Stochastic invariance of closed sets with non-Lipschitz coefficients

Eduardo Abi Jaber [*1,2], Bruno Bouchard[†1] and Camille Illand [‡2]

[1]Université Paris-Dauphine, PSL University, CNRS, CEREMADE, 75016 Paris, France.
[2]AXA Investment Managers, Multi Asset Client Solutions, Quantitative Research, 6 place de la Pyramide, 92908 Paris - La Défense, France.

June 19, 2018

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 applied to construct affine and polynomial diffusions on any arbitrary closed set.

Keywords: Stochastic differential equation, stochastic invariance, affine diffusions, polynomial 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 σ satisfy the following linear growth conditions: there exists $L > 0$ such that

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

and we consider a weak solution of the stochastic differential equation

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

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, Josef Teichmann and Damir Filipović for very fruitful discussions and comments.
‡camille.illand@axa-im.com

precise formulation.)

The first stochastic invariance results can be found in Stroock and Varadhan [38], Friedman [23] and Doss [18] . 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 [4] and Bardi and Jensen [5]. Aubin and Doss [2] used the notion of curvature, while Da Prato and Frankowska [15] provided a characterization in terms of the Stratonovich drift. For a closed convex set, the distance function was used in Da Prato and Frankowska [16], and the invariance was characterized for affine jump-diffusions in Tappe [40].
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 σ. 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 [5], the stochastic invariance is characterized by using Nagumo-type geometric conditions on the second order normal cone. Their main result states that the closed set $\mathcal{D}$ is stochastically invariant if and only if

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

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

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

Here, $\mathbb{S}^d$ stands for the cone of symmetric $d \times d$ matrices. In practice, we face two restrictions. Prior to deriving the conditions on b and σ, 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. [5, Example 1]. However, it is much more challenging in general, by lack of efficient techniques. This renders the result of [5] difficult to use in practice. This also corresponds to the positive maximum principle of Ethier and Kurtz [20].

(ii) Building on Doss [18], Da Prato and Frankowska [15] give necessary and sufficient conditions for the stochastic invariance in terms of the Stratonovich drift and the first order normal cone:

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

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

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

In practice, the first order normal cone is much simpler to compute than the second order cone used in [5], see [3] and [36]. However, the price to pay is to impose a strong regularity condition on the diffusion matrix σ, 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 σ. We make the following seemingly trivial observation: $C := \sigma\sigma^\top$ might be differentiable at a point x while σ 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 σ.

We prove that this is actually possible. Our main result, Theorem 2.3 below, states that the stochastic invariance is equivalent to the following conditions:

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

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

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

in which $\mathcal{C}^{1,1}_{loc}$ 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 σ 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 $\mathcal{N}^1_{\mathcal{D}}(0) = \mathbb{R}_-$. 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 $(CC^+)(0) = 0$.

This extension of the characterization of Da Prato and Frankowska [15] provides for the first time a unified criteria for the case where the volatility matrix may not be $\mathcal{C}^1$ on the whole domain, which is of importance in practical situations. 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 [19, 22]), or of polynomial diffusions that are characterized by a quadratic covariance matrix (see [14, 21]), 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, on a case by case basis. For instance, in their construction of affine processes on the cone of symmetric semi-definite matrices, Cuchiero *et al.* [13] start by regularizing the martingale problem before applying the stochastic invariance characterization of [15] and then pass to the limit. In Spreij and Veerman [37], 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 diffusions is given in Filipović and Larsson [21]. 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 generic example.

Our proof of the necessary condition is in the spirit of Buckdahn *et al.* [9]. 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 σ is not differentiable and $\sigma(X)$ fails to be a semi-martingale whenever an eigenvalue vanishes (see [33, Example 1.2]). We therefore need to develop new ideas. We first observe that, if σ is diagonal, then 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). If σ is not diagonal, we can essentially 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 [20, 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 [5, 9] cannot be applied to our case since σ 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 explicit stochastic invariance conditions for various typical examples of applications. 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 $x = (x^1, \ldots, x^d) \in \mathbb{R}^d$, 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 [15] to weaker regularity assumptions.

Since we are dealing with general coefficients b and σ, 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 σ: 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. [25, 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

$$\mathbb{E}\left[\|X_t - X_s\|^p\right] \leq K_{p,x}|t-s|^{\frac{p}{2}} \tag{2.1}$$

for all $0 \leq s, t \leq 1$. Hence, Kolmogorov's continuity criterion ensures that the sample paths of X

are (locally) η-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 ϕ by $\mathcal{L}\phi := D\phi\, b + \frac{1}{2}\mathrm{Tr}[\sigma\sigma^\top D^2\phi]$. Therefore, $\mathcal{Q}$ is the same if σ is replaced by $\tilde{\sigma}$ such that $\tilde{\sigma}\tilde{\sigma}^\top = \sigma\sigma^\top$, see e.g. [39, Remark 5.1.7]. Hence, we can reduce to the case where σ 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_2).

Theorem 2.3 (Invariance characterization)**.** *Let $\mathcal{D}$ be closed. Assume that b, σ and C are continuous and satisfy assumptions (H_1)-(H_2). Then, the set $\mathcal{D}$ is stochastically invariant with respect to the diffusion (1.1) if and only if*

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

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

Clearly, the regularity conditions of Theorem 2.3 are much weaker than those of Theorem 4.1 in Da Prato and Frankowska [15]. 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 $C(x)C(x)^+ = \mathbb{1}_{\{x>0\}}$ and $\mathcal{N}^1_{\mathbb{R}_+}(x) = \mathbb{1}_{\{x=0\}}\mathbb{R}_-$, 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 [15, Theorem 4.1] under their smoothness assumptions. This is the object of the next proposition (recall that, by Remark 2.1, the study can be reduced to the case $C = \sigma^2$ on $\mathcal{D}$).

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

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

Proof. Fix $x \in \mathcal{D}$ and $u \in \operatorname{Ker}\sigma(x)$. 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 $\mathcal{C}^{1,1}_{loc}$. It then follows from Proposition A.5 and the fact that $u \in \operatorname{Ker}\sigma(x)$ 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 σ 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,

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

in which the last identity follows by reproducing exactly the same computations in the reverse order with C in place of σ. □

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 $C(x) = Q(x)\mathrm{diag}\left[\lambda_1(x), \ldots, \lambda_r(x), 0, \ldots, 0\right] Q(x)^\top$, where $Q(x)Q(x)^\top = I_d$ and $\lambda_j(x) > 0$ for all $1 \le j \le r$. Hence, the r-first columns of $Q(x)$, denoted by $(q_1, \ldots, q_r) = (q_1(x), \ldots, q_r(x))$, span the image of $C(x)$ and the projection matrix on the image of $C(x)$ is given by $C(x)C(x)^+ = \sum_{j=1}^{r} q_j q_j^\top$, see Propositions A.3 and A.2 in the Appendix and recall that q_j is a column vector. Thus, by (2.3) in the proof of Proposition 2.4 and Proposition A.5 in the Appendix,*

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

so that, by Proposition A.8,

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

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

$$D_{q_j}(Cq_j)(x) := \lim_{t\to 0} \frac{C(x+tq_j)q_j - C(x)q_j}{t}.$$

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 $(q_1, ..., q_r)$. 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 an additional comment for the jump-diffusion case.

Remark 2.6 (Adding jumps)**.** *Note that jumps could be included in the dynamics of X. Based on the current work, we provide in [1] an extension of the first order characterization of Theorem 2.3 to the jump-diffusion case. We also derive an equivalent formulation in the semimartingale framework.*

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 [9]. We fix $x \in \mathcal{D}$ and 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, let X be a $\mathcal{D}$-valued solution starting from $X_0 = x$. In particular, $\phi(X_t) \leq \phi(x)$, for all $t \geq 0$. Then, if σ is sufficiently smooth, 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. [9, 11], 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 ϕ, i.e. such that $D\phi(x) = u^\top$, to deduce that (2.2a)-(2.2b) must hold when σ is differentiable, recall Proposition 2.4.

In our setting, one can however not differentiate σ^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 Λ 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 Λ 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 Λ has only (strictly) positive entries.
Note that eigenvalues and the eigenvectors can always be chosen measurable. However, multiple eigenvalues and their corresponding eigenvectors can fail to have the same regularity as C. To ensure a sufficient regularity, we therefore assume in the following Lemma that non-zero eigenvalues are distinct. The general case will be treated later, thanks to a change of variable argument, see Section 3.2 below.

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

$$C(x) = Q(x)\mathrm{diag}\left[\lambda_1(x), \ldots, \lambda_r(x), 0, \ldots, 0\right] Q(x)^\top \tag{3.1}$$

with $\lambda_1(x) > \lambda_2(x) > \cdots > \lambda_r(x) > 0$ 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 $y \mapsto \Lambda(y) := \mathrm{diag}\left[\lambda_1(y), \ldots, \lambda_d(y)\right]$ 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) $\lambda_1(y) > \lambda_2(y) > ... > \lambda_r(y) > \max\{\lambda_i(x), r+1 \leq i \leq d\} \vee 0$, *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* $\bar\Lambda = \mathrm{diag}[\lambda_1, ..., \lambda_r, 0, ..., 0]$.

Moreover, we have:

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

Proof. Note that the fact that $(q_i)_{i\leq d}$ can be chosen measurable is guaranteed when (C, Λ) is measurable by the fact that each eigenvector solves a quadratic minimization problem, see e.g. [6, Proposition 7.33(p.153)]. Moreover, the continuity of the eigenvalues follows from Weyl's perturbation theorem, [7, Corollary III.2.6], and the smoothness of $(\bar\Lambda, \bar Q)$ is a consequence of [29, Theorem 1] since all the positive eigenvalues are simple and C is $\mathcal{C}^{1,1}_{loc}(\mathbb{R}^d, \mathbb{S}^d)$.
Let us now observe that any $u \in \mathrm{Ker}(C(x))$ 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:

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

Combining the above leads to

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

which proves (3.2) by similar computations as in the proof of (2.3). □

We can now adapt the arguments of [9]. In the following we use the notion of proximal normals. A vector $u \in \mathbb{R}^d$ is said to be a proximal normal to $\mathcal{D}$ at a point x if $\|u\| = d_{\mathcal{D}}(x+u)$, where $d_{\mathcal{D}}$ is the distance function to $\mathcal{D}$. We denote by $\mathcal{N}^{1,prox}_{\mathcal{D}}(x)$ the cone spanned by all proximal normals. Note however that (2.2a)-(2.2b) holds at x for all proximal normals $u \in \mathcal{N}^{1,prox}_{\mathcal{D}}(x)$ if and only if it holds for all $u \in \mathcal{N}^1_{\mathcal{D}}(x)$. Indeed,

$$\mathcal{N}^{1,prox}_{\mathcal{D}}(x) \subset \mathcal{N}^1_{\mathcal{D}}(x) \subset \bar{\text{co}}\left(\limsup_{\mathcal{D}\ni y \to x} \mathcal{N}^{1,prox}_{\mathcal{D}}(y)\right), \tag{3.3}$$

where lim sup stands for the Painlevé-Kuratowski upper limit (see e.g. [3, 15]) and $\bar{\text{co}}$ is the closed convex hull (see also [15, Remark 4.2 (a)]).

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 $u \in \mathcal{N}^1_{\mathcal{D}}(x)$.*

Proof. It follows from the discussion before our lemma that it suffices to prove our claim for $u \in \mathcal{N}^{1,prox}_{\mathcal{D}}(x)$. 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 $\mathcal{N}^{1,prox}_{\mathcal{D}}(x) = \{0\}$ and there is nothing to prove. We therefore assume from now on that $x \in \partial\mathcal{D}$. We fix $u \in \mathcal{N}^{1,prox}_{\mathcal{D}}(x)$.
Step 1. We first claim that there exists a function $\phi \in \mathcal{C}^\infty_b(\mathbb{R}^d, \mathbb{R})$ 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 [36, 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 ρ is a $\mathcal{C}^\infty_b$ 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 $\bar{B} = \Lambda(x)\Lambda(x)^+ B = (B^1, .., B^r, 0, ..., 0)^\top$ and $\bar{B}^\perp = (I_d - \Lambda(x)\Lambda(x)^+)B = (0, ..., 0, B^{r+1}, ..., B^d)$, recall Proposition A.2. Since $Q\bar{\Lambda}^{\frac{1}{2}} = \bar{Q}\bar{\Lambda}^{\frac{1}{2}}$, the above inequality can be written in the form

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

Let $(\mathcal{F}^{\bar{B}}_s)_{s\geq 0}$ be the completed filtration generated by $\bar{B}$. By [28, Corollaries 2 and 3 of Theorem 5.13], [26, Lemma 14.2], and the fact that the martingale $\bar{B}^\perp$ is independent of $\bar{B}$, we obtain

$$\begin{aligned}
0 &\geq \int_0^t \mathbb{E}_{\mathcal{F}^{\bar{B}}_s}[\mathcal{L}\phi(X_s)]ds + \int_0^t \mathbb{E}_{\mathcal{F}^{\bar{B}}_s}[D\phi(X_s)\bar{\sigma}(X_s)]d\bar{B}_s \\
&= \int_0^t \mathbb{E}_{\mathcal{F}^{\bar{B}}_s}[\mathcal{L}\phi(X_s)]ds + \int_0^t \mathbb{E}_{\mathcal{F}^{\bar{B}}_s}[D\phi(X_s)\bar{\sigma}(X_s)]dB_s,
\end{aligned}$$

where the last equality holds because the $(d-r)$ columns of $\bar\sigma$ are 0. We now apply Lemma 3.3 below to $(D\phi\bar\sigma)(X)$ and use [28, Corollaries 2 and 3 of Theorem 5.13] and [26, Lemma 14.2] again to find a bounded adapted process η such that

$$0 \geq \int_0^t \theta_s ds + \int_0^t \left(\alpha + \int_0^s \beta_r dr + \int_0^s \gamma_r dB_r\right)^\top dB_s \tag{3.4}$$

where

$$\begin{aligned} \theta &:= \mathbb{E}_{\mathcal{F}^{\bar B}_\cdot}\left[\mathcal{L}\phi(X_\cdot)\right] \quad , \quad \alpha^\top := (D\phi\bar\sigma)(x) = u^\top Q(x)\Lambda(x)^{\frac12} \\ \beta &:= \mathbb{E}_{\mathcal{F}^{\bar B}_\cdot}\left[D(D\phi\bar\sigma)(X_\cdot)b(X_\cdot)+\eta_\cdot\right] \quad , \quad \gamma := \mathbb{E}_{\mathcal{F}^{\bar B}_\cdot}\left[D(D\phi\bar\sigma)\bar\sigma(X_\cdot)\right], \end{aligned}$$

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 ϕ has compact support. In addition, given $T>0$, the independence of the increments of $\bar B$ implies that $\theta_s = \mathbb{E}_{\mathcal{F}^{\bar B}_T}\left[\mathcal{L}\phi(X_s)\right]$ for all $s \leq T$. It follows that θ is a.s. continuous at 0.

Similarly, $\gamma = \mathbb{E}_{\mathcal{F}^{\bar B}_T}\left[D(D\phi\bar\sigma)\bar\sigma(X_\cdot)\right]$ 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, γ has ϵ-Hölder sample paths for all $0<\epsilon<\frac14$, in particular $\int_0^t \|\gamma_s-\gamma_0\|^2 ds = O(t^{1+\epsilon})$ for $0<\epsilon<\frac12$.

Step 4. In view of Step 3, we can apply Lemma 3.4 to (3.4) to deduce that $\alpha = 0$ and $\theta_0 - \frac12 \operatorname{Tr}(\gamma_0) \leq 0$. Multiplying the first equation by $\Lambda(x)^{\frac12}Q^\top(x)$ implies that $0 = \alpha^\top\Lambda(x)^{\frac12}Q^\top(x) = u^\top Q(x)\Lambda(x)^{\frac12}\Lambda(x)^{\frac12}Q^\top(x) = u^\top C(x)$, 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

$$0 \geq \mathcal{L}\phi(x) - \frac12 \operatorname{Tr}\left[\bar\sigma^\top D^2\phi\bar\sigma + (I_d\otimes u^\top)D\bar\sigma\bar\sigma\right](x) = u^\top b(x) - \frac12 \operatorname{Tr}\left[(I_d\otimes u^\top)D\bar\sigma\bar\sigma\right](x),$$

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

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 η is only locally bounded).

Lemma 3.3. *Assume that b and σ 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 η such that*

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

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) $\|f_n - f\| + \|Df_n - Df\| \le \frac{K}{n}$,

for all $n \ge 1$. 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 $\eta^n := \frac{1}{2}\mathrm{Tr}[D^2 f_n \sigma\sigma^\top](X)$. 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 [17, Theorem 1.3], there exists $(\widetilde{\eta}^n) \in \mathrm{Conv}(\eta^k, k \ge n)$ such that $\widetilde{\eta}^n \to \eta\ dt \otimes d\mathbb{P}$ almost surely. Let $N_n \ge 0$ and $(\lambda_k^n)_{n \le k \le N_n} \subset [0,1]$ be such that $\widetilde{\eta}^n = \sum_{k=n}^{N_n} \lambda_k^n \eta^k$ and $\sum_{k=n}^{N_n} \lambda_k^n = 1$. Set $\widetilde{f}_n := \sum_{k=n}^{N_n} \lambda_k^n f_k$. Then,

$$\widetilde{f}_n(X_t) = \widetilde{f}_n(x) + \int_0^t D\widetilde{f}_n(X_s)b(X_s)ds + \int_0^t \widetilde{\eta}_s^n ds + \int_0^t D\widetilde{f}_n(X_s)\sigma(X_s)dW_s. \tag{3.5}$$

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

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

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

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

in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ as $n \to \infty$, and therefore a.s. after possibly considering a subsequence. It thus remains to send $n \to \infty$ in (3.5) to obtain the required result. □

The following adapts [9, Lemma 2.1] to our setting, see also [8, 11, 12].

Lemma 3.4. *Let $(W_t)_{t\ge0}$ denote a standard d-dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge0}, \mathbb{P})$. Let $\alpha \in \mathbb{R}^d$ and $(\beta_t)_{t\ge0}$, $(\gamma_t)_{t\ge0}$ and $(\theta_t)_{t\ge0}$ be adapted processes taking values respectively in $\mathbb{R}^d$, $\mathbb{M}^d$ and $\mathbb{R}$ and satisfying*

(1) *β is bounded,*

(2) *$\int_0^t \|\gamma_s\|^2 ds < \infty$, for all $t \ge 0$,*

(3) *there exists a random variable $\eta > 0$ such that a.s.*

$$\int_0^t \|\gamma_s - \gamma_0\|^2 ds = O(t^{1+\eta}) \quad \text{for } t \to 0, \tag{3.6}$$

(4) *θ is a.s. continuous at 0.*

Suppose that for all $t \ge 0$

$$\int_0^t \theta_s ds + \int_0^t \left(\alpha + \int_0^s \beta_r dr + \int_0^s \gamma_r dW_r\right)^\top dW_s \le 0. \tag{3.7}$$

Then,

(a) $\alpha = 0$,

(b) $-\gamma_0 \in \mathbb{S}^d_+$,

(c) $\theta_0 - \frac{1}{2}\operatorname{Tr}(\gamma_0) \le 0$.

Proof. Since $(W^i_t)^2 = 2\int_0^t W^i_s dW^i_s + t$, (3.7) reduces to

$$(\theta_0 - \frac{1}{2}\operatorname{Tr}(\gamma_0))t + \sum_{i=1}^d \alpha^i W^i_t + \sum_{i=1}^d \frac{\gamma_0^{ii}}{2}(W^i_t)^2 + \sum_{1\le i\neq j\le d} \gamma_0^{ij}\int_0^t W^i_s dW^j_s + R_t \le 0,$$

where

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

In view of [9, Lemma 2.1], it suffices to show that $R_t/t \to 0$ almost surely. To see this, first note that $R^1_t = o(t)$ a.s. since θ is continuous at 0. Moreover, [11, Proposition 3.9] implies that $R^2_t = o(t)$ a.s., as β is bounded.
It remains to prove that $R^3_t = o(t)$ a.s. To see this, define $M^{ij} = \gamma^{ij} - \gamma^{ij}_0$ and $M^i = \int_0^\cdot \sum_{j=1}^d M^{ij}_r dW^j_r$ for all $1 \le i,j \le d$. The continuity assumption (3.6) implies that $\langle M^i\rangle_s = O(s^{1+\eta})$ almost surely. By the Dambis-Dubins-Schwarz theorem, $(M^i_s)_{s\ge 0}$ is therefore a time-changed Brownian motion, see e.g. [35, Theorem V.1.6]. By the law of iterated logarithm for Brownian motion $(M^i_s)^2 = O(s^{1+\frac{\eta}{2}})$ almost surely. Hence, $\langle R^3\rangle_t = O(t^{2+\frac{\eta}{2}})$ almost surely. By applying the Dambis-Dubin-Schwarz theorem and the law of iterated logarithm again, we obtain that $R^3_t = o(t)$ a.s. □

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 $u \in \mathcal{N}^1_{\mathcal{D}}(x)$.*

Proof. If x lies in the interior of $\mathcal{D}$, then $\mathcal{N}^1_{\mathcal{D}}(x) = \{0\}$ and there is nothing to prove. We therefore assume from now on that $x \in \partial\mathcal{D}$. Let Λ and Q be defined through the spectral decomposition of C, as in (3.1) but with only $\lambda_1(x) \ge \cdots \ge \lambda_d(x)$. 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

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

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

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

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_ϵ are all distinct at $x^\epsilon := A^\epsilon x$, as $C_\epsilon(x^\varepsilon) = Q(x)\mathrm{diag}\left[(1-\epsilon)\lambda_1(x), \dots, (1-\epsilon)^d \lambda_d(x)\right] Q(x)^\top$. We can therefore apply Lemma 3.2 to $(X^\epsilon, \mathcal{D}^\epsilon)$:

$$\begin{cases} C_\epsilon(x^\epsilon)u_\epsilon = 0 & \text{(3.8a)} \\ \langle u_\epsilon, b_\epsilon(x^\epsilon) - \frac{1}{2}\sum_{j=1}^{d} DC_\epsilon^j(x^\epsilon)(C_\epsilon C_\epsilon^+)^j(x^\epsilon)\rangle \leq 0 & \text{(3.8b)} \end{cases}$$

for all $u_\epsilon \in \mathcal{N}^1_{A^\epsilon \mathcal{D}}(x^\epsilon)$. We now easily verify that $\mathcal{N}^1_{A^\epsilon\mathcal{D}}(x^\epsilon) = (A^\epsilon)^{-1}\mathcal{N}^1_{\mathcal{D}}(x)$, recall the definition in (1.4). Finally, by sending $\epsilon \to 0$ in (3.8a) and (3.8b), 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 $u \in \mathcal{N}^1_{\mathcal{D}}(x)$, which ends the proof. ∎

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. [20, p165]. Then, classical arguments, mainly [20, 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 [15, 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 $u \in \mathcal{N}^1_{\mathcal{D}}(x)$. 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:

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

where $\sigma(x)^+$ is the pseudoinverse of $\sigma(x)$. Since C is locally Lipschitz and verifies condition (2.2a), [15, Proposition 2.5] combined with (3.3) implies 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 $(u, v) \in \mathcal{N}^2_{\mathcal{D}}(x)$. On the other hand, since C is $\mathcal{C}^{1,1}_{loc}$, 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 σ is symmetric. Now observe that $u \in \mathcal{N}^1_{\mathcal{D}}(x)$ whenever $(u, v) \in \mathcal{N}^2_{\mathcal{D}}(x)$. 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 $h \to 0$ before summing over $1 \le j \le d$ yields

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

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

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

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

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 $u \in \mathcal{N}^1_{\mathcal{D}}(x)$. 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, [20, 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 [10, Proposition 3.2] and [20, 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, [20, Theorem 5.3.3] implies the existence of a weak solution (X, W) such that $X_t \in \mathcal{D}$ for all $t \ge 0$ almost surely.

□

5 A generic application

We show in this section how Theorem 2.3 can be applied in various examples of application. We restrict to a two-dimensional setting for ease of computations and notations.

We first provide a generic tractable characterization for the stochastic invariance of all state spaces $\mathcal{D} \subset \mathbb{R}^2$ of the following form:

$$\mathcal{D} = \{(\bar{x}, \tilde{x}) \in \mathbb{R}^2, \bar{x} \in \mathcal{D}_1 \text{ and } \phi(\bar{x}, \tilde{x}) \in \mathcal{D}_2\}, \tag{5.1}$$

where $\mathcal{D}_1 \subset \mathbb{R}$ and $\mathcal{D}_2 \subset \mathbb{R}$ are closed subsets and ϕ is a continuously differentiable function. Then, $\mathcal{D}$ can be characterized through $\Phi : (\bar{x}, \tilde{x}) \mapsto (\bar{x}, \phi(\bar{x}, \tilde{x}))$ by

$$\mathcal{D} = \Phi^{-1}(\mathcal{D}_1 \times \mathcal{D}_2),$$

and [36, Exercise 6.7 and Proposition 6.41] provides the following description of the normal cone whenever

$$\Phi \text{ is differentiable at } x \text{ and its Jacobian } D\Phi(x) \text{ has full rank} \tag{H_x}$$

holds at any point $x \in \mathcal{D}$.

Proposition 5.1. *Fix $x = (\bar{x}, \tilde{x}) \in \mathcal{D}$ such that (H_x) holds. Then,*

$$\mathcal{N}^1_{\mathcal{D}}(x) = \left\{ \begin{pmatrix} \bar{u} + \partial_1\phi(x)\tilde{u} \\ \partial_2\phi(x)\tilde{u} \end{pmatrix}, \bar{u} \in \mathcal{N}^1_{\mathcal{D}_1}(\bar{x}) \text{ and } \tilde{u} \in \mathcal{N}^1_{\mathcal{D}_2}(\phi(\bar{x}, \tilde{x})) \right\},$$

in which $\partial_i\phi$ is the derivative with respect to the i-th component.

When x lies in the interior of $\mathcal{D}$, $\mathcal{N}^1_{\mathcal{D}}(x) = \{0\}$ and (2.2a)-(2.2b) are trivially verified. Hence, it suffices to control b and C on the boundary of the domain in order to ensure the stochastic invariance of $\mathcal{D}$ as stated by the following proposition, in which we use the notations

$$b = (\bar{b}, \tilde{b})^\top, \ C = (C_{ij})_{ij} \ \text{ and } \ \partial_u = u_2\partial_1 - u_1\partial_2. \tag{5.2}$$

Proposition 5.2. *Let $\mathcal{D}$ be as in (5.1) and $x = (\bar{x}, \tilde{x}) \in \partial\mathcal{D}$ be such that (H_x) holds. Fix $u = (u_1, u_2)^\top \in \mathcal{N}^1_{\mathcal{D}}(x)$ as in Proposition 5.1. Under the assumptions of Theorem 2.3, (2.2a)-(2.2b) are equivalent to the following:*

(a) Either $\tilde{u} \neq 0$ and

$$\begin{cases} C(x) = C_{11}(x)\begin{pmatrix} 1 & -\frac{u_1}{u_2} \\ -\frac{u_1}{u_2} & \frac{u_1^2}{u_2^2}\end{pmatrix}, & (5.3a)\\ \langle u, b(x)\rangle - \dfrac{\mathbb{1}_{\{C_{11}(x)\neq 0\}}}{2(u_1^2+u_2^2)}\left[u_1u_2\partial_u(C_{11}-C_{22})(x) + (u_2^2-u_1^2)\partial_u C_{12}(x)\right] \le 0. & (5.3b)\end{cases}$$

(b) Or, $\tilde{u} = 0$, $u_1 = \bar{u}$ and

$$\begin{cases} C(x)\mathbb{1}_{\{\bar{u}\neq 0\}} = C_{22}(x)\begin{pmatrix} 0 & 0 \\ 0 & 1\end{pmatrix}\mathbb{1}_{\{\bar{u}\neq 0\}}, & (5.4a)\\ \bar{u}\left(\bar{b}(x) - \dfrac{\mathbb{1}_{\{C_{22}(x)\neq 0\}}}{2}\partial_2 C_{12}(x)\right) \le 0. & (5.4b)\end{cases}$$

Proof. Case (a), $\tilde{u} \neq 0$: Since $D\Phi(x)$ has full rank, $\partial_2\phi(x) \neq 0$ and therefore $u_2 \neq 0$. Since $C(x) \in \mathbb{S}^2$, (2.2a) is clearly equivalent to (5.3a).
If $C_{11}(x) \neq 0$, (5.3a) implies that $u = (\bar{u} + \partial_1\phi(x)\tilde{u}, \partial_2\phi(x)\tilde{u})^\top$ spans the kernel of $C(x)$. Therefore, by Proposition A.3,

$$C(x)C(x)^+ = I_2 - \frac{1}{\|u\|^2}uu^\top = \frac{1}{u_1^2+u_2^2}\begin{pmatrix} u_2^2 & -u_1u_2 \\ -u_1u_2 & u_1^2\end{pmatrix}.$$

Straightforward computations yield

$$\langle u, \sum_{j=1}^2 DC^j(x)(CC^+)^j(x)\rangle = \frac{1}{u_1^2+u_2^2}\left[u_1u_2\partial_u(C_{11}-C_{22})(x) + (u_2^2-u_1^2)\partial_u C_{12}(x)\right],$$

recall the notations introduced in (5.2). This shows the equivalence between (2.2b) and (5.3b) when $C_{11}(x) \neq 0$.
If $C_{11}(x) = 0$, then (5.3a) implies that $C(x)C(x)^+ = 0$ and (2.2b) reads $\langle u, b(x)\rangle \le 0$.

Case (b), $\tilde{u} = 0$: If $\bar{u} = 0$, then $u = 0$ and there is nothing to prove. Otherwise, $u_1 = \bar{u} \neq 0$. Since $C(x) \in \mathbb{S}^2$, (2.2a) is clearly equivalent to $C_{11}(x) = 0$ and $C_{21}(x) = C_{12}(x) = 0$, that is (5.4a). If $C_{22}(x) \neq 0$, then (5.4a) provides

$$C(x)C(x)^+ = \begin{pmatrix} 0 & 0 \\ 0 & 1\end{pmatrix},$$

and straightforward computations yield

$$\langle u, \sum_{j=1}^2 DC^j(x)(CC^+)^j(x)\rangle = \bar{u}\partial_2 C_{12}(x),$$

which shows the equivalence between (2.2b) and (5.4b) when $C_{22}(x) \neq 0$. If $C_{22}(x) = 0$, then $C(x)C(x)^+ = 0$ and (2.2b) reads $\bar{u}\bar{b}(x) \le 0$. □

Note that $\bar{u} = 0$ when $\mathcal{D}_1 = \mathbb{R}$, which will be the case from now on. In the sequel, we impose more structure on the coefficients, as it is usually done in the construction of invariant diffusions. This permits to deduce an explicit form of (b, C) on the whole domain from the boundary conditions (5.3a)-(5.3b). As already stated, Theorem 2.3 can be directly applied to a large class of diffusions, e.g. affine diffusions [19, 22, 37] and polynomial diffusions [21, 27], not only for closed subsets of $\mathbb{R}^d$, but even when $\mathcal{D} \subset \mathbb{S}^d$ (as in [13]) since $\mathbb{S}^d$ can be identified with $\mathbb{R}^{\frac{d(d+1)}{2}}$ by using the half-vectorization operator. We start by defining these two main structures.

Definition 5.3 (Affine and polynomial diffusions). *X is a polynomial diffusion on $\mathcal{D}$ if:*

(i) There exist $\bar{b}^i, \widetilde{b}^i \in \mathbb{R}$, $0 \leq i \leq 2$, and $A^i \in \mathbb{S}^2$, $1 \leq i \leq 5$, such that $b : x \mapsto b(x) := (\bar{b}(x), \widetilde{b}(x)) \in \mathbb{R}^2$ and $C : x \mapsto C(x) \in \mathbb{S}^2$ have the following form:

$$\begin{cases} \bar{b}(x) & = \bar{b}^0 + \bar{b}^1 \bar{x} + \bar{b}^2 \widetilde{x}, \\ \widetilde{b}(x) & = \widetilde{b}^0 + \widetilde{b}^1 \bar{x} + \widetilde{b}^2 \widetilde{x}, \\ C(x) & = A^0 + A^1 \bar{x} + A^2 \widetilde{x} + A^3 \bar{x}^2 + A^4 \bar{x}\widetilde{x} + A^5 \widetilde{x}^2, \end{cases} \tag{5.5}$$

for all $x = (\bar{x}, \widetilde{x}) \in \mathcal{D}$.

(ii) $C(x) \in \mathbb{S}^d_+$, for all $x \in \mathcal{D}$.

When $A^i = 0$ for all $3 \leq i \leq 5$, we say that X is an affine diffusion.

Then, it is clear that b and C are $\mathcal{C}^\infty$ and satisfy the linear growth conditions (H_1).

In what follows, we highlight the interplay between the geometry/curvature of the boundary and the coefficients b and C. The three explicit examples below characterize the invariance for flat, convex and concave boundaries.

Example 5.4 (Canonical state space). *Fix $\mathcal{D}_1 = \mathbb{R}$, $\mathcal{D}_2 = \mathbb{R}_+$ and $\phi(\bar{x}, \widetilde{x}) = \widetilde{x}$. Then $\mathcal{D} = \mathbb{R} \times \mathbb{R}_+$ and $\mathcal{N}^1_{\mathcal{D}}(x) = \{0\} \times \mathbb{1}_{\{\widetilde{x}=0\}} \mathbb{R}_-$. Hence, (5.3a)-(5.3b) are equivalent to*

$$C(\bar{x}, 0) = C_{11}(\bar{x}, 0) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \textit{and} \quad \widetilde{b}(\bar{x}, 0) - \frac{\mathbb{1}_{\{C_{11}(\bar{x},0) \neq 0\}}}{2} \partial_1 C_{12}(\bar{x}, 0) \geq 0, \quad \textit{for all } \bar{x} \in \mathbb{R}.$$

If we now impose the structural condition (5.5), then straightforward computations lead to the characterization in [19] for affine diffusions. The case of polynomial diffusions can be treated similarly.

Example 5.5 (Parabolic convex state space). *Let us consider the following parabolic state space:*

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

Then, with the previous notations, $\mathcal{D}_1 = \mathbb{R}$, $\mathcal{D}_2 = \mathbb{R}_+$ and $\phi(\bar{x}, \widetilde{x}) = \widetilde{x} - \bar{x}^2$. Therefore, the first order normal cone given by Proposition 5.1 reads

$$\mathcal{N}^1_{\mathcal{D}}(x) = \begin{pmatrix} -2\bar{x} \\ 1 \end{pmatrix} \mathbb{R}_-, \quad \textit{for all } x = (\bar{x}, \bar{x}^2) \in \partial\mathcal{D}.$$

Conditions (5.3a)-(5.3b) are therefore equivalent to

$$\begin{cases} C(x) = C_{11}(x) \begin{pmatrix} 1 & 2\bar{x} \\ 2\bar{x} & 4\bar{x}^2 \end{pmatrix}, & (5.6a) \\ \langle u, b(x) \rangle - \dfrac{\mathbb{1}_{\{C_{11}(x) \neq 0\}}}{2(1+4\bar{x}^2)} \left[-2\bar{x}\partial_u (C_{11} - C_{22})(x) + (1 - 4\bar{x}^2)\partial_u C_{12}(x) \right] \geq 0, & (5.6b) \end{cases}$$

for all $\bar{x} \in \mathbb{R}$, $x = (\bar{x}, \bar{x}^2)$ *and* $u = (-2\bar{x}, 1)^\top$.
If we now impose an additional affine structure on the diffusion $X = (\bar{X}, \widetilde{X})$, *as in Duffie et al. [19, Section 12.2], we recover the characterization given in Gourieroux and Sufana [24, Proposition 2]. Indeed, Proposition 5.2 says that* $\mathcal{D}$ *is invariant if and only if there exists* $\alpha \geq 0$ *such that*

(a)

$$C(x) = \alpha \begin{pmatrix} 1 & 2\bar{x} \\ 2\bar{x} & 4\widetilde{x} \end{pmatrix}, \text{ for all } x = (\bar{x}, \widetilde{x}) \in \mathcal{D}, \tag{5.7}$$

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

$$\begin{cases} \widetilde{b}^2 > 2\bar{b}^1 \quad \text{and} \quad (\widetilde{b}^1 - 2\bar{b}^0)^2 \leq 4(\widetilde{b}^2 - 2\bar{b}^1)(\widetilde{b}^0 - \alpha) \\ \text{or} \\ \widetilde{b}^2 = 2\bar{b}^1, \quad \widetilde{b}^1 = 2\bar{b}^0 \quad \text{and} \quad \widetilde{b}^0 \geq \alpha. \end{cases} \tag{5.8}$$

Let us detail the computations: (a) The covariance matrix $C(x) \in \mathbb{S}^2_+$ *is of the form (5.6a) on the boundary. Since* C *is affine in* $(\bar{x}, \bar{x}^2)$, *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* α *such that* $C(x)$ *has the form (5.7) at* $x = (\bar{x}, \bar{x}^2)$, *in which* $\alpha \geq 0$ *to ensure that* $C(0) \in \mathbb{S}^2_+$. *Finally,* C *needs to have the same form (5.7) on the whole state space* $\mathcal{D}$, *since it is affine.*
(b) We now derive the form of the drift vector $b(x) = (\bar{b}(x), \widetilde{b}(x)) \in \mathbb{R}^2$ *by using (5.6b). From (5.7), elementary computations show that condition (5.6b) is equivalent to*

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

which is equivalent to (5.8), when $\alpha > 0$. *If* $\alpha = 0$, *the same conclusion holds.*
Conversely, (5.7) clearly implies (2.2a) and (ii) of Definition 5.3 since $\det(C(x)) = 4\alpha(\widetilde{x} - \bar{x}^2) \geq 0$ *and* $\widetilde{x} \geq 0$ *for all* $(\bar{x}, \widetilde{x}) \in \mathcal{D}$. *Moreover, (5.8) leads to (5.3b) by the same computations as above.*

Example 5.6 (Parabolic concave state space)**.** *We now consider the epigraph of the concave function* $\bar{x} \mapsto -\bar{x}^2$,

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

It follows that $\mathcal{D}_1 = \mathbb{R}$, $\mathcal{D}_2 = \mathbb{R}_+$, $\phi(\bar{x}, \widetilde{x}) = \widetilde{x} + \bar{x}^2$ *and*

$$\mathcal{N}^1_{\mathcal{D}}(x) = \begin{pmatrix} 2\bar{x} \\ 1 \end{pmatrix} \mathbb{R}_-, \quad \text{for all } x = (\bar{x}, -\bar{x}^2) \in \partial\mathcal{D},$$

from Proposition 5.1. Hence, conditions (5.3a)-(5.3b) are now equivalent to

$$\begin{cases} C(x) = C_{11}(x) \begin{pmatrix} 1 & -2\bar{x} \\ -2\bar{x} & 4\bar{x}^2 \end{pmatrix}, & (5.9a) \\ \langle u, b(x) \rangle - \dfrac{\mathbb{1}_{\{C_{11}(x) \neq 0\}}}{2(4\bar{x}^2 + 1)} \left[2\bar{x}\partial_u (C_{11} - C_{22})(x) + (1 - 4\bar{x}^2)\partial_u C_{12}(x) \right] \geq 0, & (5.9b) \end{cases}$$

for all $\bar{x} \in \mathbb{R}$, $x = (\bar{x}, -\bar{x}^2)$ *and* $u = (2\bar{x}, 1)^\top \in -\mathcal{N}^1_{\mathcal{D}}(x)$.
Let us first note that the above shows that we can not construct an affine diffusion living in $\mathcal{D}$, *that is not degenerate, unless it lives on the boundary only. Indeed, if* C *is affine then* $C_{11} =: \alpha$ *has to be constant, because of (5.9a), and* C *is of the form (5.9a) with* $(-\widetilde{x})$ *in place of* $\bar{x}^2$.

Since $C(x) \in \mathbb{S}^2_+$, we must have $\alpha \geq 0$ and $\det C(x) = -4\alpha^2(\widetilde{x} + \bar{x}^2) \geq 0$. Thus, $\alpha = 0$ unless we restrict to points $(\bar{x}, \widetilde{x})$ on the boundary. If we do so, it is not difficult to derive a necessary and sufficient condition on the coefficients from the identity $\widetilde{X} = -\bar{X}^2$.

We now impose a polynomial structure on the diffusion $X = (\bar{X}, \widetilde{X})$, such that $\bar{X}$ is affine on its own, i.e. $\bar{b}$ and C_{11} are of affine form and only depend on $\bar{x}$. This extends [27, Example 5.2] and entirely characterizes the stochastic invariance of $\mathcal{D}$ with respect to this structure of diffusion. By Proposition 5.1, $\mathcal{D}$ is invariant if and only if there exist $\alpha, \beta \geq 0$, such that

(a)

$$C(x) = \begin{pmatrix} \alpha & -2\alpha\bar{x} \\ -2\alpha\bar{x} & (4\alpha+\beta)\bar{x}^2 + \beta\widetilde{x} \end{pmatrix}, \text{ for all } x = (\bar{x}, \widetilde{x}) \in \mathcal{D}, \tag{5.10}$$

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

$$\begin{cases} \widetilde{b}^2 < 2\bar{b}^1 \quad \text{and} \quad (\widetilde{b}^1 + 2\bar{b}^0)^2 \leq 4(-\widetilde{b}^2 + 2\bar{b}^1)(\widetilde{b}^0 + \alpha) \\ \text{or} \\ \widetilde{b}^2 = 2\bar{b}^1, \quad \widetilde{b}^1 = -2\bar{b}^0 \quad \text{and} \quad \widetilde{b}^0 \geq -\alpha \end{cases}. \tag{5.11}$$

Let us do the computations explicitly: (a) The covariance matrix $C(x) \in \mathbb{S}^2_+$ is of the form (5.9a) on the boundary. Therefore, $C_{11}(x) \geq 0$, for all $x \in \partial\mathcal{D}$. Since C_{11} is affine and only depends on $\bar{x} \in \mathbb{R}$, then necessarily C_{11} is a non-negative constant on the whole space $\mathcal{D}$. Therefore, there exists $\alpha \geq 0$ such that $C_{11}(.) = \alpha$ on $\mathcal{D}$. Moreover, (5.5) reads on the boundary

$$C(x) = A^0 + A^1\bar{x} + (A^3 - A^2)\bar{x}^2 - A^4\bar{x}^3 + A^5\bar{x}^4, \quad \text{for all } \bar{x} \in \mathbb{R}.$$

Therefore, comparing with (5.9a) leads to $A^4 = A^5 = 0$ and the existence of β, β' such that C is of the form

$$\begin{pmatrix} \alpha & -2\alpha\bar{x} \\ -2\alpha\bar{x} & 4\alpha\bar{x}^2 \end{pmatrix} + \begin{pmatrix} 0 & \beta' \\ \beta' & \beta \end{pmatrix}(\widetilde{x} + \bar{x}^2)$$

on the whole space $\mathcal{D}$. We now use the fact that $C(\mathcal{D}) \subset \mathbb{S}^2_+$. In particular, taking $\bar{x} = 0$ shows that we must have $\alpha\beta\widetilde{x} - (\beta')^2\widetilde{x}^2 \geq 0$ for all $\widetilde{x} \geq 0$, so that $\beta' = 0$. Similarly, $4\alpha\bar{x}^2 + \beta(\widetilde{x} + \bar{x}^2) \geq 0$ must hold for all $x \in \mathcal{D}$, which is equivalent to $\beta \geq 0$.
(b) We now derive the form of the drift vector $b(x) = (\bar{b}(x), \widetilde{b}(x)) \in \mathbb{R}^2$ by using (5.9b). Since X is affine on its own, $\bar{b}^2 = 0$. From (5.10), elementary computations show that condition (5.9b) is equivalent to

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

which is equivalent to (5.11), when $\alpha > 0$. If $\alpha = 0$, the same conclusion holds.
Conversely, (5.10)-(5.11) show that X is a polynomial diffusion such that $\bar{X}$ is affine on its own since $\det(C(x)) = \alpha\beta(\widetilde{x} + \bar{x}^2) \geq 0$ and $4\alpha\bar{x}^2 + \beta(\widetilde{x} + \bar{x}^2) \geq 4\alpha\bar{x}^2 \geq 0$ for all $(\bar{x}, \widetilde{x}) \in \mathcal{D}$. (5.10) clearly implies (2.2a). Moreover, (5.11) leads to (5.3b) by the same computations as above.

We conclude with a final remark on the interplay between the local geometry of the boundary, the coefficients b and C and the structure of the diffusion.

Remark 5.7. *(i) Curvaceous boundary and covariance matrix: the curvature of the boundary plays a crucial role in determining the covariance structure. In Example 5.4, the canonical state space, which shows no curvature, imposes strict constraints on the covariance matrix. Whereas, for curved domains, as in Examples 5.5-5.6, the first order normal cone is a more complicated object and induces a richer covariance structure on the boundary.*

(ii) Convexity and drift direction: Figure 1 visualizes the direction of the drift $b(0) = (\bar{b}^0, \widetilde{b}^0)$ *with respect to the convexity of the domain. When the domain is convex, as in Example 5.5, the drift is necessarily inward pointing since* $\widetilde{b}^0 \geq \alpha$*, with* $\alpha \geq 0$ *from (5.8). However, when the domain is concave, as in Example 5.6, the drift could even be outward pointing. This follows from the fact that* $\widetilde{b}^0 \geq -\alpha$*, with* $\alpha \geq 0$ *in (5.11).*

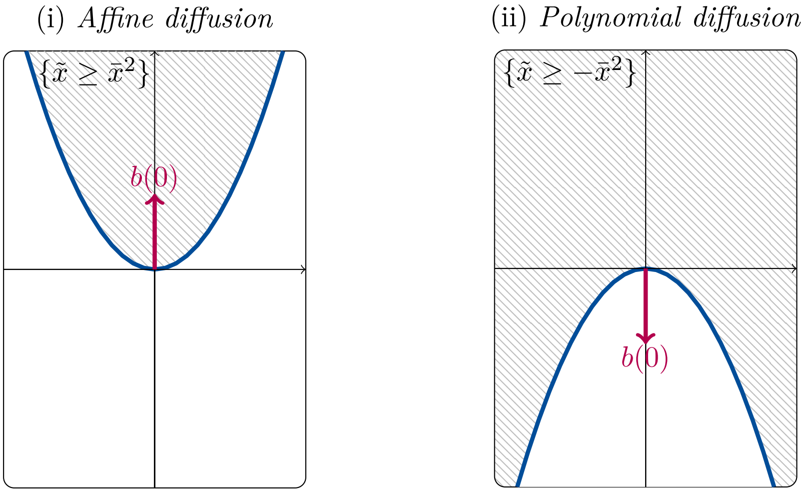

Figure 1: **Interplay between the convexity of the domain and the direction of the drift:** (i) Inward pointing drift for convex domains (Example 5.5). (ii) Possible outward pointing drift for concave domains (Example 5.6).

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 [37, Proposition 3.5] derived with the help of *McKean's argument* (see [32, Section 4]). Moreover, we extend the tractable conditions of [37, Proposition 3.7] given for affine diffusions. Our result could be easily used in the context of polynomial 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

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

We conclude by using [37, 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 [37, Proposition 3.5]). □

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 [37, Proposition 3.7].*

(iii) Jacobi diffusion: *Set* $\mathcal{D} = (0, 1]$ *and consider a polynomial 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 de 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 [23, Remark 1 page 131],* $\sigma = C^{\frac{1}{2}}$ *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 [30, 31, 34].

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* $\Lambda = \operatorname{diag}[(\lambda_i)_{i\leq d}] \in \mathbb{M}^d$*. Then,* $A^+ = Q\Lambda^+Q^\top$ *in which* $\Lambda^+ = \operatorname{diag}\left[(\lambda_i^{-1}\mathbb{1}_{\{\lambda_i\neq 0\}})_{i\leq d}\right]$*, and* $AA^+ = Q\operatorname{diag}\left[(\mathbb{1}_{\{\lambda_i\neq 0\}})_{i\leq d}\right]Q^\top$*. 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 [31, Chapter 9]).

Definition A.6 (Vectorization operator)**.** *Let $A \in \mathbb{M}^{m,n}$. The vectorization operator* 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:*

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

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] Abi Jaber, E. (2017). *Stochastic invariance of closed sets for jump-diffusions with non-Lipschitz coefficients.* Electronic Communications in Probability, vol. 22.

[2] 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.

[3] Aubin, J. P., & Frankowska, H. (1990). *Set-valued analysis.* Birkhäuser, Boston, Basel, Berlin.

[4] 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.

[5] Bardi, M., & Jensen, R. (2002). *A geometric characterization of viable sets for controlled degenerate diffusions.* Set-Valued Analysis, 10(2-3), 129-141.

[6] Bertsekas, D. P., & Shreve, S. E. (1978). *Stochastic optimal control: The discrete time case* (Vol. 23, pp. 15-24). New York: Academic Press.

[7] Bhatia, R. (1997). *Matrix analysis.* Springer-Verlag, New York, Volume 169 of Graduate Texts in Mathematics.

[8] Bruder, B. (2005). *Super-replication of European options with a derivative asset under constrained finite variation strategies.* Preprint HAL http://hal. archivesouvertes. fr/hal-00012183/fr.

[9] 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 mathematiques, 134(2), 207-214.

[10] Cheridito, P., Filipović, D., & Yor, M. (2005). *Equivalent and absolutely continuous measure changes for jump-diffusion processes.* Annals of applied probability, 1713-1732.

[11] Cheridito, P., Soner, H. M., & Touzi, N. (2005). *Small time path behavior of double stochastic integrals and applications to stochastic control.* The Annals of Applied Probability, 15(4), 2472-2495.

[12] Cheridito, P., Soner, H. M., & Touzi, N. (2005). *The multi-dimensional super-replication problem under gamma constraints.* In Annales de l'IHP Analyse non linéaire (Vol. 22, No. 5, pp. 633-666).

[13] Cuchiero, C., Filipović, D., Mayerhofer, E., & Teichmann, J. (2011). *Affine processes on positive semidefinite matrices.* The Annals of Applied Probability, 21(2), 397-463.

[14] Cuchiero, C., Keller-Ressel, M., & Teichmann, J. (2012). *Polynomial processes and their applications to mathematical finance.* Finance and Stochastics, 16(4), 711-740.

[15] Da Prato, G., & Frankowska, H. (2004). *Invariance of stochastic control systems with deterministic arguments.* Journal of differential equations, 200(1), 18-52.

[16] Da Prato, G., & Frankowska, H. (2007). *Stochastic viability of convex sets.* Journal of mathematical analysis and applications, 333(1), 151-163.

[17] 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.

[18] 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).

[19] Duffie, D., Filipović, D., & Schachermayer, W. (2003). *Affine processes and applications in finance.* Annals of applied probability, 984-1053.

[20] Ethier, S. N., & Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence.* Wiley, New York.

[21] Filipović, D., & Larsson, M. (2016). *Polynomial diffusions and applications in finance.* Finance and Stochastics, to appear.

[22] Filipović, D., & Mayerhofer, E. (2009). *Affine diffusion processes: theory and applications.* Advanced Financial Modelling, 8, 125.

[23] Friedman, A. (1976). *Stochastic differential equations and applications* (Vol. 2). Academic press.

[24] Gourieroux, C., & Sufana, R. (2006). *A Classification of Two-Factor Affine Diffusion Term Structure Models.* Journal of Financial Econometrics, 4(1), 31-52.

[25] Ikeda, N., & Watanabe, S. (1981). *Stochastic differential equations and diffusion processes.* North-Holland Publishing Company, Amsterdam.

[26] Kurtz, T. G. (2001). *Lectures on Stochastic Analysis.* Department of Mathematics and Statistics, University of Wisconsin, Madison, WI, 53706-1388.

[27] Larsson, M., and Pulido, S. (2015). *Polynomial diffusions on compact quadric sets.* Stochastic Processes and their Applications, to appear.

[28] Liptser, R., & Shiryaev, A. N. (2001). *Statistics of random Processes: I. general Theory* (Vol. 2). Springer Science & Business Media.

[29] Magnus, J. R. (1985). *On differentiating eigenvalues and eigenvectors.* Econometric Theory, 1(02), 179-191.

[30] Magnus, J. R., & Neudecker, H. (1980). *The elimination matrix: some lemmas and applications.* SIAM Journal on Algebraic Discrete Methods, 1(4), 422-449.

[31] 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.

[32] Mayerhofer, E., Pfaffel, O., & Stelzer, R. (2011). *On strong solutions for positive definite jump diffusions.* Stochastic processes and their applications, 121(9), 2072-2086.

[33] 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.

[34] Neudecker, H. (1986). *Symmetry, 0-1 matrices and jacobians.* Econometric Theory, 2, 157-190.

[35] Revuz, D., & Yor, M. (1999). *Continuous martingales and Brownian motion.* Third edition. Springer, Berlin.

[36] Rockafellar, R. T., & Wets, R. J. B. (1998). *Variational analysis* (Vol. 317). Springer-Verlag Berlin.

[37] Spreij, P., & Veerman, E. (2012). *Affine diffusions with non-canonical state space.* Stochastic Analysis and Applications, 30(4), 605-641.

[38] Stroock, D. W., & Varadhan, S. R. (1972). On the support of diffusion processes with applications to the strong maximum principle. In Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971) (Vol. 3, pp. 333-359).

[39] Stroock, D. W., & Varadhan, S. S. (1979). *Multidimensional diffusion processes.* Springer.

[40] Tappe, S. (2009). *Stochastic invariance of closed, convex sets with respect to jump-diffusions.* Preprint.