $(X, W)$  to (1.1) starting at  $X_0 = x$  such that  $X_t \in \mathcal{D}$  for all  $t \geq 0$ , almost surely.*

Our characterization of stochastic invariance reads as follows (see Propositions 3.5 and 4.2 below for the proof). From now on we use the same notation  $C$  for  $C$  defined as  $\sigma\sigma^\top$  on  $\mathcal{D}$  and for its extension defined in Assumption (H<sub>2</sub>).

**Theorem 2.3** (Invariance characterization). *Let  $\mathcal{D}$  be closed. Assume that  $b$ ,  $\sigma$  and  $C$  are continuous and satisfy assumptions (H<sub>1</sub>)-(H<sub>2</sub>). Then, the set  $\mathcal{D}$  is stochastically invariant with respect to the diffusion (1.1) if and only if*

$$\begin{cases} C(x)u = 0 & (2.2a) \\ \langle u, b(x) - \frac{1}{2} \sum_{j=1}^d DC^j(x)(CC^+)^j(x) \rangle \leq 0 & (2.2b) \end{cases}$$

for every  $x \in \mathcal{D}$  and for all  $u \in \mathcal{N}_{\mathcal{D}}^1(x)$ .

Clearly, the regularity conditions of Theorem 2.3 are weaker than those of Theorem 4.1 in Da Prato and Frankowska [13]. Let us immediately exemplify this by considering the case of the square-root process already mentioned in the introduction. Let  $\mathcal{D} = \mathbb{R}_+$ ,  $C(x) = \eta^2 x$  with  $\eta > 0$ , and consider the diffusion  $dX_t = b(X_t)dt + \eta\sqrt{X_t}dW_t$ . Since  $C(x)C(x)^+ = \mathbf{1}_{\{x>0\}}$  and  $\mathcal{N}_{\mathbb{R}_+}^1(x) = \mathbf{1}_{\{x=0\}}\mathbb{R}_-$ , Theorem 2.3 implies that  $\mathbb{R}_+$  is stochastically invariant if and only if  $b(0) \geq 0$ , while  $\sigma : x \in \mathbb{R}_+ \mapsto \eta\sqrt{x}$  is not differentiable at 0.

On the other hand, one can easily recover [13, Theorem 4.1] under their smoothness assumptions (recall Remark 2.1).

**Proposition 2.4.** *Fix  $\sigma \in \mathcal{C}_b^{1,1}(\mathbb{R}^d, \mathbb{S}^d)$  (i.e.  $\sigma$  is differentiable with a bounded and a globally Lipschitz derivative). Then  $C := \sigma^2 \in \mathcal{C}_{loc}^{1,1}(\mathbb{R}^d, \mathbb{S}_+^d)$  and*

$$\langle u, \sum_{j=1}^d D\sigma^j(x)\sigma^j(x) \rangle = \langle u, \sum_{j=1}^d DC^j(x)(CC^+)^j(x) \rangle, \quad \text{for all } x \in \mathcal{D} \text{ and } u \in \text{Ker } \sigma(x).$$

*Proof.* Fix  $x \in \mathcal{D}$  and  $u \in \text{Ker } \sigma(x)$ . By using Definition A.7 and Proposition A.8 in the Appendix, we first compute that

$$DC(x) = D(\sigma(x)^2) = (\sigma(x) \otimes I_d)D\sigma(x) + (I_d \otimes \sigma(x))D\sigma(x),$$

which clearly shows that  $C$  is  $\mathcal{C}_{loc}^{1,1}$ . It then follows from Proposition A.5 and the fact that  $u \in \text{Ker } \sigma(x)$  that

$$(I_d \otimes u^\top)DC(x)C(x)C(x)^+ = (\sigma(x) \otimes u^\top)D\sigma(x)C(x)C(x)^+.$$

Observe now that  $C(x)C(x)^+\sigma(x) = \sigma(x)$  since  $C(x) = \sigma(x)^2$  (use the spectral decomposition of  $\sigma$  as in Proposition A.2). Using Proposition A.5 again, the above implies that

$$\begin{aligned}\mathrm{Tr} \left[ (I_d \otimes u^\top) DC(x)C(x)C(x)^+ \right] &= \mathrm{Tr} \left[ \sigma(x)(I_d \otimes u^\top) D\sigma(x)C(x)C(x)^+ \right] \\ &= \mathrm{Tr} \left[ (I_d \otimes u^\top) D\sigma(x)\sigma(x) \right].\end{aligned}$$

Then, by Proposition A.5 and A.8,

$$\begin{aligned}\langle u, \sum_{j=1}^d DC^j(x)\sigma^j(x) \rangle &= \sum_{j=1}^d u^\top D(\sigma(x)e_j)\sigma(x)e_j \\ &= \sum_{j=1}^d u^\top (e_j^\top \otimes I_d) D\sigma(x)\sigma(x)e_j \\ &= \sum_{j=1}^d e_j^\top (I_d \otimes u^\top) D\sigma(x)\sigma(x)e_j \\ &= \mathrm{Tr} \left[ (I_d \otimes u^\top) D\sigma(x)\sigma(x) \right] \\ &= \mathrm{Tr} \left[ (I_d \otimes u^\top) DC(x)C(x)C(x)^+ \right] \\ &= \langle u, \sum_{j=1}^d DC^j(x)(CC^+)^j(x) \rangle,\end{aligned}\tag{2.3}$$

in which the last identity follows by reproducing exactly the same computations in the reverse order with  $C$  in place of  $\sigma$ .  $\square$

The following provides another formulation of (2.2b) that highlights the notion of projection on the image of  $C$ .

**Remark 2.5** (Interpretation of the projection formulation). *Fix  $x \in \partial\mathcal{D}$  and assume that the spectral decomposition of  $C$  at  $x$  takes the form  $C(x) = Q(x)\mathrm{diag}[\lambda_1(x), \dots, \lambda_r(x), 0, \dots, 0]Q(x)^\top$ , where  $Q(x)Q(x)^\top = I_d$  and  $\lambda_j(x) > 0$  for all  $1 \leq j \leq r$ . Hence, the  $r$ -first columns of  $Q(x)$ , denoted by  $(q_1, \dots, q_r) = (q_1(x), \dots, q_r(x))$ , span the image of  $C(x)$  and the projection matrix on the image of  $C(x)$  is given by  $C(x)C(x)^+ = \sum_{i=1}^r q_i q_i^\top$ , see Propositions A.3 and A.2 in the Appendix. Thus, by (2.3) in the proof of Proposition 2.4 and Proposition A.5 in the Appendix,*

$$\begin{aligned}\langle u, \sum_{j=1}^d DC^j(x)(CC^+)^j(x) \rangle &= \mathrm{Tr} \left[ (I_d \otimes u^\top) DC(x)C(x)C(x)^+ \right] \\ &= \sum_{j=1}^r \mathrm{Tr} \left[ (I_d \otimes u^\top) DC(x)q_i q_i^\top \right] \\ &= \sum_{j=1}^r \mathrm{Tr} \left[ q_i^\top (I_d \otimes u^\top) DC(x)q_i \right] \\ &= \sum_{j=1}^r u^\top (q_i^\top \otimes I_d) DC(x)q_i\end{aligned}$$

so that, by Proposition A.8,

$$\langle u, \sum_{j=1}^d DC^j(x)(CC^+)^j(x) \rangle = \langle u, \sum_{j=1}^r D(Cq_i)(x)q_i \rangle = \langle u, \sum_{j=1}^r D_{q_i}(Cq_i)(x) \rangle$$

in which  $D_{q_i}$  is the directional derivative with respect to  $q_i$ :

$$D_{q_i}(Cq_i)(x) := \lim_{t \rightarrow 0} \frac{C(x + tq_i)q_i - C(x)q_i}{t}.$$

Therefore (2.2b) reads  $\langle u, b(x) - \frac{1}{2} \sum_{j=1}^r D_{q_j}(Cq_j)(x) \rangle \leq 0$ . Otherwise stated,  $C$  is first projected onto the basis of the image of  $C(x)$  before being derived only in the directions of  $(q_1, \dots, q_r)$ . This is clearly consistent with (2.2a) that states that there cannot be any transverse diffusion of  $C(x)$  to the boundary. Therefore, the drift  $b(x)$  should only compensate the tangential diffusion given by the projection onto the image of  $C(x)$  in order to keep the diffusion in the domain.

Let us conclude this section with two additional comments.

**Remark 2.6.** In practical applications, the problem is usually phrased as follows: Given  $b : \mathbb{R}^d \mapsto \mathbb{R}^d$ ,  $C : \mathbb{R}^d \mapsto \mathbb{S}_+^d$ , and a closed set  $\mathcal{D}$  such that  $C(\mathcal{D}) \subset \mathbb{S}_+^d$ , can we find a  $\mathcal{D}$ -valued weak solution to (1.1), with  $\sigma\sigma^\top = C$  on  $\mathcal{D}$ . This is clearly an equivalent formulation to our problem.

**Remark 2.7** (Adding jumps). Note that jumps could be included in the dynamics of  $X$ . The extension of our main result to this setting would be similar to the extension of [13, Theorem 4.1] in Theorem A.5 of Cuchiero et al. [12].

### 3 Necessary conditions

In this section, we prove that the conditions of Theorem 2.3 are necessary for  $\mathcal{D}$  to be invariant. Our general strategy is similar to [8]. First, we consider a smooth function  $\phi : \mathbb{R}^d \mapsto \mathbb{R}$  such that  $\max_{\mathcal{D}} \phi = \phi(x)$ . Since  $\mathcal{D}$  is stochastically invariant,  $\phi(X_t) \leq \phi(x)$ , for all  $t \geq 0$ . Then, by applying Itô's Lemma twice, we obtain

$$\int_0^t \mathcal{L}\phi(X_s) ds + \int_0^t \left( D\phi\sigma(x) + \int_0^s \mathcal{L}(D\phi\sigma)(X_r) dr + \int_0^s D(D\phi\sigma)\sigma(X_r) dW_r \right)^\top dW_s \leq 0.$$

Recall Remark 2.1 for the definition of the infinitesimal generator  $\mathcal{L}$ . Given (now standard) estimates on the small time behavior of single and double stochastic integrals, see e.g. [8, 10], this readily implies

$$D\phi(x)\sigma(x) = 0 \quad \text{and} \quad \langle D\phi(x), b(x) - \frac{1}{2} \sum_{j=1}^d D\sigma^j(x)\sigma^j(x) \rangle \leq 0,$$

under appropriate regularity conditions. It remains to choose a suitable test function  $\phi$ , i.e. such that  $D\phi(x) = u^\top$ , to deduce that (2.2a)-(2.2b) must hold in the smooth case, recall Proposition 2.4.

In our setting, one can however not differentiate  $\sigma^j$  in general. To surround this problem the above can be rewritten in term of the covariance matrix  $C$ . The projection term in (2.2a)-(2.2b) will appear through a conditioning argument.

In order to separate the difficulties, we shall first consider the case where  $C$  admits a locally smooth spectral decomposition. The general case will be handled in Section 3.2 below.

#### 3.1 The case of distinct eigenvalues

As mentioned above, we shall first make profit of having distinct eigenvalues before considering the general case. The main idea consists in using the spectral decomposition of  $C$  in the form



$Q\Lambda Q^\top$  in which  $Q$  is an orthogonal matrix and  $\Lambda$  is diagonal positive semi-definite. Then, the dynamics of  $X$  can be written as

$$dX_t = b(X_t)dt + Q(X_t)\Lambda(X_t)^{\frac{1}{2}}dB_t$$

in which  $B = \int_0^\cdot Q(X_s)^\top dW_s$  is a Brownian motion. If  $Q$  and  $\Lambda$  are smooth enough, then we can apply the same ideas as the one exposed at the beginning of this section. An additional localization and conditioning argument will allow us to reduce to the case where  $\Lambda$  has only (strictly) positive entries. However, we need to ensure that  $Q$  and  $\Lambda^{\frac{1}{2}}$  are smooth enough, locally, which is essentially the case when  $C \in \mathcal{C}_{loc}^{1,1}(\mathbb{R}^d, \mathbb{S}^d)$  and  $C(x)$  has distinct and non-zero eigenvalues.

**Lemma 3.1.** *Assume that  $C \in \mathcal{C}_{loc}^{1,1}(\mathbb{R}^d, \mathbb{S}^d)$ . Let  $x \in \mathcal{D}$  be such that the spectral decomposition of  $C(x)$  is given by*

$$C(x) = Q(x)\text{diag}[\lambda_1(x), \dots, \lambda_r(x), 0, \dots, 0]Q(x)^\top \quad (3.1)$$

with  $\lambda_1(x) > \lambda_2(x) > \dots > \lambda_r(x) > 0$  and  $Q(x)Q(x)^\top = I_d$ ,  $r \leq d$ .

Then there exist an open (bounded) neighborhood  $N(x)$  of  $x$  and two measurable  $\mathbb{M}^d$ -valued functions on  $\mathbb{R}^d$ ,  $y \mapsto Q(y) := [q_1(y) \cdots q_d(y)]$  and  $y \mapsto \Lambda(y) := \text{diag}[\lambda_1(y), \dots, \lambda_d(y)]$  such that

- (i)  $C(y) = Q(y)\Lambda(y)Q(y)^\top$  and  $Q(y)Q(y)^\top = I_d$ , for all  $y \in \mathbb{R}^d$ ,
- (ii)  $\lambda_1(y) > \lambda_2(y) > \dots > \lambda_r(y) > \max\{\lambda_i(x), r+1 \leq i \leq d\} \vee 0$ , for all  $y \in N(x)$ ,
- (iii)  $\bar{\sigma} : y \mapsto \bar{Q}(y)\bar{\Lambda}(y)^{\frac{1}{2}}$  is  $C^{1,1}(N(x), \mathbb{M}^d)$ , in which  $\bar{Q} := [q_1 \cdots q_r \ 0 \cdots 0]$  and  $\bar{\Lambda} = \text{diag}[\lambda_1, \dots, \lambda_r, 0, \dots, 0]$ .

Moreover, we have:

$$\langle u, \sum_{j=1}^d D\bar{\sigma}^j(x)\bar{\sigma}^j(x) \rangle = \langle u, \sum_{j=1}^d DC^j(x)(CC^+)^j(x) \rangle, \quad \text{for all } u \in \text{Ker}(C(x)). \quad (3.2)$$

*Proof.* Note that the fact that  $(q_i)_{i \leq d}$  can be chosen measurable is guaranteed when  $(C, \Lambda)$  is measurable by the fact that each eigenvector solves a quadratic minimization problem, see e.g. [5, Proposition 7.33(p.153)]. Moreover, the continuity of the eigenvalues follows from Weyl's perturbation theorem, [6, Corollary III.2.6], and the smoothness of  $(\bar{\Lambda}, \bar{Q})$  is a consequence of [27, Theorem 1] since all the positive eigenvalues are simple and  $C$  is  $\mathcal{C}_{loc}^{1,1}(\mathbb{R}^d, \mathbb{S}^d)$ .

Let us now observe that any  $u \in \text{Ker}(C(x))$  satisfies

$$u^\top \bar{Q}(x) = u^\top \bar{\sigma}(x) = 0.$$

Since  $\bar{C} := \bar{\sigma}\bar{\sigma}^\top$  is differentiable at  $x$ , the product rule of Proposition A.8 combined with Proposition A.5 yields

$$\begin{aligned} (I_d \otimes u^\top)D\bar{C}(x) &= (I_d \otimes u^\top) \left[ (\bar{\sigma}(x) \otimes I_d)D\bar{\sigma}(x) + (I_d \otimes \bar{\sigma}(x))D\bar{\sigma}(x)^\top \right] \\ &= (\bar{\sigma}(x) \otimes u^\top)D\bar{\sigma}(x) \\ &= \bar{\sigma}(x)(I_d \otimes u^\top)D\bar{\sigma}(x). \end{aligned}$$

Observing that  $\bar{C} = \bar{\sigma}\bar{\sigma}^\top = C\bar{Q}\bar{Q}^\top$  and that  $\bar{Q}(x)\bar{Q}(x)^\top = C(x)C(x)^+$ , we get by similar computations:

$$\begin{aligned} (I_d \otimes u^\top)D\bar{C}(x) &= (I_d \otimes u^\top) \left[ (C(x)C(x)^+ \otimes I_d)DC(x) + (I_d \otimes C(x))D(\bar{Q}\bar{Q}^\top)(x) \right] \\ &= C(x)C(x)^+(I_d \otimes u^\top)DC(x). \end{aligned}$$

Combining the above leads to

$$\mathrm{Tr} \left[ (I_d \otimes u^\top) D\bar{\sigma}(x)\bar{\sigma}(x) \right] = \mathrm{Tr} \left[ (I_d \otimes u^\top) DC(x)C(x)C(x)^\top \right],$$

which proves (3.2) by similar computations as in the proof of (2.3).  $\square$

We can now adapt the arguments of [8]. In the following we use the notion of proximal cone:

$$\mathcal{N}_{\mathcal{D}}^{1,prox}(x) := \{u \in \mathbb{R}^d, \|u\| = d_{\mathcal{D}}(x+u)\},$$

in which  $d_{\mathcal{D}}$  is the distance function to  $\mathcal{D}$ . Note however that (2.2a)-(2.2b) holds at  $x$  for all proximal normals  $u \in \mathcal{N}_{\mathcal{D}}^{1,prox}(x)$  if and only if it holds for all  $u \in \mathcal{N}_{\mathcal{D}}^1(x)$ . Indeed,  $\mathcal{N}_{\mathcal{D}}^{1,prox}(x) \subset \mathcal{N}_{\mathcal{D}}^1(x)$  while  $\mathcal{N}_{\mathcal{D}}^1(x)$  is included in the closed convex envelope of the limsup of  $\mathcal{N}_{\mathcal{D}}^{1,prox}$  at  $x$ , see e.g. [2].

**Lemma 3.2.** *Assume that  $\mathcal{D}$  is stochastically invariant with respect to the diffusion (1.1). Let  $x \in \mathcal{D}$  and  $C$  be as in Lemma 3.1. Then, (2.2a) and (2.2b) hold at  $x$  for all  $u \in \mathcal{N}_{\mathcal{D}}^1(x)$ .*

*Proof.* It follows from the discussion before our lemma that it suffices to prove our claim for  $u \in \mathcal{N}_{\mathcal{D}}^{1,prox}(x)$ . Let  $(X, W)$  denote a weak solution starting at  $X_0 = x$  such that  $X_t \in \mathcal{D}$  for all  $t \geq 0$ . If  $x \notin \partial\mathcal{D}$ , then  $\mathcal{N}_{\mathcal{D}}^{1,prox}(x) = \{0\}$  and there is nothing to prove. We therefore assume from now on that  $x \in \partial\mathcal{D}$ . We fix  $u \in \mathcal{N}_{\mathcal{D}}^{1,prox}(x)$ .

Step 1. We first claim that there exists a function  $\phi \in \mathcal{C}_b^\infty(\mathbb{R}^d, \mathbb{R})$  with compact support in  $N(x)$  such that  $\max_{\mathcal{D}} \phi = \phi(x) = 0$  and  $D\phi(x) = u^\top$ . Indeed, it follows from [34, Chapter 6.E]

that one can find  $\kappa > 0$  such that  $\langle u, y - x \rangle \leq \frac{\kappa}{2} \|y - x\|^2$  for all  $y \in \mathcal{D}$ . Then, one can set  $\psi := \langle u, \cdot - x \rangle - \frac{\kappa}{2} \|\cdot - x\|^2$  and define  $\phi := \psi\rho$  in which  $\rho$  is a  $\mathcal{C}_b^\infty$  function with values in  $[0, 1]$ , compact support included in  $N(x)$ , and satisfying  $\rho = 1$  in a neighborhood of  $x$ .

Step 2. Since  $\mathcal{D}$  is invariant under the diffusion  $X$ ,  $\phi(X_t) \leq \phi(x)$ , for all  $t \geq 0$ . From now on, we use the notations of Lemma 3.1. By the above and Itô's lemma:

$$0 \geq \int_0^t \mathcal{L}\phi(X_s)ds + \int_0^t D\phi(X_s)\sigma(X_s)dW_s = \int_0^t \mathcal{L}\phi(X_s)ds + \int_0^t (D\phi Q\Lambda^{\frac{1}{2}}Q^\top)(X_s)dW_s$$

in which  $\mathcal{L}$  is the infinitesimal generator of  $X$ . Let us define the Brownian motion  $B = \int_0^\cdot Q(X_s)^\top dW_s$ , recall that  $Q$  is orthogonal, together with  $\bar{B} = \Lambda(x)\Lambda(x)^+B = (B^1, \dots, B^r, 0, \dots, 0)^\top$  and  $\bar{B}^\perp = (I_d - \Lambda(x)\Lambda(x)^+)B = (0, \dots, 0, B^{r+1}, \dots, B^d)$ , recall Proposition A.2. Since  $Q\Lambda^{\frac{1}{2}} = \bar{Q}\bar{\Lambda}^{\frac{1}{2}}$ , the above inequality can be written in the form

$$0 \geq \int_0^t \mathcal{L}\phi(X_s)ds + \int_0^t D\phi(X_s)\bar{\sigma}(X_s)d\bar{B}_s + \int_0^t (D\phi Q\Lambda^{\frac{1}{2}})(X_s)d\bar{B}_s^\perp.$$

Let  $(\mathcal{F}_s^{\bar{B}})_{s \geq 0}$  be the completed filtration generated by  $\bar{B}$ . By [26, Corollaries 2 and 3 of Theorem 5.13], [24, Lemma 14.2], and the fact that the martingale  $\bar{B}^\perp$  is independent of  $\bar{B}$ , we obtain

$$0 \geq \int_0^t \mathbb{E}_{\mathcal{F}_s^{\bar{B}}}[\mathcal{L}\phi(X_s)]ds + \int_0^t \mathbb{E}_{\mathcal{F}_s^{\bar{B}}}[D\phi(X_s)\bar{\sigma}(X_s)]d\bar{B}_s.$$

We now apply Lemma 3.3 below to  $(D\phi\bar{\sigma})(X)$  and use [26, Corollaries 2 and 3 of Theorem 5.13] and [24, Lemma 14.2] again to find a bounded adapted process  $\eta$  such that

$$0 \geq \int_0^t \theta_s ds + \int_0^t \left( \alpha + \int_0^s \beta_r dr + \int_0^s \gamma_r dB_r \right)^\top dB_s \quad (3.3)$$

where

$$\begin{aligned}\theta &:= \mathbb{E}_{\mathcal{F}^{\bar{B}}} [\mathcal{L}\phi(X.)] \quad , \quad \alpha^\top := (D\phi\bar{\sigma})(x) = u^\top Q(x)\Lambda(x)^{\frac{1}{2}} \\ \beta &:= \mathbb{E}_{\mathcal{F}^{\bar{B}}} [D(D\phi\bar{\sigma})(X.)b(X.) + \eta.] \quad , \quad \gamma := \mathbb{E}_{\mathcal{F}^{\bar{B}}} [D(D\phi\bar{\sigma})\bar{\sigma}(X.)],\end{aligned}$$

recall from Step 1 that  $D\phi(x) = u^\top$ .

**Step 3.** We now check that we can apply Lemma 3.4 below. First note that all the above processes are bounded. This follows from Lemma 3.1,  $(H_1)$  and the fact that  $\phi$  has compact support. In addition, given  $T > 0$ , the independence of the increments of  $\bar{B}$  implies that  $\theta_s = \mathbb{E}_{\mathcal{F}_T^{\bar{B}}} [\mathcal{L}\phi(X_s)]$  for all  $s \leq T$ . It follows that  $\theta$  is a.s. continuous at 0.

Similarly,  $\gamma = \mathbb{E}_{\mathcal{F}_T^{\bar{B}}} [D(D\phi\bar{\sigma})\bar{\sigma}(X.)]$  on  $[0, T]$ . Moreover, since  $D\phi\bar{\sigma}$  is  $\mathcal{C}^{1,1}$ ,  $F := D(D\phi\bar{\sigma})\bar{\sigma}$  is Lipschitz and Jensen's inequality combined with (2.1) implies that we can find  $L' > 0$  such that

$$\mathbb{E} \left[ \|\gamma_s - \gamma_r\|^4 \right] \leq \mathbb{E} \left[ \|F(X_s) - F(X_r)\|^4 \right] \leq L'|s - r|^2, \quad \text{for all } 0 \leq s, r \leq 1.$$

By Kolmogorov's continuity criterion, up to considering a suitable modification,  $\gamma$  has  $\epsilon$ -Hölder sample paths for all  $0 < \epsilon < \frac{1}{4}$ , in particular  $\int_0^t \|\gamma_s - \gamma_0\|^2 ds = O(t^{1+\epsilon})$  for  $0 < \epsilon < \frac{1}{2}$ .

**Step 4.** In view of Step 3, we can apply Lemma 3.4 to (3.3) to deduce that  $\alpha = 0$  and  $\theta_0 - \frac{1}{2} \text{Tr}(\gamma_0) \leq 0$ . The first equation implies that  $\alpha^\top \Lambda(x)^{\frac{1}{2}} Q^\top(x) = u^\top C(x) = 0$ , or equivalently  $C(x)u = 0$  since  $C(x)$  is symmetric. The second identity combined with  $D\phi(x) = u^\top$  and Proposition A.8 shows that

$$0 \geq \mathcal{L}\phi(x) - \frac{1}{2} \text{Tr} \left[ \bar{\sigma}^\top D^2\phi\bar{\sigma} + (I_d \otimes u^\top) D\bar{\sigma}\bar{\sigma} \right] (x) = u^\top b(x) - \frac{1}{2} \text{Tr} \left[ (I_d \otimes u^\top) D\bar{\sigma}\bar{\sigma} \right] (x),$$

which is equivalent to (2.2b) by (3.2) and similar computations as in the proof of (2.3).  $\square$

The rest of this section is dedicated to the proof of the two technical lemmas that were used above. Our first result is a slight extension of Itô's lemma to only  $\mathcal{C}^{1,1}$  function. It is based on a simple application of Komlós lemma (note that the assumption that  $f$  has a compact support in the following is just for convenience, it can obviously be removed by a localization argument, in which case the process  $\eta$  is only locally bounded).

**Lemma 3.3.** *Assume that  $b$  and  $\sigma$  are continuous and that there exists a solution  $(X, W)$  to (1.1). Let  $f \in \mathcal{C}^{1,1}(\mathbb{R}^d, \mathbb{R})$  have compact support. Then, there exists an adapted bounded process  $\eta$  such that*

$$f(X_t) = f(x) + \int_0^t (Df(X_s)b(X_s) + \eta_s) ds + \int_0^t Df(X_s)\sigma(X_s)dW_s$$

for all  $t \geq 0$ .

*Proof.* Since  $f \in \mathcal{C}^{1,1}$  has a compact support, we can find a sequence  $(f_n)_n$  in  $\mathcal{C}^\infty$  with compact support (uniformly) and a constant  $K > 0$  such that

- (i)  $\|D^2 f_n\| \leq K$ ,
- (ii)  $\|f_n - f\| + \|Df_n - Df\| \leq \frac{K}{n}$ ,

for all  $n \geq 1$ . This is obtained by considering a simple mollification of  $f$ . By applying Itô's Lemma to  $f_n(X)$ , we get

$$f_n(X_t) = f_n(x) + \int_0^t Df_n(X_s)b(X_s)ds + \int_0^t \eta_s^n ds + \int_0^t Df_n(X_s)\sigma(X_s)dW_s$$

in which  $\eta^n := \frac{1}{2}\text{Tr}[D^2 f_n \sigma \sigma^\top](X)$ . Since  $\sigma \sigma^\top$  is continuous, (i) above implies that  $(\eta^n)_n$  is uniformly bounded in  $L^\infty(dt \times d\mathbb{P})$ . By [15, Theorem 1.3], there exists  $(\tilde{\eta}^n) \in \text{Conv}(\eta^k, k \geq n)$  such that  $\tilde{\eta}^n \rightarrow \eta$   $dt \otimes d\mathbb{P}$  almost surely. Let  $N_n \geq 0$  and  $(\lambda_k^n)_{n \leq k \leq N_n} \subset [0, 1]$  be such that  $\tilde{\eta}^n = \sum_{k=n}^{N_n} \lambda_k^n \eta^k$  and  $\sum_{k=n}^{N_n} \lambda_k^n = 1$ . Set  $\tilde{f}_n := \sum_{k=n}^{N_n} \lambda_k^n f_k$ . Then,

$$\tilde{f}_n(X_t) = \tilde{f}_n(x) + \int_0^t D\tilde{f}_n(X_s)b(X_s)ds + \int_0^t \tilde{\eta}_s^n ds + \int_0^t D\tilde{f}_n(X_s)\sigma(X_s)dW_s. \quad (3.4)$$

By dominated convergence,  $\int_0^t \tilde{\eta}_s^n ds$  converges a.s. to  $\int_0^t \eta_s ds$ . Moreover, (ii) implies that

$$\|\tilde{f}_n(X_t) - f(X_t)\| \leq \sum_{k=n}^{N_n} \lambda_k^n \|\tilde{f}_k(X_t) - f(X_t)\| \leq \sum_{k=n}^{N_n} \lambda_k^n \frac{K}{k} \leq \frac{K}{n},$$

so that  $\tilde{f}_n(X_t)$  converges a.s. to  $f(X_t)$ . Similarly,

$$\int_0^t D\tilde{f}_n(X_s)\sigma(X_s)dW_s \rightarrow \int_0^t f(X_s)\sigma(X_s)dW_s \text{ and } \int_0^t D\tilde{f}_n(X_s)b(X_s)ds \rightarrow \int_0^t Df(X_s)b(X_s)ds$$

in  $L^2(\Omega, \mathcal{F}, \mathbb{P})$  as  $n \rightarrow \infty$ , and therefore a.s. after possibly considering a subsequence. It thus remains to send  $n \rightarrow \infty$  in (3.4) to obtain the required result.  $\square$

The following adapts [8, Lemma 2.1] to our setting, see also [7, 10, 11].

**Lemma 3.4.** *Let  $(W_t)_{t \geq 0}$  denote a standard  $d$ -dimensional Brownian motion on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ . Let  $\alpha \in \mathbb{R}^d$  and  $(\beta_t)_{t \geq 0}$ ,  $(\gamma_t)_{t \geq 0}$  and  $(\theta_t)_{t \geq 0}$  be adapted processes taking values respectively in  $\mathbb{R}^d$ ,  $\mathbb{M}^d$  and  $\mathbb{R}$  and satisfying*

- (1)  $\beta$  is bounded,
- (2)  $\int_0^t \|\gamma_s\|^2 ds < \infty$ , for all  $t \geq 0$ ,
- (3) there exists a random variable  $\eta > 0$  such that a.s.

$$\int_0^t \|\gamma_s - \gamma_0\|^2 ds = O(t^{1+\eta}) \quad \text{for } t \rightarrow 0, \quad (3.5)$$

- (4)  $\theta$  is a.s. continuous at 0.

Suppose that for all  $t \geq 0$

$$\int_0^t \theta_s ds + \int_0^t \left( \alpha + \int_0^s \beta_r dr + \int_0^s \gamma_r dW_r \right)^\top dW_s \leq 0. \quad (3.6)$$

Then,

- (a)  $\alpha = 0$ ,
- (b)  $-\gamma_0 \in \mathbb{S}_+^d$ ,
- (c)  $\theta_0 - \frac{1}{2} \text{Tr}(\gamma_0) \leq 0$ .

*Proof.* Since  $(W_t^i)^2 = 2 \int_0^t W_s^i dW_s^i + t$ , (3.6) reduces to

$$(\theta_0 - \frac{1}{2} \text{Tr}(\gamma_0))t + \sum_{i=1}^d \alpha^i W_t^i + \sum_{i=1}^d \frac{\gamma_0^{ii}}{2} (W_t^i)^2 + \sum_{1 \leq i \neq j \leq d} \gamma_0^{ij} \int_0^t W_s^i dW_s^j + R_t \leq 0,$$

where

$$\begin{aligned} R_t &= \int_0^t (\theta_s - \theta_0) ds + \int_0^t \left( \int_0^s \beta_r dr \right)^\top dW_s + \int_0^t \left( \int_0^s (\gamma_r - \gamma_0) dW_r \right)^\top dW_s \\ &=: R_t^1 + R_t^2 + R_t^3. \end{aligned}$$

In view of [8, Lemma 2.1], it suffices to show that  $R_t/t \rightarrow 0$  almost surely. To see this, first note that  $R_t^1 = o(t)$  a.s. since  $\theta$  is continuous at 0. Moreover, [10, Proposition 3.9] implies that  $R_t^2 = o(t)$  a.s., as  $\beta$  is bounded.

It remains to prove that  $R_t^3 = o(t)$  a.s. To see this, define  $M^{ij} = \gamma^{ij} - \gamma_0^{ij}$  and  $M^i = \int_0^\cdot \sum_{j=1}^d M_r^{ij} dW_r^j$  for all  $1 \leq i, j \leq d$ . The continuity assumption (3.5) implies that  $\langle M^i \rangle_s = O(s^{1+\eta})$  almost surely. By the Dambis-Dubins-Schwarz theorem,  $(M_s^i)_{s \geq 0}$  is therefore a time-changed Brownian motion, see e.g. [33, Theorem V.1.6]. By the law of iterated logarithm for Brownian motion  $(M_s^i)^2 = O(s^{1+\frac{\eta}{2}})$  almost surely. Hence,  $\langle R^3 \rangle_t = O(t^{2+\frac{\eta}{2}})$  almost surely. By applying the Dambis-Dubins-Schwarz theorem and the law of iterated logarithm again, we obtain that  $R_t^3 = o(t)$  a.s.  $\square$

### 3.2 The general case

We can now turn to the general case.

**Proposition 3.5** (Necessary conditions of Theorem 2.3). *Let the conditions of Theorem 2.3 hold and assume that  $\mathcal{D}$  is stochastically invariant with respect to the diffusion (1.1). Then conditions (2.2a) and (2.2b) hold for all  $x \in \mathcal{D}$  and  $u \in \mathcal{N}_{\mathcal{D}}^1(x)$ .*

*Proof.* If  $x$  lies in the interior of  $\mathcal{D}$ , then  $\mathcal{N}_{\mathcal{D}}^1(x) = \{0\}$  and there is nothing to prove. We therefore assume from now on that  $x \in \partial\mathcal{D}$ . Let  $\Lambda$  and  $Q$  be defined through the spectral decomposition of  $C$ , as in (3.1) but with only  $\lambda_1(x) \geq \dots \geq \lambda_d(x)$ . We shall perform a change of variable to reduce to the conditions of Lemma 3.2. To do this, we fix  $0 < \epsilon < 1$  and define

$$A^\epsilon = Q(x) \text{diag} \left[ \sqrt{(1-\epsilon)}, \sqrt{(1-\epsilon)^2}, \dots, \sqrt{(1-\epsilon)^d} \right] Q(x)^\top.$$

Since  $\mathcal{D}$  is invariant with respect to the diffusion  $X$ ,  $\mathcal{D}^\epsilon := A^\epsilon \mathcal{D}$  is invariant with respect to the diffusion  $X^\epsilon := A^\epsilon X$ . Note that

$$dX^\epsilon = b_\epsilon(X^\epsilon) dt + C_\epsilon(X^\epsilon)^{\frac{1}{2}} dW$$

in which

$$b_\epsilon := A^\epsilon b((A^\epsilon)^{-1} \cdot) \quad \text{and} \quad C_\epsilon := A^\epsilon C((A^\epsilon)^{-1} \cdot) (A^\epsilon)^\top$$

have the same regularity and growth as  $b$  and  $C$ . Moreover, the positive eigenvalues of  $C_\epsilon$  are all distinct at  $x^\epsilon := A^\epsilon x$ , as  $C_\epsilon(x^\epsilon) = Q(x) \text{diag} \left[ (1-\epsilon)\lambda_1(x), \dots, (1-\epsilon)^d \lambda_d(x) \right] Q(x)^\top$ . We can therefore apply Lemma 3.2 to  $(X^\epsilon, \mathcal{D}^\epsilon)$ :

$$\begin{cases} C_\epsilon(x^\epsilon) u_\epsilon = 0 & (3.7a) \\ \langle u_\epsilon, b_\epsilon(x^\epsilon) - \frac{1}{2} \sum_{j=1}^d DC_\epsilon^j(x^\epsilon) (C_\epsilon C_\epsilon^+)^j(x^\epsilon) \rangle \leq 0 & (3.7b) \end{cases}$$

for all  $u_\epsilon \in \mathcal{N}_{A^\epsilon \mathcal{D}}^1(x^\epsilon)$ . We now easily verify that  $\mathcal{N}_{A^\epsilon \mathcal{D}}^1(x^\epsilon) = (A^\epsilon)^{-1} \mathcal{N}_{\mathcal{D}}^1(x)$ , recall the definition in (1.4). Finally, by sending  $\epsilon \rightarrow 0$  in (3.7a) and (3.7b), we get by continuity:

$$\begin{cases} C(x)u = 0 \\ \langle u, b(x) - \frac{1}{2} \sum_{j=1}^d DC^j(x)(CC^+)^j(x) \rangle \leq 0, \end{cases}$$

for all  $u \in \mathcal{N}_{\mathcal{D}}^1(x)$ , which ends the proof.  $\square$

## 4 Sufficient conditions

In this section, we prove that the necessary conditions of Proposition 3.5 are also sufficient. We start by showing in Proposition 4.1 that (2.2a) and (2.2b) imply that the generator  $\mathcal{L}$  of  $X$  satisfies the *positive maximum principle*:  $\mathcal{L}\phi(x) \leq 0$  for any  $x \in \mathcal{D}$  and any function  $\phi \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$  such that  $\max_{\mathcal{D}} \phi = \phi(x) \geq 0$ , see e.g. [18, p165]. Then, classical arguments, mainly [18, Theorem 4.5.4], yield the existence of a solution to the corresponding martingale problem that stays in  $\mathcal{D}$ , see Proposition 4.2 below.

The following proposition is inspired by [13, Remark 5.6].

**Proposition 4.1.** *Under the assumptions of Theorem 2.3, assume that (2.2a)-(2.2b) hold for all  $x \in \mathcal{D}$  and  $u \in \mathcal{N}_{\mathcal{D}}^1(x)$ . Then, the generator  $\mathcal{L}$  satisfies the positive maximum principle.*

*Proof.* We fix  $x \in \mathcal{D}$ . For  $1 \leq j \leq d$ , let us consider the following deterministic control system:

$$\begin{cases} y'(t) = C(y(t))\sigma(x)^+ e_j \\ y(0) = x, \end{cases} \quad (4.1)$$

where  $\sigma(x)^+$  is the pseudoinverse of  $\sigma(x)$ . Since  $C$  is locally Lipschitz and verifies condition (2.2a), [13, Proposition 2.5] and the discussion before Lemma 3.2 imply that  $\mathcal{D}$  is invariant with respect to the deterministic control system (4.1). Then, by definition of the second order normal cone in (1.2),

$$\langle u, y(\sqrt{h}) - x \rangle + \frac{1}{2} \langle v(y(\sqrt{h}) - x), y(\sqrt{h}) - x \rangle \leq o(\|y(\sqrt{h}) - x\|^2)$$

for any  $(u, v) \in \mathcal{N}_{\mathcal{D}}^2(x)$ . On the other hand, since  $C$  is  $\mathcal{C}_{loc}^{1,1}$ , a Taylor expansion around 0 yields

$$y(\sqrt{h}) = x + \sqrt{h}C(x)\sigma(x)^+ e_j + \frac{h}{2}(e_j^\top \sigma(x)^+ \otimes I_d)DC(x)C(x)\sigma(x)^+ e_j + o(h),$$

recall Proposition A.8 and note that  $(\sigma^+)^{\top} = \sigma^+$  since  $\sigma$  is symmetric. Now observe that  $u \in \mathcal{N}_{\mathcal{D}}^1(x)$  whenever  $(u, v) \in \mathcal{N}_{\mathcal{D}}^2(x)$ . In particular,  $u^\top C(x) = 0$  under (2.2a). Combining the above, and recalling Proposition A.5 then leads to

$$\frac{h}{2} e_j^\top (\sigma(x)^+ \otimes u^\top) DC(x)C(x)\sigma(x)^+ e_j + \frac{h}{2} e_j^\top \sigma(x)^+ C(x)vC(x)\sigma(x)^+ e_j \leq o(h).$$

Note that  $\sigma^+ \sigma^+ = C^+$  and that  $C\sigma^+ \sigma^+ C = CC^+ C = C$ , see e.g. Definition A.1 and Proposition A.2, and recall that  $(\sigma(x)^+ \otimes u^\top) = \sigma(x)^+(I_d \otimes u^\top)$  by Proposition A.5. Then, dividing the above by  $h/2$  and sending  $h \rightarrow 0$  before summing over  $1 \leq j \leq d$  yields

$$\text{Tr} \left( (I_d \otimes u^\top) DC(x)C(x)C(x)^+ \right) + \text{Tr} (vC(x)) \leq 0.$$

In view of (2.2b) and (2.3), this shows that

$$\langle b(x), u \rangle + \frac{1}{2} \text{Tr}(vC(x)) \leq \langle u, b(x) - \frac{1}{2} \sum_{j=1}^d DC^j(x)(CC^+)^j(x) \rangle \leq 0$$

for all  $(u, v) \in \mathcal{N}_{\mathcal{D}}^2(x)$ . To conclude, it remains to observe that  $(D\phi(x), D^2\phi(x)) \in \mathcal{N}_{\mathcal{D}}^2(x)$  whenever  $\phi \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$  is such that  $\max_{\mathcal{D}} \phi = \phi(x) \geq 0$ . Hence,  $\mathcal{L}\phi(x) \leq 0$ .  $\square$

**Proposition 4.2** (Sufficient conditions of Theorem 2.3). *Under the assumptions of Theorem 2.3, assume that conditions (2.2a) and (2.2b) hold for all  $x \in \mathcal{D}$  and  $u \in \mathcal{N}_{\mathcal{D}}^1(x)$ . Then,  $\mathcal{D}$  is stochastically invariant with respect to the diffusion (1.1).*

*Proof.* We already know from Proposition 4.1 that  $\mathcal{L}$  satisfies the positive maximum principle. Then, [18, Theorem 4.5.4] yields the existence of a solution to the martingale problem associated to  $\mathcal{L}$  with sample paths in the space of càdlàg functions with values in  $\mathcal{D}^{\Delta} := \mathcal{D} \cup \{\Delta\}$ , the one-point compactification of  $\mathcal{D}$ . The discussion preceding [9, Proposition 3.2] and [18, Proposition 5.3.5], recall our linear growth conditions ( $H_1$ ), then shows that the solution has a modification with continuous sample paths in  $\mathcal{D}$ . Finally, [18, Theorem 5.3.3] implies the existence of a weak solution  $(X, W)$  such that  $X_t \in \mathcal{D}$  for all  $t \geq 0$  almost surely.  $\square$

## 5 A simple example of application

As already stated, Theorem 2.3 can be directly applied to a large class of diffusions, e.g. affine diffusions [17, 20, 35] and polynomial preserving diffusions [19, 25], not only for closed subsets of  $\mathbb{R}^d$ , but even when  $\mathcal{D} = \mathbb{S}^d$  (as in [12]) since  $\mathbb{S}^d$  can be identified with  $\mathbb{R}^{\frac{d(d+1)}{2}}$  by using the half-vectorization operator. In this section, we treat one single, yet rich, application in order to highlight the simplicity of our methodology. Namely, we explicit the conditions (2.2a) and (2.2b) for affine diffusions living in the parabolic domain:

$$\mathcal{D} = \{(\bar{x}, \tilde{x}) \in \mathbb{R}^2 : \tilde{x} \geq \bar{x}^2\}.$$

This example was initially introduced in Duffie *et al.* [17, Section 12.2] and was completely treated later in Gourieroux and Sufana [22]. Our aim is therefore to derive [22, Proposition 2] by using our main Theorem 2.3 and simple calculus.

We say that  $b : x \mapsto b(x) := (\bar{b}(x), \tilde{b}(x)) \in \mathbb{R}^2$  and  $C : x \mapsto C(x) \in \mathbb{S}^2$  are of affine form if there exists  $\bar{b}^i, \tilde{b}^i \in \mathbb{R}$  and  $A^i \in \mathbb{S}^2$ ,  $1 \leq i \leq 3$ , such that

$$\begin{aligned} \bar{b}(x) &= \bar{b}^0 + \bar{b}^1 \bar{x} + \bar{b}^2 \tilde{x} \\ \tilde{b}(x) &= \tilde{b}^0 + \tilde{b}^1 \bar{x} + \tilde{b}^2 \tilde{x} \\ C(x) &= A^0 + A^1 \bar{x} + A^2 \tilde{x}, \end{aligned}$$

for all  $x = (\bar{x}, \tilde{x}) \in \mathbb{R}^2$ . Then it is clear that  $b$  and  $C$  are  $\mathcal{C}^\infty$  and satisfy the linear growth conditions ( $H_1$ ). Moreover, the first order normal cone at a boundary element  $x = (\bar{x}, \bar{x}^2)$  can easily be computed as

$$\mathcal{N}_{\mathcal{D}}^1(x) = \begin{pmatrix} -2\bar{x} \\ 1 \end{pmatrix} \mathbb{R}_-.$$

In this case, Theorem 2.3 says that  $\mathcal{D}$  is invariant if and only if (ii) of the following proposition holds.

**Proposition 5.1.** *Let  $\mathcal{D} = \{(\bar{x}, \tilde{x}) \in \mathbb{R}^2 : \tilde{x} \geq \bar{x}^2\}$ . Then, the following assertions are equivalent:*

(i)  *$b$  and  $C$  are of affine form and conditions  $(H_2)$ , (2.2a) and (2.2b) hold for all  $x \in \mathcal{D}$  and  $u \in \mathcal{N}_{\mathcal{D}}^1(x)$ .*

(ii) *There exists  $\alpha \geq 0$  such that*

(a)

$$C(x) = \alpha \begin{pmatrix} 1 & 2\bar{x} \\ 2\bar{x} & 4\tilde{x} \end{pmatrix}, \text{ for all } x = (\bar{x}, \tilde{x}) \in \mathcal{D}, \quad (5.1)$$

(b)  $\bar{b}^2 = 0$  and

$$\begin{cases} \tilde{b}^2 > 2\bar{b}^1 & \text{and} & (\tilde{b}^1 - 2\bar{b}^0)^2 < 4(\tilde{b}^2 - 2\bar{b}^1)(\tilde{b}^0 - \alpha) \\ \text{or} & & \\ \tilde{b}^2 = 2\bar{b}^1, & \tilde{b}^1 = 2\bar{b}^0 & \text{and} & \tilde{b}^0 \geq \alpha \end{cases}. \quad (5.2)$$

*Proof.* Step 1. We first show that (i) implies (ii).

(1.a) The form of the diffusion matrix  $C(x) \in \mathbb{S}_+^2$  is derived by using  $(H_2)$  and (2.2a). Indeed, let  $x = (\bar{x}, \tilde{x}) \in \partial\mathcal{D}$ , i.e.  $\tilde{x} = \bar{x}^2$ . Then, (2.2a) implies that the two rows  $C_1(x)$  and  $C_2(x)$  of  $C(x)$  must satisfy

$$C_1(x)^\top, C_2(x)^\top \in \{(w_1, w_2) \in \mathbb{R}^2, w_2 = 2\bar{x}w_1\}.$$

Since  $C(x)$  is symmetric, its components  $(C_{ij}(x))_{i,j \leq 2}$  should verify  $C_{22}(x) = 4\bar{x}^2 C_{11}(x)$ . Since  $C$  is affine in  $(\bar{x}, \tilde{x}^2)$ , then necessarily  $C_{11}(x)$  is constant (or else  $C_{22}(x)$  would have at least a polynomial dependence of order 3 in  $\bar{x}$ ). Therefore, there exists  $\alpha$  such that  $C(x)$  has the form (5.1) at  $x$ , in which  $\alpha \geq 0$  to ensure that  $C(0) \in \mathbb{S}_+^2$ . Finally,  $C$  needs to have the same form (5.1) on the whole state space  $\mathcal{D}$ , since it is affine.

(1.b) We now derive the form of the drift vector  $b(x) = (\bar{b}(x), \tilde{b}(x)) \in \mathbb{R}^2$  by using (2.2a). Given  $\bar{x} \in \mathbb{R}$ ,  $x = (\bar{x}, \bar{x}^2) \in \partial\mathcal{D}$ . If  $\alpha > 0$ , then  $u(x) = (-2\bar{x}, 1)^\top$  spans the kernel of  $C(x)$ . Therefore, by Proposition A.3,

$$C(x)C(x)^+ = I_2 - \frac{1}{\|u(x)\|^2} u(x)u(x)^\top = \frac{1}{1 + 4\bar{x}^2} \begin{pmatrix} 1 & 2\bar{x} \\ 2\bar{x} & 4\bar{x}^2 \end{pmatrix}.$$

Moreover, the Jacobian of the two columns of  $C(x)$  are given by

$$DC^1(x) = \alpha \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix} \text{ and } DC^2(x) = \alpha \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}.$$

Let us now fix  $\mu \in \mathbb{R}_-$ . Then  $u := \mu \begin{pmatrix} -2\bar{x} \\ 1 \end{pmatrix} \in \mathcal{N}_{\mathcal{D}}^1(x)$  and elementary computations show that condition (2.2b) is equivalent to

$$-2\bar{b}^2 \bar{x}^3 + (\tilde{b}^2 - 2\bar{b}^1) \bar{x}^2 + (\tilde{b}^1 - 2\bar{b}^0) \bar{x} + \tilde{b}^0 - \alpha \geq 0, \quad \text{for all } \bar{x} \in \mathbb{R},$$

which is equivalent to (5.2).

If  $\alpha = 0$ , then  $C(x)C(x)^+ = 0$  and the same conclusion holds.

Step 2. We now show that (ii) implies (i). The condition (5.1) clearly implies (2.2a) and  $(H_2)$  since  $\det(C(x)) = 4\alpha(\tilde{x} - \bar{x}^2) \geq 0$  and  $\tilde{x} \geq 0$  for all  $(\bar{x}, \tilde{x}) \in \mathcal{D}$ . Moreover, (5.2) leads to (2.2b) by the same computations as above.  $\square$



## 6 Additional remark on the boundary non-attainment

In this last section, we provide a sufficient condition for the stochastic invariance of the interior of  $\mathcal{D}$ , when  $\mathcal{D}$  has a smooth boundary. The result is a direct implication of [35, Proposition 3.5] derived with the help of *McKean's argument* (see [30, Section 4]). Moreover, we extend the tractable conditions of [35, Proposition 3.7] given for affine diffusions. Our result could be easily used in the context of polynomial preserving diffusions for instance.

**Proposition 6.1.** *Let  $\mathcal{D} \subset \mathbb{R}^d$  be closed with a non-empty interior  $\mathring{\mathcal{D}}$  that is a maximal connected subset of  $\{x, \Phi(x) < 0\}$  where  $\Phi \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$  such that  $\partial\mathcal{D} = \Phi^{-1}(0)$ . Assume that  $b$  and  $C$  are continuous and satisfy assumptions  $(H_1)$ - $(H_2)$ . Moreover, assume that  $C \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{S}_+^d)$ . Then  $\mathring{\mathcal{D}}$  is stochastically invariant if there exists  $v \in \mathbb{R}^d$  such that*

$$\begin{cases} D\Phi(x)C(x) = \Phi(x)v^\top & (6.1a) \\ \langle D\Phi(x), b(x) - \frac{1}{2} \sum_{j=1}^d DC^j(x)e_j \rangle \leq 0 & (6.1b) \end{cases}$$

for all  $x \in \mathcal{D}$ .

*Proof.* Fix  $x \in \mathring{\mathcal{D}}$ . By differentiating (6.1a) with the help of Propositions A.8 and A.5, we obtain

$$vD\Phi(x) = (C(x) \otimes I_1)D^2\Phi(x) + (I_d \otimes D\Phi(x))DC(x) = C(x)D^2\Phi(x) + (I_d \otimes D\Phi(x))DC(x),$$

which, combined with (6.1b), leads to

$$\begin{aligned} \langle D\Phi(x), b(x) \rangle &\leq \frac{1}{2} \text{Tr} \left[ (I_d \otimes D\Phi(x)^\top) DC(x) \right] \\ &= -\frac{1}{2} \text{Tr} \left[ C(x) D^2\Phi(x) \right] + \frac{1}{2} D\Phi(x)v \\ &= -\frac{1}{2} \text{Tr} \left[ C(x) D^2\Phi(x) \right] + \frac{1}{2} \Phi(x)^{-1} D\Phi(x) C(x) D\Phi(x)^\top. \end{aligned}$$

We conclude by using [35, Proposition 3.5] (after a change of the sign, since  $\mathcal{D}$  is assumed to be a connected subset of  $\{x, \Phi(x) > 0\}$  in [35, Proposition 3.5]).  $\square$

**Example 6.2.** (i) Square root process: *Let us consider again the process defined by  $dX_t = b(X_t)dt + \eta\sqrt{X_t}dW_t$ , for some  $\eta > 0$ , on  $\mathcal{D} = \mathbb{R}_+$ . Then,  $\Phi : x \mapsto -x$  and (6.1a)-(6.1b) are equivalent to  $v = \eta^2$  and  $b(0) \geq \frac{\eta^2}{2}$ . These are the well known conditions for the boundary non-attainment of the square-root process.*

(ii) Affine diffusions: *More generally, let  $\mathcal{D} \subset \mathbb{R}^d$  satisfy the assumptions of Proposition 6.1 and take  $C(x) = A^0 + \sum_{j=1}^d A^j x^j$  for some  $A^j \in \mathbb{S}^d$ ,  $1 \leq j \leq d$ . Then differentiating  $C$  shows that condition (6.1b) is equivalent to  $\langle D\Phi(x), b(x) - \frac{1}{2} \sum_{j=1}^d (A^j)^j \rangle \leq 0$  yielding [35, Proposition 3.7].*

(iii) Jacobi diffusion: *Set  $\mathcal{D} = (0, 1]$  and consider a polynomial preserving diffusion  $X$  on  $\mathcal{D}$ , i.e.  $b$  is affine and  $C$  is a polynomial of degree two. Theorem 2.3 applied on  $[0, 1]$  immediately yields that the dynamics of  $X$  must be of the form  $dX_t = \kappa(\theta - X_t)dt + \eta\sqrt{X_t(1 - X_t)}dW_t$  where  $\kappa, \eta \geq 0$  and  $0 \leq \theta \leq 1$ . Now a localized version of Proposition 6.1 shows that  $\mathcal{D} = (0, 1]$  is stochastically invariant under the additional condition that  $\kappa\theta \geq \frac{\eta^2}{2}$ .*

Proposition 6.1 is important in practice since it gives, in many cases, the existence and the uniqueness of a global *strong solution* to (1.1) as discussed in the following remark.

**Remark 6.3.** Let  $\mathcal{D}$  be as in Proposition 6.1. Assume that  $C \in \mathcal{C}^2(\mathring{\mathcal{D}}, \mathbb{S}_+^d)$  and that  $b$  is locally Lipschitz (which is clearly the case for affine and polynomial diffusions). By [21, Remark 1 page 131] the square-root of  $C$  is locally Lipschitz on  $\mathring{\mathcal{D}}$ . Therefore, when the boundary is never attained, (1.1) starting from any element  $x \in \mathring{\mathcal{D}}$  admits a global strong solution and pathwise-uniqueness holds.

## A Matrix tools

For the reader's convenience, we collect in this Appendix some definitions and properties of matrix tools intensively used in the proofs throughout the article. For a complete review and proofs we refer to [28, 29, 32].

We start by recalling the definition of the Moore-Penrose pseudoinverse which generalizes the concept of invertibility of square matrices, to non-singular and non-square matrices. In the following, we denote by  $\mathbb{M}^{m,n}$  the collection of  $m \times n$  matrices.

**Definition A.1** (Moore-Penrose pseudoinverse). Fix  $A \in \mathbb{M}^{m,n}$ . The Moore-Penrose pseudoinverse of  $A$  is the unique  $n \times m$  matrix  $A^+$  satisfying:  $AA^+A = A$ ,  $A^+AA^+ = A^+$ ,  $AA^+$  and  $A^+A$  are Hermitian.

**Proposition A.2.** If  $A \in \mathbb{M}^d$  has the spectral decomposition  $Q\Lambda Q^\top$  for some orthogonal matrix  $Q \in \mathbb{M}^d$  and a diagonal matrix  $\Lambda = \text{diag}[(\lambda_i)_{i \leq d}] \in \mathbb{M}^d$ . Then,  $A^+ = Q\Lambda^+Q^\top$  in which  $\Lambda^+ = \text{diag}[(\lambda_i^{-1}\mathbb{1}_{\{\lambda_i \neq 0\}})_{i \leq d}]$ , and  $AA^+ = Q\text{diag}[(\mathbb{1}_{\{\lambda_i \neq 0\}})_{i \leq d}]Q^\top$ . If moreover  $A$  is positive semi-definite and  $B = A^{\frac{1}{2}}$ , then  $B^+ = Q(\Lambda^+)^{\frac{1}{2}}Q^\top$ .

**Proposition A.3.** If  $A \in \mathbb{M}^{m,n}$ , then  $AA^+$  is the orthogonal projection on the image of  $A$ .

We now collect some useful identities on the Kronecker product.

**Definition A.4** (Kronecker product). Let  $A = (a_{ij})_{i \leq m_1, j \leq n_1} \in \mathbb{M}^{m_1, n_1}$  and  $B \in \mathbb{M}^{m_2, n_2}$ . The Kronecker product  $(A \otimes B)$  is defined as the  $m_1 m_2 \times n_1 n_2$  matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n_1}B \\ \vdots & & \vdots \\ a_{m_1 1}B & \cdots & a_{m_1 n_1}B \end{pmatrix}$$

**Proposition A.5.** Let  $A$  and  $B$  be as in Definition A.4,  $C \in \mathbb{M}^{n_1, n_3}$  and  $D \in \mathbb{M}^{n_2, n_4}$ . Then,

$$\begin{aligned} (A \otimes B)(C \otimes D) &= (AC \otimes BD) \\ A \otimes B &= A(I_{n_1} \otimes B) \text{ if } m_2 = 1 \\ A \otimes B &= B(A \otimes I_{n_2}) \text{ if } m_1 = 1. \end{aligned}$$

The following definitions extend the concept of Jacobian matrix and show how to nicely stack the partial derivatives of a matrix-valued function  $F : \mathbb{M}^{n,q} \mapsto \mathbb{M}^{m,p}$  by using the vectorization operator (see [29, Chapter 9]).

**Definition A.6** (Vectorization operator). Let  $A \in \mathbb{M}^{m,n}$ . The vectorization operator  $\text{vec}$  transforms the matrix into a vector in  $\mathbb{R}^{mn}$  by stacking all the columns of the matrix  $A$  one underneath the other.

**Definition A.7** (Jacobian matrix). *Let  $F$  be a differentiable map from  $\mathbb{M}^{n,q}$  to  $\mathbb{M}^{m,p}$ . The Jacobian matrix  $DF(X)$  of  $F$  at  $X$  is defined as the following  $mp \times nq$  matrix:*

$$DF(X) = \frac{\partial \text{vec}(F(X))}{\partial \text{vec}(X)^\top}$$

**Proposition A.8** (Product rule). *Let  $G$  be a differentiable map from  $\mathbb{M}^{n,q}$  to  $\mathbb{M}^{m,p}$  and  $H$  be a differentiable map from  $\mathbb{M}^{n,q}$  to  $\mathbb{M}^{p,l}$ . Then,  $D(GH) = (H^\top \otimes I_m)DG + (I_l \otimes G)DH$ .*

## References

- [1] AUBIN, J. P., & DOSS, H. (2003). *Characterization of stochastic viability of any nonsmooth set involving its generalized contingent curvature*. Stochastic analysis and applications, 21(5), 955-981.
- [2] AUBIN, J. P., & FRANKOWSKA, H. (1990). *Set-valued analysis*. Birkhäuser, Boston, Basel, Berlin.
- [3] BARDI, M., & GOATIN, P. (1999). *Invariant sets for controlled degenerate diffusions: a viscosity solutions approach*. In Stochastic analysis, control, optimization and applications (pp. 191-208). Birkhäuser Boston.
- [4] BARDI, M., & JENSEN, R. (2002). *A geometric characterization of viable sets for controlled degenerate diffusions*. Set-Valued Analysis, 10(2-3), 129-141.
- [5] BERTSEKAS, D. P., & SHREVE, S. E. (1978). *Stochastic optimal control: The discrete time case* (Vol. 23, pp. 15-24). New York: Academic Press.
- [6] BHATIA, R. (1997). *Matrix analysis*. Springer-Verlag, New York, Volume 169 of Graduate Texts in Mathematics.
- [7] BRUDER, B. (2005). *Super-replication of European options with a derivative asset under constrained finite variation strategies*. Preprint HAL <http://hal.archives-ouvertes.fr/hal-00012183/fr>.
- [8] BUCKDAHN, R., QUINCAMPOIX, M., RAINER, C., & TEICHMANN, J. (2010). *Another proof for the equivalence between invariance of closed sets with respect to stochastic and deterministic systems*. Bulletin des sciences mathématiques, 134(2), 207-214.
- [9] CHERIDITO, P., FILIPOVIĆ, D., & YOR, M. (2005). *Equivalent and absolutely continuous measure changes for jump-diffusion processes*. Annals of applied probability, 1713-1732.
- [10] CHERIDITO, P., SONER, H. M., & TOUZI, N. (2005a). *Small time path behavior of double stochastic integrals and applications to stochastic control*. The Annals of Applied Probability, 15(4), 2472-2495.
- [11] CHERIDITO, P., SONER, H. M., & TOUZI, N. (2005b). *The multi-dimensional super-replication problem under gamma constraints*. In Annales de l'IHP Analyse non linéaire (Vol. 22, No. 5, pp. 633-666).
- [12] CUCHIERO, C., FILIPOVIĆ, D., MAYERHOFER, E., & TEICHMANN, J. (2011). *Affine processes on positive semidefinite matrices*. The Annals of Applied Probability, 21(2), 397-463.

- [13] DA PRATO, G., & FRANKOWSKA, H. (2004). *Invariance of stochastic control systems with deterministic arguments*. Journal of differential equations, 200(1), 18-52.
- [14] DA PRATO, G., & FRANKOWSKA, H. (2007). *Stochastic viability of convex sets*. Journal of mathematical analysis and applications, 333(1), 151-163.
- [15] DELBAEN, F., & SCHACHERMAYER, W. (1999). *A compactness principle for bounded sequences of martingales with applications*. In Seminar on stochastic analysis, random fields and applications (pp. 137-173). Birkhäuser Basel.
- [16] DOSS, H. (1977). *Liens entre équations différentielles stochastiques et ordinaires*. In Annales de l'IHP Probabilités et statistiques (Vol. 13, No. 2, pp. 99-125).
- [17] DUFFIE, D., FILIPOVIĆ, D., & SCHACHERMAYER, W. (2003). *Affine processes and applications in finance*. Annals of applied probability, 984-1053.
- [18] ETHIER, S. N., & KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York.
- [19] FILIPOVIĆ, D., & LARSSON, M. (2014). *Polynomial preserving diffusions and applications in finance*. Swiss Finance Institute Research Paper, (14-54).
- [20] FILIPOVIĆ, D., & MAYERHOFER, E. (2009). *Affine diffusion processes: theory and applications*. Advanced Financial Modelling, 8, 125.
- [21] FRIEDMAN, A. (1976). *Stochastic differential equations and applications* (Vol. 2). Academic press.
- [22] GOURIEROUX, C., & SUFANA, R. (2006). *A Classification of Two-Factor Affine Diffusion Term Structure Models*. Journal of Financial Econometrics, 4(1), 31-52.
- [23] IKEDA, N., & WATANABE, S. (1981). *Stochastic differential equations and diffusion processes*. North-Holland Publishing Company, Amsterdam.
- [24] KURTZ, T. G. (2001). *Lectures on Stochastic Analysis*. Department of Mathematics and Statistics, University of Wisconsin, Madison, WI, 53706-1388.
- [25] LARSSON, M., AND PULIDO, S. (2015). *Polynomial preserving diffusions on compact quadric sets*. arXiv preprint arXiv:1511.03554.
- [26] LIPTSER, R., & SHIRYAEV, A. N. (2001). *Statistics of random Processes: I. general Theory* (Vol. 2). Springer Science & Business Media.
- [27] MAGNUS, J. R. (1985). *On differentiating eigenvalues and eigenvectors*. Econometric Theory, 1(02), 179-191.
- [28] MAGNUS, J. R., & NEUDECKER, H. (1980). *The elimination matrix: some lemmas and applications*. SIAM Journal on Algebraic Discrete Methods, 1(4), 422-449.
- [29] MAGNUS, J. R., & NEUDECKER, H. (1988). *Matrix differential calculus with applications in statistics and econometrics*. Wiley series in probability and mathematical statistics Show all parts in this series.
- [30] MAYERHOFER, E., PFAFFEL, O., & STELZER, R. (2011). *On strong solutions for positive definite jump diffusions*. Stochastic processes and their applications, 121(9), 2072-2086.

- [31] MIJATOVIĆ, A., & URUSOV, M. (2014). *On the loss of the semimartingale property at the hitting time of a level.* Journal of Theoretical Probability, 1-31.
- [32] NEUDECKER, H. (1986). *Symmetry, 0-1 matrices and jacobians.* Econometric Theory, 2, 157-190.
- [33] REVUZ, D., & YOR, M. (1999). *Continuous martingales and Brownian motion.* Third edition. Springer, Berlin.
- [34] ROCKAFELLAR, R. T., & WETS, R. J. B. (1998). *Variational analysis* (Vol. 317). Springer-Verlag Berlin.
- [35] SPREIJ, P., & VEERMAN, E. (2012). *Affine diffusions with non-canonical state space.* Stochastic Analysis and Applications, 30(4), 605-641.
- [36] STROOCK, D. W., & VARADHAN, S. S. (1979). *Multidimensional diffusion processes.* Springer.
- [37] TAPPE, S. (2009). *Stochastic invariance of closed, convex sets with respect to jump-diffusions.* Preprint.