The Annals of Applied Probability
2017, Vol. 27, No. 4, 2383–2418
DOI: 10.1214/16-AAP1262

A STRONG ORDER 1/2 METHOD FOR MULTIDIMENSIONAL SDES WITH DISCONTINUOUS DRIFT

BY GUNTHER LEOBACHER[1] AND MICHAELA SZÖLGYENYI[2]

University of Graz and Vienna University of Economics and Business WU

In this paper, we consider multidimensional stochastic differential equations (SDEs) with discontinuous drift and possibly degenerate diffusion coefficient. We prove an existence and uniqueness result for this class of SDEs and we present a numerical method that converges with strong order $1/2$. Our result is the first one that shows existence and uniqueness as well as strong convergence for such a general class of SDEs.

The proof is based on a transformation technique that removes the discontinuity from the drift such that the coefficients of the transformed SDE are Lipschitz continuous. Thus the Euler–Maruyama method can be applied to this transformed SDE. The approximation can be transformed back, giving an approximation to the solution of the original SDE.

As an illustration, we apply our result to an SDE the drift of which has a discontinuity along the unit circle and we present an application from stochastic optimal control.

1. Introduction. We consider a d-dimensional time-homogeneous stochastic differential equation (SDE):

$$dX = \mu(X)\,dt + \sigma(X)\,dW, \qquad X_0 = x, \tag{1}$$

where $\mu : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \longrightarrow \mathbb{R}^{d\times d}$ are measurable functions and $W = (W_t)_{t\ge 0}$ is a d-dimensional standard Brownian motion on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathbb{P})$.

If both μ and σ are Lipschitz, then existence and uniqueness is guaranteed by Picard iteration. Furthermore, (1) can be solved numerically with, for example, the

Received July 2016.

[1]Supported by the Austrian Science Fund (FWF): Project F5508-N26, which is part of the Special Research Program "Quasi-Monte Carlo Methods: Theory and Applications". The paper was written while G. Leobacher was member of the Department of Financial Mathematics and Applied Number Theory, Johannes Kepler University Linz, 4040 Linz, Austria.

[2]Supported by the Vienna Science and Technology Fund (WWTF): Project MA14-031. A part of this paper was written while M. Szölgyenyi was member of the Department of Financial Mathematics and Applied Number Theory, Johannes Kepler University Linz, 4040 Linz, Austria. During this time, M. Szölgyenyi was supported by the Austrian Science Fund (FWF): Project F5508-N26, which is part of the Special Research Program "Quasi-Monte Carlo Methods: Theory and Applications".

MSC2010 subject classifications. Primary 60H10, 65C30, 65C20; secondary 65L20.

Key words and phrases. Stochastic differential equations, discontinuous drift, degenerate diffusion, existence and uniqueness of solutions, numerical methods for stochastic differential equations, strong convergence rate.

Euler–Maruyama method, which then converges with strong order $1/2$; see [10], Theorem 10.2.2.

However, in applications one is frequently confronted with SDEs where μ is non-Lipschitz, for example, in stochastic control theory. There, whenever an optimal control of bang-bang type appears, meaning that the strategy is of the form $\mathbf{1}_S(X)$ for some measurable set $S \subseteq \mathbb{R}^d$, the drift of the controlled underlying system is discontinuous. Furthermore, for example, in setups with incomplete information, which are currently heavily under study, for example, for applications in mathematical finance, the underlying systems have degenerate diffusion coefficients. Therefore, the class of SDEs that we study in this paper appears frequently in applied mathematics and we shall elaborate our contributions to this kind of problems later in the paper.

The question of existence and uniqueness of solutions to SDEs with non-Lipschitz drift has been studied by various authors.

For the case where μ is only bounded and measurable and σ is bounded, Lipschitz and satisfies a certain uniform ellipticity condition, Zvonkin [24] and Veretennikov [21–23] prove existence and uniqueness of a solution by removing the drift coefficient in a way such that the Lipschitz condition of the diffusion coefficient is preserved.

But uniform ellipticity is a strong assumption which is—as mentioned above—frequently violated in applications.

In Leobacher et al. [15], an existence and uniqueness result for (1) is presented for the case where the drift is potentially discontinuous at a hyperplane, or a special hypersurface, but well behaved everywhere else and where the diffusion coefficient is potentially degenerate. In that paper, not the whole drift is removed, but only the discontinuity is removed locally from the drift.

Due to the weaker requirements on the diffusion coefficient, the restriction to homogeneous SDEs does not pose any loss of generality. In Shardin and Szölgyenyi [18], the authors extend the result from [15] to the time-inhomogeneous case.

In Leobacher and Szölgyenyi [13], an existence and uniqueness result, as well as a numerical method are presented for the one-dimensional case with piecewise Lipschitz drift coefficient. There the coefficients are globally transformed into Lipschitz ones. Both computation of the transformed coefficients and inversion can be done efficiently. This leads to a numerical method for one-dimensional SDEs through application of the Euler–Maruyama scheme on the transformed equation and transforming the approximation back. We present a simplified version of this result in Section 2.

However, extending the result from [13] to the d-dimensional case is far from being straightforward. One problem is that there is no immediate generalization of the concept of a piecewise Lipschitz function with several variables that suits our needs. The second problem is that it is more difficult to obtain a transform that is a Lipschitz diffeomorphism $\mathbb{R}^d \longrightarrow \mathbb{R}^d$. We use Hadamard's global inverse function theorem to prove that our transform is of this kind. Moreover, we need to show that the transform and its inverse are sufficiently well behaved for Itô's formula to hold.

The coefficients of the SDE obtained by transforming the original one are shown to be Lipschitz, such that we can apply the Euler–Maruyama method to the transformed SDE. An approximation to the original SDE is then obtained by applying the inverse transform to the approximation of the transformed solution. For this scheme, we show strong convergence with order $1/2$. One might ask whether the results of Zvonkin and Veretennikov give rise to a similar method. However, in order to apply their method one would have to solve a system of parabolic partial differential equations (in each step). Further, for using this solution in a numerical method like ours, one would also have to find its inverse function. Therefore, such a method, if it exists at all, would be rather costly from the computational perspective.

In the present paper, we present a transform for the multidimensional case which allows to prove an existence and uniqueness result for d-dimensional SDEs with discontinuous drift and degenerate diffusion coefficient under conditions significantly weaker than those in the literature. The essential geometric condition in our setup is that the diffusion must have a component orthogonal to the set of discontinuities of the drift.

Furthermore, we present a numerical method for such SDEs based on the ideas outlined above. To the authors' knowledge, there is no other numerical method that can deal with such a general class of SDEs and gives strong convergence, much less giving a strong convergence *rate*.

We are now going to review the literature on numerical methods for SDEs with nonglobally Lipschitz drift coefficient. In Berkaoui [1], strong convergence of the Euler–Maruyama scheme is proven under the assumption that the drift is of class C^1. For an SDE with continuously differentiable but nonglobally Lipschitz drift, Hutzenthaler et al. [7] introduce a new explicit numerical scheme—the tamed Euler scheme—and prove its strong convergence. Sabanis [17] proves strong convergence of the tamed Euler scheme for SDEs with one-sided Lipschitz drift. For the Euler–Maruyama scheme, Gyöngy [5] proves almost sure convergence for the case that the drift satisfies a monotonicity condition. A different approach is introduced by Halidias and Kloeden [6], who show that the Euler–Maruyama scheme converges strongly for SDEs with a discontinuous monotone drift coefficient, especially mentioning the case in which the drift is a Heaviside function. Kohatsu-Higa et al. [11] show weak convergence of a method where they first regularize the drift and then apply the Euler–Maruyama scheme. They allow the drift to be discontinuous. Étoré and Martinez [2, 3] introduce an exact simulation algorithm for one-dimensional SDEs that have a bounded drift coefficient being discontinuous in one point, but differentiable everywhere else.

This paper is organized as follows. In Section 2, we present the one-dimensional result and algorithm in a form that can be generalized to multiple dimensions, which is subsequently done in Section 3. In Section 4, we give two numerical examples: one where the drift coefficient has discontinuities along the unit circle in $\mathbb{R}^2$ and an example from stochastic optimal control.

Some of the more technical and geometrical proofs have been moved to the Appendix.

2. The one-dimensional problem. Here, we consider the one-dimensional version of SDE (1) and give simple conditions for existence and uniqueness of a solution and a strong order $1/2$ algorithm. For this, we recall the following definition.

DEFINITION 2.1. Let $I \subseteq \mathbb{R}$ be an interval. We say a function $f : I \longrightarrow \mathbb{R}$ is piecewise Lipschitz, if there are finitely many points $\xi_1 < \cdots < \xi_m \in I$ such that f is Lipschitz on each of the intervals $(-\infty, \xi_1) \cap I$, $(\xi_m, \infty) \cap I$ and (ξ_k, ξ_{k+1}), $k = 1, \dots, m$.

We make the following assumptions on the coefficients.

ASSUMPTION 2.1. The drift coefficient $\mu : \mathbb{R} \longrightarrow \mathbb{R}$ is piecewise Lipschitz.

ASSUMPTION 2.2. The diffusion coefficient $\sigma : \mathbb{R} \longrightarrow \mathbb{R}$ is Lipschitz with $\sigma(\xi) \neq 0$ whenever $\mu(\xi+) \neq \mu(\xi-)$.

For simplicity, we derive the result for $\mu : \mathbb{R} \longrightarrow \mathbb{R}$ that is Lipschitz with the exception of only a single point ξ where μ is allowed to jump. We are going to construct a transform $G : \mathbb{R} \longrightarrow \mathbb{R}$ such that the process formally defined by $Z = G(X)$ satisfies an SDE with Lipschitz coefficients and, therefore, has a solution by Itô's classical theorem on existence and uniqueness of solutions; see [8].

For this, define the following bump function on $\mathbb{R}$, which we need to localize the impact of the transform G:

$$\phi(u) = \begin{cases} (1+u)^3(1-u)^3 & \text{if } |u| \leq 1, \\ 0 & \text{else.} \end{cases} \tag{2}$$

The function ϕ has the following properties:

1. ϕ defines a C^2 function on all of $\mathbb{R}$;
2. $\phi(0) = 1$, $\phi'(0) = 0$, $\phi''(0) = -6$;
3. $\phi(u) = \phi'(u) = \phi''(u) = 0$ for all $|u| \geq 1$.

We define the transform $G : \mathbb{R} \longrightarrow \mathbb{R}$ by

$$G(x) = x + \alpha\phi\left(\frac{x-\xi}{c}\right)(x-\xi)|x-\xi|, \qquad x \in \mathbb{R}, \tag{3}$$

where $\alpha \neq 0$ and $c > 0$ are some constants.

LEMMA 2.2. *Let* $c < \frac{1}{6|\alpha|}$.

Then $G'(x) > 0$ *for all* $x \in \mathbb{R}$. *Furthermore*, $G'(x) = 1$ *for all* $|x - \xi| > c$. *Therefore*, G *has a global inverse* G^{-1}.

PROOF. Differentiating G for $|x - \xi| \le c$ yields

$$G'(x) = 1 - \frac{6\alpha}{c^2}(x-\xi)^2|x-\xi|\left(1 + \frac{x-\xi}{c}\right)^2\left(1 - \frac{x-\xi}{c}\right)^2 \\ + 2\alpha|x-\xi|\left(1 + \frac{x-\xi}{c}\right)^3\left(1 - \frac{x-\xi}{c}\right)^3.$$

For positive α, this is positive, if $c < \frac{1}{6|\alpha|}$. For negative α, it is positive, if $c < \frac{1}{2|\alpha|}$. Altogether a sufficient condition for G' to be positive is $c < \frac{1}{6|\alpha|}$. $\square$

W.l.o.g. we always choose $c < \frac{1}{6|\alpha|}$, such that G has a global inverse.

REMARK 2.3. In [13], the function G is constructed differently. There G is piecewise cubic, such that G^{-1} is piecewise radical, and hence admits exact inversion, which is advantageous for the numerical treatment.

In fact, G can be made piecewise cubic by still using equation (3), but with a different choice for ϕ. Actually, any function ϕ with support contained in $[-1, 1]$ satisfying properties 1, 2, 3 from page 2386 will give rise to a transform G sufficient for our purpose, with a similar condition on the constant c for G to be invertible. The form chosen here is simple in the one-dimensional case and has a direct multidimensional analog.

Formally define $Z = G(X)$. Abbreviating $\bar{\phi}(x) := \phi(\frac{x-\xi}{c})(x-\xi)|x-\xi|$, we have

$$\begin{aligned} dZ &= dX + \alpha\bar{\phi}'(X)\,dX + \frac{1}{2}\alpha\bar{\phi}''(X)\,d[X] \\ &= \left(\mu(X) + \alpha\bar{\phi}'(X)\mu(X) + \frac{1}{2}\alpha\bar{\phi}''(X)\sigma(X)^2\right)dt \\ &\quad + \big(\sigma(X) + \alpha\bar{\phi}'(X)\sigma(X)\big)\,dW \\ &= \tilde{\mu}(Z)\,dt + \tilde{\sigma}(Z)\,dW, \end{aligned} \tag{4}$$

where

$$\tilde{\mu}(z) = \mu\big(G^{-1}(z)\big) + \alpha\bar{\phi}'\big(G^{-1}(z)\big)\mu\big(G^{-1}(z)\big) + \frac{1}{2}\alpha\bar{\phi}''\big(G^{-1}(z)\big)\sigma\big(G^{-1}(z)\big)^2,$$

$$\tilde{\sigma}(z) = \sigma\big(G^{-1}(z)\big) + \alpha\bar{\phi}'\big(G^{-1}(z)\big)\sigma\big(G^{-1}(z)\big).$$

We now show that, for an appropriate choice of α, the transformed drift $\tilde{\mu}$ is Lipschitz. For this, we need the following elementary lemma from [13].

LEMMA 2.4. *Let $f:\mathbb{R}\longrightarrow\mathbb{R}$ be piecewise Lipschitz and continuous. Then f is Lipschitz on $\mathbb{R}$.*

From Lemma 2.4 and $\lim_{h\to 0}\bar\phi'(h)=0$, we see that the mapping $z\mapsto \bar\phi'(G^{-1}(z))\mu(G^{-1}(z))$ is Lipschitz. In order to make the mapping $z\mapsto \mu(G^{-1}(z))+\frac{1}{2}\alpha\bar\phi''(G^{-1}(z))\sigma(G^{-1}(z))^2$ continuous, we need to choose α so that

$$\mu\bigl(G^{-1}(\xi+)\bigr)+\frac{1}{2}\alpha\bar\phi''\bigl(G^{-1}(\xi+)\bigr)\sigma\bigl(G^{-1}(\xi+)\bigr)^2 = \mu\bigl(G^{-1}(\xi-)\bigr)+\frac{1}{2}\alpha\bar\phi''\bigl(G^{-1}(\xi-)\bigr)\sigma\bigl(G^{-1}(\xi-)\bigr)^2,$$

i.e.

$$\mu(\xi+)+\frac{1}{2}\alpha\bar\phi''(\xi+)\sigma(\xi)^2=\mu(\xi-)+\frac{1}{2}\alpha\bar\phi''(\xi-)\sigma(\xi)^2.$$

Thus we get, for the choice

$$\alpha=-2\frac{\mu(\xi+)-\mu(\xi-)}{(\bar\phi''(\xi+)-\bar\phi''(\xi-))\sigma(\xi)^2}=\frac{\mu(\xi-)-\mu(\xi+)}{2\sigma(\xi)^2}$$

that $\tilde\mu$ is continuous. Note that at this point we need nondegeneracy of σ in ξ.

Since $\tilde\mu$ is continuous with the appropriate choice of α, it is Lipschitz as well by Lemma 2.4.

One may worry about the quadratic occurrence of σ in the expression for $\tilde\mu$. Note, however, that $\bar\phi''$ vanishes outside $[-c,c]$.

To prove that $\tilde\sigma$ is Lipschitz as well, we need the following lemma.

LEMMA 2.5. *Let $f:\mathbb{R}\longrightarrow\mathbb{R}$ be Lipschitz. Then $f\phi'$ is Lipschitz.*

PROOF. Let L_f be a Lipschitz constant for f. Note that 6 is a Lipschitz constant for ϕ'. If $|x|,|y|\le 1$, then

$$\begin{aligned}\bigl|f(x)\phi'(x)-f(y)\phi'(y)\bigr| &\le \bigl|f(x)\phi'(x)-f(y)\phi'(x)\bigr|\\ &\quad+\bigl|f(y)\phi'(x)-f(y)\phi'(y)\bigr|\\ &\le L_f|x-y|\max_{z\in[-1,1]}\bigl|\phi'(z)\bigr|\\ &\quad+6|x-y|\max_{z\in[-1,1]}\bigl|f(z)\bigr|\\ &\le K|x-y|,\end{aligned}$$

where $K=L_f|x-y|\max_{z\in[-1,1]}|\phi'(z)|+6|x-y|\max_{z\in[-1,1]}|f(z)|$. For $-1\le x\le 1<y$, we have

$$\begin{aligned}\bigl|f(x)\phi'(x)-f(y)\phi'(y)\bigr|&=\bigl|f(x)\phi'(x)\bigr|=\bigl|f(x)\phi'(x)-f(1)\phi'(1)\bigr|\bigr|\\ &\le K|x-1\le K|x-y|.\end{aligned}$$

The same estimate holds for the case $x < -1 \le y \le 1$. For $|x|, |y| > 1$, we have $|f(x)\phi'(x) - f(y)\phi'(y)| = 0 \le K|x - y|$. $\square$

Thus, $\tilde{\sigma}$ is Lipschitz by Lemma 2.5 and the fact that the composition of Lipschitz functions is Lipschitz.

Altogether we have that the SDE (4) for Z has Lipschitz coefficients $\tilde{\mu}$ and $\tilde{\sigma}$.

The generalization to finitely many discontinuities of μ in the points $\xi_1 < \cdots < \xi_m$ is now straightforward: define

$$G(x) := x + \sum_{k=1}^{m} \alpha_k \phi\left(\frac{x - \xi_k}{c}\right)(x - \xi_k)|x - \xi_k|,$$

with

$$\alpha_k = \frac{\mu(\xi_k-) - \mu(\xi_k+)}{2\sigma(\xi_k)^2} \quad \text{and} \quad c < \min\left(\min_{1\le k\le m} \frac{1}{6|\alpha_k|}, \min_{1\le k\le m-1} \frac{\xi_{k+1} - \xi_k}{2}\right).$$

We are ready to prove existence and uniqueness of a solution to the one-dimensional SDE (1).

THEOREM 2.6 (cf. [20], Theorem 2.2). *Let Assumptions* 2.1, *and* 2.2 *be satisfied*, *that is*, μ *is piecewise Lipschitz with finitely many jump points*, σ *is Lipschitz and* $\forall \xi : \mu(\xi+) \ne \mu(\xi-) \Longrightarrow \sigma(\xi) \ne 0$.

Then the one-dimensional SDE (1) *has a unique global strong solution.*

PROOF. Since the SDE (4) for Z has Lipschitz coefficients, it follows that (4) with initial condition $Z_0 = G(x)$ has a unique global strong solution. Furthermore, G has a global inverse G^{-1}, which inherits the smoothness from G. Although $G^{-1} \notin C^2$, Itô's formula holds for G^{-1}; see [9], 5. Problem 7.3. Applying Itô's formula to G^{-1}, we obtain that $G^{-1}(Z)$ satisfies

$$dX = \mu(X)\,dt + \sigma(X)\,dW, \qquad X_0 = x.$$

Setting $X = G^{-1}(Z)$ yields the desired result. $\square$

For approximating the solution to the one-dimensional SDE (1), we propose the following numerical method. Let $Z_T^{(\delta)}$ be the Euler–Maruyama approximation of the solution to SDE (4) with step size smaller than $\delta > 0$.

ALGORITHM 2.7. Go through the following steps:

1. Set $Z_0^{(\delta)} = G(x)$.
2. Apply the Euler–Maruyama method to the SDE (4) to obtain $Z_T^{(\delta)}$.
3. Set $\bar{X} = G^{-1}(Z_T^{(\delta)})$.

THEOREM 2.8 (cf. [20], Theorem 3.1). *Let Assumptions 2.1, and 2.2 be satisfied.*

Then Algorithm 2.7 converges with strong order $1/2$ to the solution X of the one-dimensional SDE (1).

PROOF. We estimate the L^2-error of the approximation. For every $T > 0$ there is a constant C, such that

$$\begin{aligned}\mathbb{E}\big((X_T - \bar{X}_T)^2\big) &= \mathbb{E}\big((G^{-1}(Z_T) - G^{-1}(Z_T^{(\delta)}))^2\big)\\ &\le L_{G^{-1}}^2 \mathbb{E}\big((Z_T - Z_T^{(\delta)})^2\big) = L_{G^{-1}}^2 C\delta\end{aligned}$$

for every sufficiently small step size δ, where $L_{G^{-1}}$ is the Lipschitz constant of G^{-1} and where we applied [10], Theorem 10.2.2, for the L^2-convergence of the Euler–Maruyama scheme for SDEs with Lipschitz coefficients. □

3. The multidimensional problem. We now consider the multidimensional case. Like in dimension one, we will have to make assumptions on the drift so that it is Lipschitz apart from—relatively few—locations of discontinuity. That is, we need a concept similar to that of "piecewise Lipschitz" in the one-dimensional case. We will develop such a concept now.

In contrast to the one-dimensional case, we shall have to make additional assumptions on the behaviour of the drift close to its points of discontinuity, which shall all lie in a hypersurface Θ.

Regarding the diffusion coefficient we need to find a condition corresponding to Assumption 2.2.

Note that most of these assumptions are automatically satisfied, or can at least be weakened, if Θ is compact. We will treat the case of compact Θ in Section 3.6.

3.1. *Piecewise Lipschitz functions.* For a continuous curve $\gamma : [0, 1] \longrightarrow \mathbb{R}^d$, let $\ell(\gamma)$ denote its length,

$$\ell(\gamma) := \sup_{n, 0 \le t_1 < \cdots < t_n \le 1} \sum_{k=1}^{n} \|\gamma(t_k) - \gamma(t_{k-1})\| \in [0, \infty].$$

DEFINITION 3.1. Let $A \subseteq \mathbb{R}^d$. The *intrinsic metric d* on A is given by

$$\begin{aligned}\rho(x, y) := \inf\big\{\ell(\gamma) : \gamma : [0, 1] \to A \text{ is a continuous curve}\\ \text{satisfying } \gamma(0) = x, \gamma(1) = y\big\},\end{aligned}$$

where $\rho(x, y) := \infty$, if there is no continuous curve from x to y.

DEFINITION 3.2. Let $A \subseteq \mathbb{R}^d$. Let $f : A \longrightarrow \mathbb{R}^m$ be a function. We say that f is *intrinsic Lipschitz*, if it is Lipschitz w.r.t. the intrinsic metric on A, that is, if there exists a constant L such that

$$\forall x, y \in A: \qquad \|f(x) - f(y)\| \le L\rho(x, y).$$

REMARK 3.3. Note that for a function $f:\mathbb{R}\longrightarrow\mathbb{R}$ we have that f is piecewise Lipschitz, iff f is intrinsic Lipschitz on $\mathbb{R}\setminus B$, where B is a finite subset of $\mathbb{R}$.

This motivates the following definition.

DEFINITION 3.4. A function $f:\mathbb{R}^d\longrightarrow\mathbb{R}^m$ is *piecewise Lipschitz*, if there exists a hypersurface[3] Θ with finitely many components and with the property, that the restriction $f|_{\mathbb{R}^d\setminus\Theta}$ is intrinsic Lipschitz. We call Θ an *exceptional set* for f.

The definition is more general than the more obvious requirement that $\mathbb{R}^d$ can be partitioned into finitely many patches in a way such that f is Lipschitz on all of the patches. This is illustrated by the following example.

EXAMPLE 3.5. Consider the function $f:\mathbb{R}^2\longrightarrow\mathbb{R}$, $f(x)=\|x\|\arg(x)$. Then f is not Lipschitz, since $\lim_{h\to 0+} f(\cos(\pi-h),\sin(\pi-h))=\pi$ and $\lim_{h\to 0+} f(\cos(\pi+h),\sin(\pi+h))=-\pi$ for $x_1<0$.

It is readily checked, however, that f is intrinsic Lipschitz on $A=\mathbb{R}^2\setminus\{x\in\mathbb{R}^2:x_1<0,x_2=0\}$ and $\{x\in\mathbb{R}^2:x_1<0,x_2=0\}$ is obviously a one-dimensional submanifold of $\mathbb{R}^2$.

Thus, f is piecewise Lipschitz in the sense of Definition 3.4.

The following lemma is a multidimensional generalization of Lemma 2.4.

LEMMA 3.6. *Let $f:\mathbb{R}^d\longrightarrow\mathbb{R}^m$ be a function. If*:

1. *f is continuous in every point $x\in\mathbb{R}^d$*;
2. *f is piecewise Lipschitz with exceptional set Θ*;
3. *for $x,y\in\mathbb{R}^d$ and $\eta>0$ there exists a continuous curve γ from x to y with $\ell(\gamma)<\|x-y\|+\eta$ such that $\#(\gamma\cap\Theta)<\infty$.*

Then f is Lipschitz on $\mathbb{R}^d$ w.r.t. the Euclidean metric, *and with the same Lipschitz constant*.

PROOF. Let L be the intrinsic Lipschitz constant of f, that is, $\|f(y)-f(x)\|\le L\rho(x,y)$ for all $x,y\in\mathbb{R}^d$, and let $x,y\in\mathbb{R}^d$. If $\rho(x,y)=\|x-y\|$, then clearly $\|f(x)-f(y)\|\le L\rho(x,y)=L\|x-y\|$.

If $\rho(x,y)>\|x-y\|$, then the line segment $s(x,y):=\{(1-\lambda)x+\lambda y:\lambda\in[0,1]\}$ has nonempty intersection with Θ.

Consider first the case where $s(x,y)\cap\Theta=\{z_1,\dots,z_n\}$, that is, we have finite intersection. There exist $\lambda_1,\dots,\lambda_n$ such that $z_k=(1-\lambda_k)x+\lambda_k y$. Define $g:[0,1]\longrightarrow\mathbb{R}^m$ by $g(\lambda):=f((1-\lambda)x+\lambda y)$.

[3]By a hypersurface, we mean a $(d-1)$-dimensional submanifold of the $\mathbb{R}^d$.

Set $z_0 = x$, $z_{n+1} = y$, $\lambda_0 = 0$, $\lambda_{n+1} = 1$. W.l.o.g., $\lambda_0 < \cdots < \lambda_{n+1}$. Now

$$
\begin{aligned}
\|f(y) - f(x)\| &= \left\| \sum_{k=1}^{n+1} f(z_k) - f(z_{k-1}) \right\| \\
&\le \sum_{k=1}^{n+1} \|f(z_k) - f(z_{k-1})\| = \sum_{k=1}^{n+1} \|g(\lambda_k) - g(\lambda_{k-1})\| \\
&= \lim_{h\to 0+} \sum_{k=1}^{n+1} \|g(\lambda_k - h) - g(\lambda_{k-1} + h)\| \\
&\le \lim_{h\to 0+} \sum_{k=1}^{n+1} L\rho\big(\big((1-\lambda_k + h)x + (\lambda_k - h)y\big), \\
&\quad \big((1-\lambda_{k-1} - h)x + (\lambda_{k-1} + h)y\big)\big) \\
&= \lim_{h\to 0+} \sum_{k=1}^{n+1} L\big\|\big((1-\lambda_k + h)x + (\lambda_k - h)y\big) \\
&\quad - \big((1-\lambda_{k-1} - h)x + (\lambda_{k-1} + h)y\big)\big\| \\
&= \sum_{k=1}^{n+1} L\|z_k - z_{k-1}\| = L\|y - x\|,
\end{aligned}
$$

where we have used the continuity of f and g, and that the intrinsic metric coincides with the Euclidean metric for pairs of points for which the connecting line segment has empty intersection with Θ.

If $s(x, y) \cap \Theta$ contains infinitely many points, we can replace $s(x, y)$ by γ, which is only slightly longer than $s(x, y)$, but has only finitely many intersections with Θ. A slight modification of the argument above then gives that $\|f(y) - f(x)\| < L\|y - x\| + \varepsilon$ for any $\varepsilon > 0$, and thus the desired result. □

CONJECTURE 3.7. *Item 3 of the assumptions of Lemma 3.6 is not necessary to prove the assertion of the lemma.*

We will later give sufficient conditions for item 3 of the assumptions of Lemma 3.6 to hold; see Lemma 3.11. These conditions are satisfied in our applications.

It is well known that differentiable functions with bounded derivative are Lipschitz w.r.t. the Euclidean metric. The same holds true for the intrinsic metric.

LEMMA 3.8. *Let $A \subseteq \mathbb{R}^d$ be open and let $f : A \longrightarrow \mathbb{R}^m$ be differentiable with $\|f'\| \le K$.*

Then f is intrinsic Lipschitz with constant K.

PROOF. Let $x, y \in A$ and let γ be a continuous curve of finite length with $\gamma(0) = x$ and $\gamma(1) = y$. [If no such curve exists, we trivially have $\|f(y) - f(x)\| \le K\rho(x, y) = \infty$.] Let $0 = t_0 < \cdots < t_n = 1$. Without loss of generality, the t_k can be chosen such that the line segment spanned by $\gamma(t_{k-1})$ and $\gamma(t_k)$ is in A for every k. Then

$$\begin{aligned}\|f(y) - f(x)\| &\le \sum_{k=1}^{n} \|f(\gamma(t_k)) - f(\gamma(t_{k-1}))\| \\ &\le \sum_{k=1}^{n} \sup_{t\in(t_{k-1},t_k)} \|f'(\gamma(t))\| \|\gamma(t_k) - \gamma(t_{k-1})\| \\ &\le K \sum_{k=1}^{n} \|\gamma(t_k) - \gamma(t_{k-1})\| \le K\ell(\gamma).\end{aligned}$$

□

Furthermore, we prove that the composition of an intrinsic Lipschitz function with a Lipschitz function is intrinsic Lipschitz.

LEMMA 3.9. *Let $A \subseteq \mathbb{R}^d$ be open. Let $g : \mathbb{R}^d \longrightarrow A$ be Lipschitz with constant L_g. Let $f : A \longrightarrow \mathbb{R}^m$ be intrinsic Lipschitz with constant L_f.*
Then $f \circ g$ is intrinsic Lipschitz with constant $L_f L_g$.

PROOF. Let γ be a continuous curve of finite length with $\gamma(0) = x$ and $\gamma(1) = y$. [If no such curve exists, we trivially have $\|f(y) - f(x)\| \le L_g\rho(x, y) = \infty$.] Let $0 = t_0 < \cdots < t_n = 1$. For every $\delta > 0$, there are $0 = \bar{t}_0 < \cdots < \bar{t}_{\bar{n}} = 1$ such that $\rho(g(x), g(y)) < \sum_{k=1}^{\bar{n}} \|g(\bar{t}_k) - g(\bar{t}_{k-1})\| + \delta/L_f$. So

$$\begin{aligned}\sum_{k=1}^{n} \|f \circ g(\gamma(t_k)) - f \circ g(\gamma(t_{k-1}))\| &\le L_f \sum_{k=1}^{n} \|g(t_k) - g(t_{k-1})\| \\ &\le L_f \rho(g(x), g(y)) \\ &< L_f \left(\sum_{k=1}^{\bar{n}} \|g(\bar{t}_k) - g(\bar{t}_{k-1})\| + \delta/L_f \right) \\ &< L_f \left(L_g \sum_{k=1}^{\bar{n}} \|\bar{t}_k - \bar{t}_{k-1}\| + \delta/L_f \right) \\ &\le L_f L_g \ell(\gamma) + \delta.\end{aligned}$$

Since $\delta > 0$ was arbitrary, we obtain the result. □

3.2. *The form of the set of discontinuities.* We are going to generalize the idea of transforming a discontinuous drift into a Lipschitz one to general dimensions.

For this, we assume that the drift coefficient μ is piecewise Lipschitz in the sense of Definition 3.4, that is, there exists a hypersurface Θ with finitely many components such that $\mu|_{\mathbb{R}^d \setminus \Theta}$ is intrinsic Lipschitz. The assumption on the drift that will make our method work therefore encompasses assumptions on Θ.

ASSUMPTION 3.1. The drift coefficient μ is a piecewise Lipschitz function $\mathbb{R}^d \longrightarrow \mathbb{R}^d$. Its exceptional set Θ is a C^3 hypersurface.

A consequence of Assumption 3.1 is that locally there exists a C^2 orthonormal vector, that is, for every sufficiently small open and connected $B \subseteq \Theta$ there exists an orthonormal vector on B, that is, a C^2-function $n : B \longrightarrow \mathbb{R}^d$ such that for all $\xi \in B$ the vector $n(\xi)$ is orthogonal to the tangent space of Θ in ξ and $\|n(\xi)\| = 1$. It is well known, that there are in general two possible choices for n and that one can take $B = \Theta$ only if Θ is orientable. But given n on B, the only other orthonormal vector is $-n$.

Define the distance $d(x, \Theta)$ between a point x and the hypersurface Θ in the usual way, $d(x, \Theta) := \inf\{\|x - y\| : y \in \Theta\}$. For every $\varepsilon > 0$, we define $\Theta^\varepsilon := \{x \in \mathbb{R}^d : d(x, \Theta) < \varepsilon\}$.

ASSUMPTION 3.2. There exists $\varepsilon_0 > 0$ such that Θ^{ε_0} has the *unique closest point property*, that is, for every $x \in \mathbb{R}^d$ with $d(x, \Theta) < \varepsilon_0$ there is a unique $p \in \Theta$ with $d(x, \Theta) = \|x - p\|$.

A set possessing the property described in Assumption 3.2 is called a *set of positive reach. The reach of a set Θ is the supremum over all $\varepsilon_0 > 0$ such that Θ^{ε_0} has the* unique closest point property. This and the notion of *unique closest point property* can be found in [12].

LEMMA 3.10. *Let Θ be a C^3-hypersurface.*
If Θ is of positive reach, then $\|n'\|$ is bounded:

$$\|n'(\xi)\| \leq 2 \frac{d-1}{\operatorname{reach}(\Theta)}$$

for all $\xi \in \Theta$.

The proof of Lemma 3.10 can be found in the Appendix.

Note that one can find examples of hypersurfaces with bounded $\|n'\|$ which are not of positive reach; see Figure 1.

Due to Assumption 3.2, there exists an $\varepsilon_0 > 0$ for which we may define a mapping $p : \Theta^{\varepsilon_0} \longrightarrow \Theta$ assigning to each x the point $p(x)$ in Θ closest to x.

LEMMA 3.11. *If Θ is a C^3-hypersurface that satisfies Assumption 3.2, then item 3 of Lemma 3.6 is satisfied, that is, for $x, y \in \mathbb{R}^d$ and $\eta > 0$ there exists a continuous curve γ from x to y with $\ell(\gamma) < \|x - y\| + \eta$ such that $\#(\gamma \cap \Theta) < \infty$.*

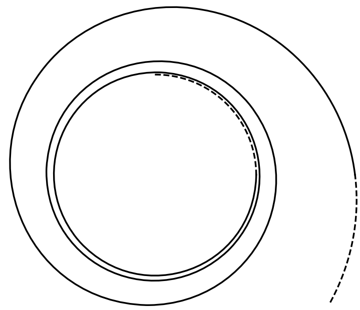

FIG. 1. *A hypersurface in $\mathbb{R}^2$ with bounded $\|n'\|$ that is not of positive reach.*

The rather technical proof of this lemma can be found in the Appendix. Note that for many examples, like a (hyper-)sphere or hyperplane, item 3 of Lemma 3.6 is obviously satisfied. So in these cases there is no need to resort to Lemma 3.11. However, it is an interesting fact that this condition is automatically satisfied under our assumptions on Θ.

3.3. *Construction of the transform G.* As before, we construct a transform G with the property that the SDE for $G(X)$ has Lipschitz coefficients.

For this to be well-defined, we make the following assumption.

ASSUMPTION 3.3. There is a constant $c_0 > 0$ such that $\|\sigma(\xi)^\top n(\xi)\| \geq c_0$ for all $\xi \in \Theta$.

REMARK 3.12. Assumption 3.3 is a *nonparallelity condition*, meaning that for all $\xi \in \Theta$, $\sigma(\xi)$ must not be parallel to Θ, in the sense that there exists some $x \in \mathbb{R}^d$ such that $\sigma(\xi)x$ is not in the tangent space of Θ in ξ.

Assumption 3.3 is by far weaker than uniform ellipticity. For the practical example we study in Section 4, it is satisfied, whereas uniform ellipticity clearly is not.

For defining the transform, we first switch to a local setting. Suppose $\tilde{x} \in \mathbb{R}^d$ is close to Θ, that is, $d(\tilde{x}, \Theta) < \varepsilon_0$. Let $B \subseteq \Theta$ be an open environment of $p(x)$ in Θ and n an orthonormal vector. It follows that the set

$$U = \{y_1 n(\xi) + \xi : y_1 \in (-\varepsilon_0, \varepsilon_0), \xi \in B\}$$

is an open environment of $\tilde{x}$, and every point $x \in U$ can be uniquely represented in the form $x = y_1 n(\xi) + \xi$, $y_1 \in (-\varepsilon_0, \varepsilon_0)$, $\xi \in B$.

We are now ready to locally define the transform $G : U \longrightarrow \mathbb{R}^d$ by

$$G(x) = x + \tilde{\phi}(x)\alpha\big(p(x)\big), \tag{5}$$

where $\tilde{\phi}(x) = (x - p(x)) \cdot n(p(x)) \|x - p(x)\| \phi(\frac{\|x-p(x)\|}{c})$, with ϕ as in (2) and where

$$\alpha(\xi) := \lim_{h \to 0} \frac{\mu(\xi - hn(\xi)) - \mu(\xi + hn(\xi))}{2n(\xi)^\top \sigma(\xi)\sigma(\xi)^\top n(\xi)}, \qquad \xi \in B. \tag{6}$$

One important point to note is the following proposition.

PROPOSITION 3.13. *The value of the function G does not depend on the choice of the orthonormal vector.*

PROOF. Both $\alpha(p(x))$ and $\tilde{\phi}(x)$ depend on the parametrization only through the direction of the normal vector $n(p(x))$. But from the definitions of $\tilde{\phi}$ and α, we see that if $n(p(x))$ is replaced by $-n(p(x))$, then $\tilde{\phi}(x)$ and $\alpha(p(x))$ both change sign. Therefore, $\tilde{\phi}(x)\alpha(p(x))$ does not depend on the particular choice of the orthonormal vector. $\square$

The only reason why we defined G locally at first was that for a nonorientable hypersurface we do not have, by definition, a global orthonormal vector. However, since the value of the locally defined function G does not depend on the particular choice of the orthonormal vector, we can use the same equations (5) and (6) for defining G globally on Θ^{ε_0}. That is, the function $G : \mathbb{R}^d \longrightarrow \mathbb{R}^d$,

$$G(x) = \begin{cases} x + \tilde{\phi}(x)\alpha\big(p(x)\big), & x \in \Theta^{\varepsilon_0}, \\ x, & x \in \mathbb{R}^d \setminus \Theta^{\varepsilon_0} \end{cases}$$

is well-defined. Note further that, if we require $c \leq \varepsilon_0$, then from $d(x, \Theta) > \varepsilon_0$ it follows that $d(x, \Theta) > c$ and, therefore, $\phi(\frac{\|x-p(x)\|}{c}) = 0$ with a C^2-smooth paste to 0 in all points x satisfying $d(x, \theta) = c$.

3.4. *Properties of G.* We need to prove the following:

1. c can be chosen in a way such that G is a diffeomorphism $\mathbb{R}^d \longrightarrow \mathbb{R}^d$;
2. Itô's formula holds for G^{-1};
3. the SDE for $G(X)$ has Lipschitz coefficients.

ASSUMPTION 3.4. There is a constant a such that every locally defined function α as defined in (6) is C^3 and all derivatives up to order 3 are bounded by a.

THEOREM 3.14. *Let Assumptions 3.1–3.4 be satisfied. If the constant $c > 0$ appearing in the definition of $\tilde{\phi}$ is sufficiently small, then G is a diffeomorphism $\mathbb{R}^d \longrightarrow \mathbb{R}^d$.*

For proving Theorem 3.14, we first need to prove two technical lemmas. For every $\xi \in \Theta$, denote by $\tau(\xi)$ the tangent space of Θ in ξ.

LEMMA 3.15. *For $\xi \in \Theta$, n' is a linear mapping from $\tau(\xi)$ into $\tau(\xi)$.*

PROOF. n' is by definition a linear mapping $\tau(\xi) \longrightarrow \mathbb{R}^d$. Furthermore, we have $\|n\| = 1$, so that for any curve γ in Θ:

$$0 = \frac{d}{dt}\|n(\gamma(t))\|^2 = 2n(\gamma(t)) \cdot (n'(\gamma(t))\gamma'(t)).$$

If $b \in \tau(\xi)$, we can find a curve γ in Θ such that $\gamma(0) = \xi$ and $\gamma'(0) = b$. Thus, $n(\xi) \cdot (n'(\xi)b) = 0$, that is, $n'(\xi)b \in \tau(\xi)$. $\square$

REMARK 3.16. If Θ is C^3 and of positive reach ε_0, then we may choose $0 < \varepsilon < \varepsilon_0$ such that, whenever $y_1 \in \mathbb{R}$ with $|y_1| < \varepsilon$, then $\mathrm{id}_{\tau(\xi)} + y_1 n'(\xi)$ is invertible.

Indeed, let K be a bound on $\|n'\|$ and let $\varepsilon = \frac{\varepsilon_0}{\varkappa \max(K,1)}$ for some fixed $\varkappa > 1$. Then for $|y_1| < \varepsilon$ we have $\|y_1 n'(\xi)\| = |y_1|\|n'(\xi)\| < \frac{1}{\varkappa} < 1$, such that $\mathrm{id}_{\tau(\xi)} + y_1 n'(\xi)$ is invertible by the subsequent well-known Lemma 3.17.

LEMMA 3.17. *Let $\mathcal{A}$ be a linear operator on a subspace $V \subseteq \mathbb{R}^d$ and let $\mathcal{A}$ have (operator) norm smaller than* 1.

Then $\mathrm{id}_V + \mathcal{A}$ is invertible and $\|(\mathrm{id}_V + \mathcal{A})^{-1}\| \le (1 - \|\mathcal{A}\|)^{-1}$.

PROOF. Consider the Neumann series $\mathcal{B} = \sum_{k=0}^{\infty}(-\mathcal{A})^k$ which converges in operator norm and satisfies $\|\mathcal{B}\| \le (1 - \|\mathcal{A}\|)^{-1}$. Then

$$(\mathrm{id}_V + \mathcal{A})\mathcal{B} = \sum_{k=0}^{\infty}(-\mathcal{A})^k - \sum_{k=0}^{\infty}(-\mathcal{A})(-\mathcal{A})^k = \sum_{k=0}^{\infty}(-\mathcal{A})^k - \sum_{k=1}^{\infty}(-\mathcal{A})^k = \mathrm{id}_V.$$

Thus, $\mathcal{B}$ is the inverse of $\mathrm{id}_V + \mathcal{A}$. $\square$

PROOF OF THEOREM 3.14. Fix some $\varkappa > 1$ and set $\varepsilon = \frac{\varepsilon_0}{\varkappa \max(K,1)}$, where K is a bound on $\|n'\|$, which exists by Lemma 3.10.

Let $0 < c < \varepsilon$.

For $\tilde{x} \notin \Theta^c$, differentiability of G in $\tilde{x}$ is obvious.

For $\tilde{x} \in \Theta^c$, choose an open subset B of Θ (as before) and an orthonormal vector n such that $U \subset \mathbb{R}^d$ is an open set with $U \cap \Theta = B$ and every $x \in U$ can uniquely be written in the form $x = y_1 n(\xi) + \xi$ with $\xi = p(x)$. Θ can be parametrized locally by a one-one mapping $\psi : R \longrightarrow \mathbb{R}^d$, where $R \subseteq \mathbb{R}^{d-1}$ is an open rectangle in $\mathbb{R}^{d-1}$, and there is a point $(\tilde{y}_2, \dots, \tilde{y}_d) \in R$ such that $\psi(\tilde{y}_2, \dots, \tilde{y}_d) = p(\tilde{x})$. By making R and/or B smaller, if necessary, we may w.l.o.g. assume that $B = \psi(R)$.

Thus, we have a bijective mapping $\mathscr{T} : (-\varepsilon, \varepsilon) \times R \longrightarrow U$,

$$\mathscr{T}(y_1, \dots, y_d) := y_1 n(\psi(y_2, \dots, y_d)) + \psi(y_2, \dots, y_d), \qquad y \in (-\varepsilon, \varepsilon) \times R.$$

Note that $p(\mathscr{T}(y)) = \psi(y_2, \dots, y_d)$ for all $y \in (-\varepsilon, \varepsilon) \times R$.

We have

$$
\begin{aligned}
G \circ \mathscr{T}(y) &= y_1 n\big(\psi(y_2,\dots,y_d)\big) + \psi(y_2,\dots,y_d) \\
&\quad + y_1|y_1|\phi\bigg(\frac{|y_1|}{c}\bigg)\alpha\big(\psi(y_2,\dots,y_d)\big) \\
&= y_1 n\big(\psi(y_2,\dots,y_d)\big) + \psi(y_2,\dots,y_d) \\
&\quad + \bar\phi(y_1)\alpha\big(\psi(y_2,\dots,y_d)\big),
\end{aligned}
$$

where $\bar\phi = y|y|\phi(\frac{y}{c})$, and thus

$$
\begin{aligned}
\frac{\partial(G\circ\mathscr{T})}{\partial y_1}(y) &= n\big(\psi(y_2,\dots,y_d)\big) + \bar\phi'(y_1)\alpha\big(\psi(y_2,\dots,y_d)\big), \quad \text{and} \\
\frac{\partial(G\circ\mathscr{T})}{\partial y_j}(y) &= y_1\frac{\partial(n\circ\psi)}{\partial y_j}(y_2,\dots,y_d) + \frac{\partial\psi}{\partial y_j}(y_2,\dots,y_d) \\
&\quad + \bar\phi(y_1)\frac{\partial(\alpha\circ\psi)}{\partial y_j}(y_2,\dots,y_d).
\end{aligned}
$$

Now note that

$$
\begin{aligned}
\frac{\partial(G\circ\mathscr{T})}{\partial y_1}(y) &= G'\big(\mathscr{T}(y)\big)\frac{\partial\mathscr{T}}{\partial y_1}(y) = G'\big(\mathscr{T}(y)\big)n\big(\psi(y_2,\dots,y_d)\big), \quad \text{and} \\
\frac{\partial(G\circ\mathscr{T})}{\partial y_j}(y) &= G'\big(\mathscr{T}(y)\big)\frac{\partial\mathscr{T}}{\partial y_j}(y) \\
&= G'\big(\mathscr{T}(y)\big)\bigg(y_1\frac{\partial(n\circ\psi)}{\partial y_j}(y_2,\dots,y_d) + \frac{\partial\psi}{\partial y_j}(y_2,\dots,y_d)\bigg)
\end{aligned}
$$

for all $j \neq 1$. Further,

$$
\begin{aligned}
\frac{\partial(n\circ\psi)}{\partial y_j}(y_2,\dots,y_d) &= n'\big(\psi(y_2,\dots,y_d)\big)\frac{\partial\psi}{\partial y_j}(y_2,\dots,y_d), \quad \text{and} \\
\frac{\partial(\alpha\circ\psi)}{\partial y_j}(y_2,\dots,y_d) &= \alpha'\big(\psi(y_2,\dots,y_d)\big)\frac{\partial\psi}{\partial y_j}(y_2,\dots,y_d).
\end{aligned}
$$

Recall that for any $\xi\in\Theta$, we have that $n'(\xi)$ and $\alpha'(\xi)$ are linear mappings from the tangent space of Θ in ξ into the $\mathbb{R}^d$. For $\xi = \psi(y_2,\dots,y_d)$, it then follows that

$$
\begin{aligned}
&G'\big(\mathscr{T}(y)\big)\big(\mathrm{id}_{\tau(\xi)} + y_1 n'(\xi)\big)\frac{\partial\psi}{\partial y_j}(y_2,\dots,y_d) \\
&\qquad = \big(\mathrm{id}_{\tau(\xi)} + y_1 n'(\xi) + \bar\phi(y_1)\alpha'(\xi)\big)\frac{\partial\psi}{\partial y_j}(y_2,\dots,y_d).
\end{aligned}
$$

Since this equation holds for all $\frac{\partial\psi}{\partial y_j}$, $j = 2,\dots,d$, it also holds for every vector b in the tangent space, that is,

$$
G'\big(\mathscr{T}(y)\big)\big(\mathrm{id}_{\tau(\xi)} + y_1 n'(\xi)\big)b = \big(\mathrm{id}_{\tau(\xi)} + y_1 n'(\xi) + \bar\phi(y_1)\alpha'(\xi)\big)b.
$$

For $|y_1| \leq \varepsilon$, the mapping $\mathrm{id}_{\tau(\xi)} + y_1 n'(\xi)$ is invertible by the argument from Remark 3.16. Denote the inverse of $\mathrm{id}_{\tau(\xi)} + y_1 n'(\xi)$ by $\mathcal{I}_\xi(y)$.

Then for any $b \in \tau(\xi)$, we can write $b = (\mathrm{id}_{\tau(\xi)} + y_1 n'(\xi)) b_1$ with $b_1 = \mathcal{I}_\xi(y) b$ and, therefore,

$$G'(\mathscr{T}(y))b = b + \bar{\phi}(y_1)\alpha'(\xi)\mathcal{I}_\xi(y)b.$$

For a general vector $b \in \mathbb{R}^d$, we have that $(b \cdot n)n = nn^\top b$ is orthogonal to the tangent space and $(\mathrm{id}_{\mathbb{R}^d} - nn^\top)b$ is in the tangent space.

We abbreviate $G' = G'(\tilde{x})$, $p = p(\tilde{x})$, $d = \|\tilde{x} - p(\tilde{x})\|$, $n = n(p(\tilde{x}))$, $n' = n'(p(\tilde{x}))$, $\mathcal{I}_\xi = \mathcal{I}_\xi(\mathscr{T}^{-1}(\tilde{x}))$. Then we have for $b \in \mathbb{R}^d$:

$$\begin{aligned}
G'b &= G'\big((b\cdot n)n + \big(b - (b\cdot n)n\big)\big)\\
&= (b\cdot n)G'n + G'\big(b - (b\cdot n)n\big)\\
&= (b\cdot n)\big(n + \bar{\phi}'(d)\alpha(p)\big) + \big(b - (b\cdot n)n\big) + \bar{\phi}(d)\alpha'(p)\mathcal{I}_\xi\big(b - (b\cdot n)n\big)\\
&= b + \bar{\phi}'(d)\alpha(p)n^\top b + \bar{\phi}(d)\alpha'(p)\mathcal{I}_\xi\big(\mathrm{id}_{\mathbb{R}^d} - nn^\top\big)b.
\end{aligned}$$

Therefore,

$$G' = \mathrm{id}_{\mathbb{R}^d} + \bar{\phi}'(d)\alpha(p)n^\top + \bar{\phi}(d)\alpha'(p)\mathcal{I}_\xi\big(\mathrm{id}_{\mathbb{R}^d} - nn^\top\big),$$

or, more explicitly,

$$\begin{aligned}
G'(\tilde{x}) = {}& \mathrm{id}_{\mathbb{R}^d} + \bar{\phi}'\big(\|\tilde{x} - p(\tilde{x})\|\big)\alpha\big(p(\tilde{x})\big)n\big(p(\tilde{x})\big)^\top\\
&+ \bar{\phi}\big(\|\tilde{x} - p(\tilde{x})\|\big)\alpha'\big(p(\tilde{x})\big)\mathcal{I}_\xi\big(\mathscr{T}^{-1}(\tilde{x})\big)\big(\mathrm{id}_{\mathbb{R}^d} - n\big(p(\tilde{x})\big)n\big(p(\tilde{x})\big)^\top\big).
\end{aligned} \tag{7}$$

In order to apply Hadamard's global inverse function theorem [16], Theorem 2.2, and thus to show that G is a diffeomorphism $\mathbb{R}^d \longrightarrow \mathbb{R}^d$, we need to show that G is C^1, $G'(x)$ is invertible for all $x \in \mathbb{R}^d$, and $\lim_{\|x\|\to\infty} \|G(x)\| = \infty$.

We have already proven differentiability of G in $\tilde{x}$. If c is sufficiently small, $G'(\tilde{x})$ is invertible, since $\bar{\phi}'$ and $\bar{\phi}$ are uniformly bounded with a bound that tends to 0 for $c \to 0$. For c small enough, it is therefore guaranteed that $G'(\tilde{x})$ is close to the identity and, therefore, invertible by Lemma 3.17. We show in the separate Lemma 3.18 that $c > 0$ can be chosen uniformly for all $\tilde{x}$ such that $G'(\tilde{x})$ is invertible.

Since $G(x) = x + \bar{\phi}(x)\alpha(x)$ and both $\bar{\phi}$ and α are bounded by the definition of $\bar{\phi}$ and Assumption 3.4, we also have the third requirement of Hadamard's global inverse function theorem. G is therefore a diffeomorphism. $\square$

We will see that c can always be chosen sufficiently small in the proof of Theorem 3.14.

LEMMA 3.18. *Fix $\varkappa > 1$ and define $A = \frac{256(\varkappa-1)}{27\varkappa(d-1)d}$.*

Let $c < \min(\frac{\varepsilon_0}{\varkappa \max(K,1)}, \min_{1\le i,j\le d} b_{i,j})$ *where for* $i, j \in \{1,\dots,d\}$ *we define*

$$b_{i,j} := \begin{cases} \dfrac{1}{2d^2|\alpha_i(p(x))|} & \textit{if } \dfrac{\partial\alpha_i(p(x))}{\partial x_j} = 0, \\ -Ad\left|\dfrac{\alpha_i(p(x))}{\frac{\partial\alpha_i(p(x))}{\partial x_j}}\right| & \\ \quad + \sqrt{A^2d^4\left|\dfrac{\alpha_i(p(x))}{\frac{\partial\alpha_i(p(x))}{\partial x_j}}\right|^2 + \dfrac{A}{|\frac{\partial\alpha_i(p(x))}{\partial x_j}|}} & \textit{if } \dfrac{\partial\alpha_i(p(x))}{\partial x_j} \neq 0. \end{cases}$$

With this choice of c, *we have that* $G'(x)$ *is invertible for every* $x \in \mathbb{R}^d$.

PROOF. Recall equation (7) from the proof of Theorem 3.14:

$$\begin{aligned} G'(x) &= \mathrm{id}_{\mathbb{R}^d} + \bar\phi'(\|x-p(x)\|)\alpha(p(x))n(p(x))^\top \\ &\quad + \bar\phi(\|x-p(x)\|)\alpha'(p(x))\mathcal{I}_\xi(\mathscr{T}^{-1}(x))(\mathrm{id}_{\mathbb{R}^d} - n(p(x))n(p(x))^\top) \\ &=: 1 + \mathcal{A}. \end{aligned}$$

We begin by estimating the operator norm of $\mathcal{A}$:

$$\begin{aligned} \|\mathcal{A}\| &\le \sum_{i=1}^d \|\bar\phi'(\|x-p(x)\|)\| \sum_{j=1}^d |\alpha_i(p(x))n_j(p(x))| \\ &\quad + d(d-1)\left|\frac{\partial\alpha_i(p(x))}{\partial x_j}\right| |\bar\phi(\|x-p(x)\|)| \|\mathcal{I}_\xi\| \|\mathrm{id}_{\mathbb{R}^d} - n(p(x))n(p(x))^\top\| \\ &\le \max_{1\le i,j\le d} d^2|\alpha_i(p(x))| \|\bar\phi'\| \|x-p(x)\| \\ &\quad + d(d-1)\left|\frac{\partial\alpha_i(p(x))}{\partial x_j}\right| |\bar\phi(\|x-p(x)\|)| \|\mathcal{I}_\xi\| \|\mathrm{id}_{\mathbb{R}^d} - n(p(x))n(p(x))^\top\| \\ &\le \max_{1\le i,j\le d} 2cd^2|\alpha_i(p(x))| + \frac{27c^2}{256}d(d-1)\left|\frac{\partial\alpha_i(p(x))}{\partial x_j}\right| \frac{1}{1+|y_1|\|n'\|}, \end{aligned}$$

where we used that $\|x-p(x)\| \le c$ and $\|\bar\phi'\| \le 2$ for $x \in \Theta^c$, $|\bar\phi(\|x-p(x)\|)|$ attains its maximum in $\frac{27c^2}{256}$, and $\|\mathrm{id}_{\mathbb{R}^d} - n(p(x))n(p(x))^\top\| \le 1$. Furthermore, $\|\mathcal{I}_\xi\| \le \frac{1}{1+|y_1|\|n'\|}$, since $\|y_1 n'\| < \frac{1}{\varkappa}$ by Lemma 3.17 and Remark 3.16. Hence,

$$\frac{1}{1+|y_1|\|n'\|} \le \frac{\varkappa}{\varkappa - 1}.$$

To complete the proof, we have to solve the quadratic inequality:

$$\frac{27c^2\varkappa d(d-1)}{256(\varkappa-1)}\left|\frac{\partial\alpha_i(p(x))}{\partial x_j}\right| + 2cd^2|\alpha_i(p(x))| < 1$$

in c to get the second upper bound for c, such that $G'(x)$ is invertible for $x \in \Theta^c$ by Lemma 3.17. For $x \in \mathbb{R}^d \setminus \Theta^c$, $G'(x) = \mathrm{id}_{\mathbb{R}^d}$. $\square$

W.l.o.g. we always choose c like in Lemma 3.18.

We proceed with proving that, although $G \notin C^2$, Itô's formula holds for G and G^{-1}.

THEOREM 3.19. *Let Assumptions 3.1–3.4 be satisfied. Then Itô's formula holds for G and G^{-1}.*

PROOF. If $x \in \mathbb{R}^d \setminus \Theta$, then since $G, G^{-1} \in C^2$ on $\mathbb{R}^d \setminus \Theta$, Itô's formula holds for G and G^{-1} until the first time X hits Θ. So the only interesting case is $x \in \Theta$.

For this, there exists an open rectangle $R \in \mathbb{R}^{d-1}$ and a local parametrization $\psi : R \longrightarrow \mathbb{R}^d$ of Θ. Let $B = \psi(R)$. Moreover,

$$U = \{y_1 n(\psi(y_2, \dots, y_d)) + \psi(y_2, \dots, y_d) : y_1 \in (-\varepsilon, \varepsilon), (y_2, \dots, y_d) \in R\}.$$

Let $\mathscr{T} : (-\varepsilon, \varepsilon) \times R \longrightarrow U$ be defined as in the proof of Theorem 3.14. Note that $\mathscr{T} \in C^2$, because Θ is C^3 by Assumption 3.1, so Itô's formula holds for $\mathscr{T}$. $\mathscr{T}$ is locally invertible with $\det \mathscr{T}' \neq 0$, so $\mathscr{T}^{-1} \in C^2$ as well. If we can show that Itô's formula holds for $G \circ \mathscr{T}$, then it also holds for $G = G \circ \mathscr{T} \circ \mathscr{T}^{-1}$.

$G \circ \mathscr{T}$ fits the assumptions of [15], Theorem 2.9, (we get boundedness of the derivatives by localizing to a bounded domain), so Itô's formula holds for $G \circ \mathscr{T}$ and, therefore, also for G.

$\tilde{G} = \mathscr{T}^{-1} \circ G \circ \mathscr{T}$ is a function with continuous first and second derivatives, with the sole exception of $\frac{\partial^2 \tilde{G}}{\partial y_1^2}$, which is bounded, but may be discontinuous for $y_1 = 0$. Since $\det \tilde{G}' \neq 0$ on an environment of x, this property transfers to the inverse, which is $\tilde{G}^{-1} = \mathscr{T}^{-1} \circ G^{-1} \circ \mathscr{T}$. Thus, again by [15], Theorem 2.9, Itô's formula holds for $\tilde{G}^{-1}$, and a fortiori for G^{-1}. $\square$

Now we are ready to show that the coefficients of the transformed SDE for $G(X)$ are Lipschitz.

ASSUMPTION 3.5. We assume the following for μ and σ:

1. the diffusion coefficient σ is Lipschitz;
2. μ and σ are bounded on Θ^ε.

THEOREM 3.20. *Let Assumptions 3.1–3.5 be satisfied. Then the SDE for $G(X)$ has Lipschitz coefficients.*

PROOF. We first show that the drift of $G(X)$ is continuous in Θ. Let B, R, ψ and $\mathscr{T}$ be defined as in the proof of Theorem 3.14. Suppose now, we have a locally defined process X in U. Then there exists a locally defined process Y in $(-\varepsilon,\varepsilon)\times R$ with

$$X = Y_1 n\big(\psi(Y_2,\dots,Y_d)\big) + \psi(Y_2,\dots,Y_d),$$

that is, $X=\mathcal{T}(Y)$.

If Y is a locally defined solution to $dY=\nu(Y)\,dt+\omega(Y)\,dW$, then by Itô's formula,

$$dX = \mathcal{T}'(Y)\nu(Y)\,dt + \mathcal{T}'(Y)\omega(Y)\,dW + \frac{1}{2}\operatorname{tr}\big(\omega^\top(Y)\mathcal{T}''(Y)\omega(Y)\big)\,dt,$$

where $\mathcal{T}'$ and $\mathcal{T}''$ denote the Jacobian and the Hessian of $\mathcal{T}$, and tr denotes the trace of a matrix. We want $\mathcal{T}'\omega=\sigma$, or more precisely $\mathcal{T}'(Y)\omega(Y)=\sigma(\mathcal{T}(Y))$, that is, $\omega=(\mathcal{T}')^{-1}\sigma$. For brevity, write $\mathcal{S}=\mathcal{T}^{-1}$. Now

$$\big(\omega\omega^\top\big)_{1,1} = \omega_{1,1}^2+\cdots+\omega_{1,d}^2 = e_1^\top\omega\omega^\top e_1 = e_1^\top\big(\mathcal{S}'\sigma\sigma^\top\big(\mathcal{S}'\big)^\top\big)e_1.$$

We show that $(\mathcal{S}')^\top e_1=n$. It is not hard to see that the Jacobian $\mathcal{T}'$ of $\mathcal{T}$ in a point $\xi\in\Theta$ is given by

$$\mathcal{T}' = \left(n, \frac{\partial\psi}{\partial y_2},\dots,\frac{\partial\psi}{\partial y_d}\right),$$

such that

$$\begin{aligned} & e_1^\top\big(\mathcal{T}'\big)^{-1} = e_1^\top\big(\big(\mathcal{T}'\big)^{-1}\big) = n^\top \\ \iff\quad & e_1^\top = n^\top\mathcal{T}' = n^\top\left(n, \frac{\partial\psi}{\partial y_2},\dots,\frac{\partial\psi}{\partial y_d}\right) = \big(\|n\|^2,0,\dots,0\big) = e_1^\top. \end{aligned}$$

Therefore, we have $\omega_{1,1}^2+\cdots+\omega_{1,d}^2 = n^\top\sigma\sigma^\top n$ on Θ.

The drift coefficient ν of the SDE for Y has only discontinuities in the set $\{y\in\mathbb{R}^d : y_1=0\}$. Further,

$$\begin{aligned} dY &= d\big(\mathcal{S}(X)\big) \\ &= \mathcal{S}'(X)\mu(X)\,dt + \mathcal{S}'(X)\sigma(X)\,dW + \frac{1}{2}\operatorname{tr}\big(\sigma^\top(X)\mathcal{S}''(X)\sigma(X)\big)\,dt, \end{aligned}$$

that is, $\nu(y)=\mathcal{S}'(\mathcal{T}(y))\mu(\mathcal{T}(y))+\frac{1}{2}\operatorname{tr}(\sigma^\top(\mathcal{T}(y))\mathcal{S}''(\mathcal{T}(y))\sigma(\mathcal{T}(y)))$. The second term is continuous, so that

$$\begin{aligned} &\lim_{h\to 0}\big(\nu(-h,y_2,\dots,y_d)-\nu(h,y_2,\dots,y_d)\big) \\ &\quad = \mathcal{S}'\big(\mathcal{T}(0,y_2,\dots,y_d)\big) \\ &\qquad \times \lim_{h\to 0}\big(\mu\big(\mathcal{T}(-h,y_2,\dots,y_d)\big)-\mu\big(\mathcal{T}(h,y_2,\dots,y_d)\big)\big) \\ &\quad = \mathcal{S}'\big(\mathcal{T}(0,y_2,\dots,y_d)\big) \end{aligned}$$

$$
\begin{aligned}
&\times \lim_{h\to 0}\big(\mu\big(\mathscr{T}(0,y_2,\dots,y_d)-hn\big(\mathscr{T}(0,y_2,\dots,y_d)\big)\big)\\
&\quad-\mu\big(\mathscr{T}(0,y_2,\dots,y_d)+hn\big(\mathscr{T}(0,y_2,\dots,y_d)\big)\big)\big)\\
&=\mathscr{S}'\big(\mathscr{T}(0,y_2,\dots,y_d)\big)2\alpha\big(\mathscr{T}(0,y_2,\dots,y_d)\big)\\
&\quad\times \big(n^\top\sigma\sigma^\top n\big)\big(\mathscr{T}(0,y_2,\dots,y_d)\big)\\
&=\mathscr{S}'\big(\mathscr{T}(0,y_2,\dots,y_d)\big)2\alpha\big(\mathscr{T}(0,y_2,\dots,y_d)\big)\\
&\quad\times \big(\omega\omega^\top\big)_{11}(0,y_2,\dots,y_d).
\end{aligned}
\tag{8}
$$

Consider

$$
\begin{aligned}
(G\circ\mathscr{T})(y)&=\mathscr{T}(y)+\tilde\phi\big(\mathscr{T}(y)\big)\alpha\big(p\big(\mathscr{T}(y)\big)\big)\\
&=y_1n\big(\mathscr{T}(0,y_2,\dots,y_d)\big)+\mathscr{T}(0,y_2,\dots,y_d)\\
&\quad+y_1|y_1|\phi\left(\frac{y_1}{c}\right)\alpha\big(\mathscr{T}(0,y_2,\dots,y_d)\big),
\end{aligned}
$$

and

$$
\begin{aligned}
(\mathscr{S}\circ G\circ\mathscr{T})(y)=\mathscr{S}\bigg(&y_1n\big(\mathscr{T}(0,y_2,\dots,y_d)\big)+\mathscr{T}(0,y_2,\dots,y_d)\\
&+y_1|y_1|\phi\left(\frac{y_1}{c}\right)\alpha\big(\mathscr{T}(0,y_2,\dots,y_d)\big)\bigg).
\end{aligned}
$$

Differentiation yields

$$
\begin{aligned}
&\frac{\partial}{\partial y_1}(\mathscr{S}\circ G\circ\mathscr{T})(y)\\
&\quad=\mathscr{S}'\big((G\circ\mathscr{T})(y)\big)\frac{\partial}{\partial y_1}\bigg(y_1n\big(\mathscr{T}(0,y_2,\dots,y_d)\big)+\mathscr{T}(0,y_2,\dots,y_d)\\
&\qquad+y_1|y_1|\phi\left(\frac{y_1}{c}\right)\alpha\big(\mathscr{T}(0,y_2,\dots,y_d)\big)\bigg)\\
&\quad=\mathscr{S}'\big((G\circ\mathscr{T})(y)\big)\bigg(n\big(\mathscr{T}(0,y_2,\dots,y_d)\big)\\
&\qquad+\left(2|y_1|\phi\left(\frac{y_1}{c}\right)+y_1|y_1|\phi'\left(\frac{y_1}{c}\right)\frac{1}{c}\right)\alpha\big(\mathscr{T}(0,y_2,\dots,y_d)\big)\bigg).
\end{aligned}
$$

We look at the second derivative w.r.t. y_1:

$$
\begin{aligned}
&\frac{\partial^2}{\partial y_1^2}(\mathscr{S}\circ G\circ\mathscr{T})(y)\\
&\qquad=\text{something continuous}\\
&\qquad\quad+\mathscr{S}'\big((G\circ\mathscr{T})(y)\big)\left(2\operatorname{sign}(y_1)\phi\left(\frac{y_1}{c}\right)\alpha\big(\mathscr{T}(0,y_2,\dots,y_d)\big)\right).
\end{aligned}
$$

Since $G(x)=x$ for $x\in\Theta$, we have that $G(\mathscr{T}(y))=\mathscr{T}(y)$ for $y_1=0$, and thus

$$
\begin{aligned}
&\lim_{h\to 0+}\Bigl(\frac{\partial^2}{\partial y_1^2}(\mathscr{S}\circ G\circ\mathscr{T})(-h,y_2,\dots,y_d)\\
&\qquad -\frac{\partial^2}{\partial y_1^2}(\mathscr{S}\circ G\circ\mathscr{T})(h,y_2,\dots,y_d)\Bigr)\\
&\quad=-4\mathscr{S}'\bigl((G\circ\mathscr{T})(0,y_2,\dots,y_d)\bigr)\alpha\bigl(\mathscr{T}(0,y_2,\dots,y_d)\bigr)\\
&\quad=-4\mathscr{S}'\bigl(\mathscr{T}(0,y_2,\dots,y_d)\bigr)\alpha\bigl(\mathscr{T}(0,y_2,\dots,y_d)\bigr).
\end{aligned}
\tag{9}
$$

Consider the drift coefficient of $(\mathscr{S}\circ G\circ\mathscr{T})_k(Y)$, which is

$$
\begin{aligned}
\tilde\nu_k(y):=&\sum_{j=1}^d\frac{\partial}{\partial y_j}(\mathscr{S}\circ G\circ\mathscr{T})_k(y)\nu_j(y)\\
&+\frac12\sum_{i,j=1}^d\frac{\partial^2}{\partial y_i\,\partial y_j}(\mathscr{S}\circ G\circ\mathscr{T})_k(y)\sum_{l=1}^d\omega_{li}(y)\omega_{lj}(y).
\end{aligned}
$$

$(\mathscr{S}\circ G\circ\mathscr{T})'(0,y_2,\dots,y_d)=\mathrm{id}_{\mathbb{R}^d}$, thus $\frac{\partial}{\partial y_j}(\mathscr{S}\circ G\circ\mathscr{T})_k(0,y_2,\dots,y_d)=(e_k)_j$. Further, note that $\frac{\partial^2}{\partial y_i\,\partial y_j}(\mathscr{S}\circ G\circ\mathscr{T})_k$ is continuous for all pairs (i,j) except $(i,j)=(1,1)$.

Thus, using (8) and (9), we have

$$
\begin{aligned}
&\lim_{h\to 0+}\bigl(\tilde\nu_k(-h,y_2,\dots,y_d)-\tilde\nu_k(h,y_2,\dots,y_d)\bigr)\\
&\quad=\lim_{h\to 0+}\Bigl(\nu_k(-h,y_2,\dots,y_d)+\frac12\frac{\partial^2}{\partial y_1^2}(\mathscr{S}\circ G\circ\mathscr{T})_k\\
&\qquad\times(-h,y_2,\dots,y_d)\bigl(\omega\omega^\top\bigr)_{11}(0,y_2,\dots,y_d)\\
&\qquad-\nu_k(h,y_2,\dots,y_d)-\frac12\frac{\partial^2}{\partial y_1^2}(\mathscr{S}\circ G\circ\mathscr{T})_k\\
&\qquad\times(h,y_2,\dots,y_d)\bigl(\omega\omega^\top\bigr)_{11}(0,y_2,\dots,y_d)\Bigr)\\
&\quad=\mathscr{S}'\bigl(\mathscr{T}(0,y_2,\dots,y_d)\bigr)2\alpha\bigl(\mathscr{T}(0,y_2,\dots,y_d)\bigr)\bigl(\omega\omega^\top\bigr)_{11}(0,y_2,\dots,y_d)\\
&\qquad-2\mathscr{S}'\bigl(\mathscr{T}(0,y_2,\dots,y_d)\bigr)\alpha\bigl(\mathscr{T}(0,y_2,\dots,y_d)\bigr)\bigl(\omega\omega^\top\bigr)_{11}(0,y_2,\dots,y_d)\\
&\quad=0.
\end{aligned}
$$

Therefore, $\tilde\nu$ is continuous on the whole of $\mathbb{R}^d$.

Now the drift coefficient of the SDE for the process $G(X)$ is continuous as well: $G(X)=\mathscr{T}\circ(\mathscr{S}\circ G\circ\mathscr{T})\circ\mathscr{S}(X)$ and compounding with $\mathscr{T}$ and $\mathscr{S}$ preserves continuity of the drift since $\mathscr{T},\mathscr{S}\in C^2$.

The kth coordinate of the transformed drift $\tilde{\mu}$ has the form:

$$\tilde{\mu}_k(z) = G_k'\big(G^{-1}(z)\big)\mu\big(G^{-1}(z)\big) + \frac{1}{2}\operatorname{tr}\big(\sigma^\top\big(G^{-1}(z)\big)G_k''\big(G^{-1}(z)\big)\sigma\big(G^{-1}(z)\big)\big)$$

and we have just seen that it is continuous in all $z \in \Theta$. It remains to show that $\tilde{\mu}$ is intrinsic Lipschitz on $\mathbb{R}^d \setminus \Theta$. For $z \in \mathbb{R}^d \setminus \Theta^c$, we have $\tilde{\mu}(z) = \mu(z)$. μ is intrinsic Lipschitz on $\mathbb{R}^d \setminus \Theta$ and, therefore, also on $\mathbb{R}^d \setminus \Theta^c$.

On $\Theta^c \setminus \Theta$, we have that G' is differentiable with bounded derivative and is therefore intrinsic Lipschitz by Lemma 3.8. μ is intrinsic Lipschitz on $\mathbb{R}^d \setminus \Theta$ by Assumption 3.1 and μ is bounded on Θ^c by Assumption 3.5, item 2. Moreover, G^{-1} is Lipschitz on $\mathbb{R}^d$, and thus the mapping $x \mapsto G_k'(G^{-1}(z))\mu(G^{-1}(z))$ is intrinsic Lipschitz by Lemma 3.9.

In the same way, we see that G'' is differentiable with bounded derivative on $\Theta^c \setminus \Theta$ and is therefore intrinsic Lipschitz by Lemma 3.8. σ is Lipschitz on $\mathbb{R}^d$ and, therefore, intrinsic Lipschitz on $\Theta^c \setminus \Theta$. Moreover, both G'' and σ are bounded on $\Theta^c \setminus \Theta$, thus $z \mapsto \frac{1}{2}\operatorname{tr}(\sigma^\top(G^{-1}(z))G_k''(G^{-1}(z))\sigma(G^{-1}(z)))$ is intrinsic Lipschitz by Lemma 3.9.

Now $\tilde{\mu}$ is intrinsic Lipschitz as a sum of intrinsic Lipschitz functions.

Altogether we have shown that $\tilde{\mu}$ is piecewise Lipschitz and continuous, and hence Lipschitz by Lemma 3.6 and Lemma 3.11.

The transformed diffusion coefficient is given by

$$\tilde{\sigma}(z) = G'\big(G^{-1}(z)\big)\sigma\big(G^{-1}(z)\big).$$

Since G^{-1}, G' and σ are Lipschitz, the mappings $z \mapsto G'(G^{-1}(z))$ and $z \mapsto \sigma(G^{-1}(z))$ are Lipschitz. Moreover, they are both bounded on Θ^ε (and thus on Θ^c), such that their product is Lipschitz. $\square$

3.5. *Main results.* Finally, we are ready to prove the two main results of this paper.

For this, define

$$dZ = dG(X) = \tilde{\mu}(Z)\,dt + \tilde{\sigma}(Z)\,dW, \qquad Z_0 = G(x), \tag{10}$$

where $\tilde{\mu}$ and $\tilde{\sigma}$ are defined in the proof of Theorem 3.20.

THEOREM 3.21. *Let Assumptions 3.1–3.5 be satisfied. Then the d-dimensional SDE (1) has a unique global strong solution.*

PROOF. Since by Theorem 3.20 SDE (10) has Lipschitz coefficients, it follows that it has a unique global strong solution for the initial value $G(x)$. Due to Theorem 3.14, the transformation G has a global inverse G^{-1}. Itô's formula holds for G^{-1} by Theorem 3.19. Applying Itô's formula to G^{-1}, we obtain that $G^{-1}(Z)$ satisfies

$$dX = \mu(X)\,dt + \sigma(X)\,dW, \qquad X_0 = x.$$

Setting $X = G^{-1}(Z)$ closes the proof. $\square$

For calculating the solution to the d-dimensional SDE (1), the same algorithm as for the one-dimensional case works, if applied using the transformations from the d-dimensional case. Let $Z_T^{(\delta)}$ be the Euler–Maruyama approximation of the solution to SDE (10) with step size smaller than $\delta > 0$.

ALGORITHM 3.22. Go through the following steps:

1. Set $Z_0^{(\delta)} = G(x)$.
2. Apply the Euler–Maruyama method to SDE (10) to obtain $Z_T^{(\delta)}$.
3. Set $\bar{X} = G^{-1}(Z_T^{(\delta)})$.

THEOREM 3.23. *Let Assumptions 3.1–3.5 be satisfied.*
Then Algorithm 3.22 converges with strong order $1/2$ *to the solution* X *of the* d*-dimensional SDE* (1).

PROOF. We estimate the L^2-error of the approximation. For every $T > 0$, there is a constant C, such that

$$\begin{aligned}\mathbb{E}(\|X_T - \bar{X}_T\|^2) &= \mathbb{E}(\|G^{-1}(Z_T) - G^{-1}(Z_T^{(\delta)})\|^2)\\ &\le L_{G^{-1}}^2 \mathbb{E}(\|Z_T - Z_T^{(\delta)}\|^2) = L_{G^{-1}}^2 C\delta\end{aligned}$$

for every sufficiently small step size δ, where $L_{G^{-1}}$ is the Lipschitz constant of G^{-1}. We used [10], Theorem 10.2.2, for the L^2-convergence of order $1/2$ of the Euler–Maruyama scheme for SDEs with Lipschitz coefficients. □

3.6. *Compact set of discontinuities*. To be able to prove our main results, we had to make a number of assumptions on the coefficient functions μ and σ. At least one of those is indispensable for our method to work, that is, Assumption 3.1, which demands that μ is piecewise Lipschitz and that its set of discontinuities Θ is a C^3 hypersurface.

There are two more assumptions on Θ and several on the behaviour of the coefficients close to Θ. In this subsection, we shall find out which assumptions are automatically satisfied in the case where Θ is compact.

For compact Θ, Assumption 3.2 is also automatically satisfied. This follows from a lemma in [4].

LEMMA 3.24. *Let* $\Theta \subseteq \mathbb{R}^d$ *be a compact* C^k *submanifold with* $k \ge 2$.
Then Θ *has a neighbourhood* $U = \Theta^\varepsilon$ *with the unique closest point property, and the projection map* $p : U \longrightarrow \Theta$ *is* C^{k-1}.

Assumption 3.3 prescribes a certain geometrical relation between Θ and directions of the diffusion coefficient. This will not be satisfied automatically only from making additional assumptions on Θ, of course. But for the case of compact Θ, Assumption 3.3 follows easily from weaker requirements on σ.

PROPOSITION 3.25. *Let Θ be a compact C^2 hypersurface and let $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be Lipschitz.*

If $\sigma(\xi)^\top n(\xi) \neq 0$ for all $\xi \in \Theta$, then there exists a constant $c_0 > 0$ such that $\|\sigma^\top(\xi)n(\xi)\| \geq c_0$ for all $\xi \in \Theta$.

PROOF. Let $B \subseteq \Theta$ be a bounded, open and connected subset with the property that there exists an orthonormal vector n on B. Since $\sigma^\top n$ is continuous on the closure $\overline{B}$, there exists $c > 0$ such that $\|\sigma(\xi)^\top(\xi)\| \geq c$ for all $\xi \in B$.

By compactness, Θ can be covered by finitely many sets $B_1, \dots, B_n$ with lower bounds $c_1, \dots, c_n$ and we can take $c_0 := \min(c_1, \dots, c_n)$ for the conclusion to hold. $\square$

Note that $\sigma(\xi)^\top n(\xi) \neq 0$ also follows from $\det(\sigma(\xi)) \neq 0$. So in particular, regularity of σ implies Assumption 3.3 for compact Θ.

Finally, consider Assumption 3.4 which asserts boundedness of the first three derivatives of the locally defined function α on Θ. Similar to what we have done in the proof of Proposition 3.25, we can conclude boundedness of the derivatives from their continuity.

Assumption 3.5(2) is also automatically satisfied for compact Θ.

4. Numerical examples. In this section, we present concrete examples. We compute the transform G as well as the coefficients $\tilde{\mu}$, $\tilde{\sigma}$ of the transformed SDE to which we apply the Euler–Maruyama scheme. Furthermore, we examine the quality of the approximation by considering the estimated L^2-error.

Discontinuity on the unit circle. Let Θ be the unit circle in $\mathbb{R}^2$, that is, the drift of our SDE is discontinuous only in $\Theta = \{x \in \mathbb{R}^2 : \|x\| = 1\}$. We want to solve the following SDE numerically:

$$\begin{pmatrix} dX \\ dY \end{pmatrix} = \mu(X, Y)\,dt + \sigma(X, Y)\,dW_t, \qquad \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}, \tag{11}$$

where

$$\mu(x, y) = \begin{cases} (-x, -y)^\top, & x^2 + y^2 > 1, \\ (x, 0)^\top, & x^2 + y^2 < 1, \end{cases}$$

$\sigma \equiv \mathrm{id}_{\mathbb{R}^2}$, and W is a two-dimensional standard Brownian motion. Note that the nonparallelity condition, Assumption 3.3 is satisfied with $c_0 = 1$ (σ is even uniformly elliptic).

We have that $p(x,y)=n(x,y)=(\sqrt{x^2+y^2})^{-1}(x,y)^\top$ yielding the transform:

$$G(x,y)=\begin{cases}\left(1+\frac{(\sqrt{x^2+y^2}-1)|\sqrt{x^2+y^2}-1|}{\sqrt{x^2+y^2}}\phi\left(\frac{|1-\sqrt{x^2+y^2}|}{c}\right)\right)\begin{pmatrix}x\\y\end{pmatrix}, \\ \qquad (1+c)^2>x^2+y^2\geq 1, \\ \left(1+\frac{(\sqrt{x^2+y^2}-1)|\sqrt{x^2+y^2}-1|}{2\sqrt{x^2+y^2}}\phi\left(\frac{|1-\sqrt{x^2+y^2}|}{c}\right)\right)\begin{pmatrix}x\\y\end{pmatrix}, \\ \qquad (1-c)^2<x^2+y^2<1,\end{cases}$$

and $G=\mathrm{id}_{\mathbb{R}^2}$, if $x^2+y^2\geq(1+c)^2$, or $x^2+y^2\leq(1-c)^2$, where we have chosen $c=1/2$.

Then the drift of the transformed SDE is given by

$$\tilde\mu\big(G^{-1}(x,y)\big) = \begin{cases}\nabla G(x,y)(-x,-y)^\top+\frac{1}{2}\operatorname{tr}\big(G''(x,y)\big), & (1+c)^2>x^2+y^2\geq 1,\\ \nabla G(x,y)(x,0)^\top+\frac{1}{2}\operatorname{tr}\big(G''(x,y)\big), & (1-c)^2<x^2+y^2<1,\end{cases}$$

and $\tilde\mu(x,y)=(-x,-y)^\top$, if $x^2+y^2\geq(1+c)^2$, and $\tilde\mu(x,y)=(x,0)^\top$, if $x^2+y^2\leq(1-c)^2$. Furthermore, $\tilde\sigma(G^{-1}(x,y))=\nabla G(x,y)$. G^{-1} has to be evaluated numerically.

Figure 2 shows the deviation of the first component of G from the identity. Figure 3 shows the first component of μ, $\tilde\mu$, and σ_{11}, $\tilde\sigma_{11}$. All other components look similar.

We apply Algorithm 3.22 to solve SDE (11). Figure 4 shows the estimated L^2-error of the approximation of our G-transformed Euler–Maruyama method (GM),

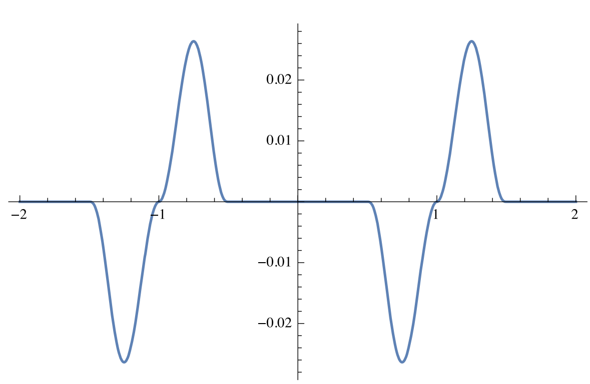

FIG. 2. *The function* $(x,y)\mapsto G_1(x,y)-x$.

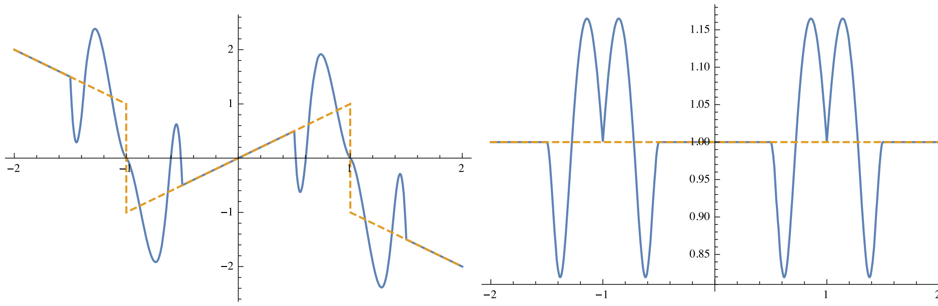

FIG. 3. *The functions* $\tilde{\mu}_1$ *and* $\tilde{\sigma}_{11}$ (*blue line*) *and* μ_1 *and* σ_{11} (*yellow dashed*).

compared to the Euler–Maruyama (EM) scheme:

$$\mathrm{err}_k := \log_2\Big(d\sqrt{\hat{E}\big(\|X_T^{(k)} - X_T^{(k-1)}\|^2\big)}\Big)$$

plotted over $\log_2 \delta^{(k)}$, where $X_T^{(k)}$ is the numerical approximation with step size $\delta = \delta^{(k)}$, $\hat{E}$ is an estimator of the mean value using 1024 paths and d is a normalizing constant so that $\mathrm{err}_1 = \sqrt{1/2}$.

We observe that our G-transformed (GM) method converges roughly with order $1/2$, and the crude Euler–Maruyama (EM) method seems to converge even at a higher rate. Note however that, even though the Euler–Maruyama method is extensively used in practice, it is not even known whether the method converges strongly for SDEs of the kind considered here. Especially we cannot conclude whether for even smaller step-size the error of the Euler–Maruyama method will still become smaller, will flatten out or whether it will even explode.

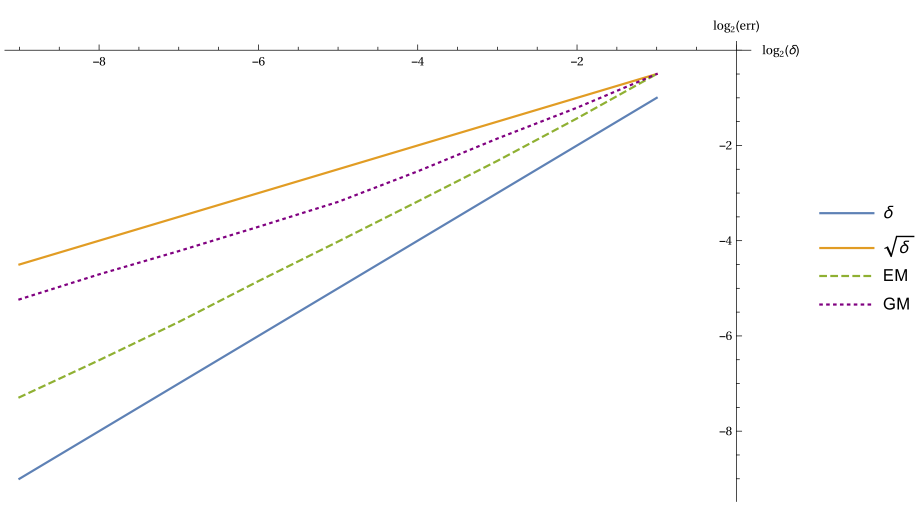

FIG. 4. *The estimated* L^2*-error for the example where* Θ *is the unit circle.*

Dividend maximization. In [20], the dividend maximization problem from actuarial mathematics, that is, the problem of maximizing the expected discounted future dividend payments until the time of ruin τ of an insurance company, is studied. In actuarial mathematics, the solution of this optimization problem serves as a risk measure. The problem is studied in a setup with incomplete information, where the drift of the underlying surplus process of the insurance company from which dividends are paid is driven by an unobservable Markov chain, the states of which represent different phases of the economy; an assumption that makes the model more realistic. In order to solve the optimization problem, the underlying surplus process has to be replaced by a multidimensional process consisting of filter probabilities of the states of the hidden Markov chain and the surplus written in terms of the filter probabilities. The resulting system is

$$\begin{aligned} dR_t &= (\bar{\alpha}_t - u_t)\,dt + \beta\,dW_t, \\ d\pi_i(t) &= \left(q_{di} + \sum_{j=1}^{d-1}(q_{ji} - q_{di})\pi_j(t)\right)dt \\ &\quad + \pi_i(t)\frac{\alpha_i - \bar{\alpha}_t}{\beta}\,dW_t, \qquad i = 1, \dots, d-1, \end{aligned} \tag{12}$$

where $\bar{\alpha}_t := \alpha_d + \sum_{i=1}^{d-1}(\alpha_i - \alpha_d)\pi_i(t)$ and where $(u_t)_{t\geq 0} \in [0, \bar{u}]$ is the dividend strategy, $R = (R_t)_{t\geq 0}$ is the surplus process and the $(\pi_i(t))_{t\geq 0}$, $i = 1, \dots, d-1$, are the conditional probabilities that the underlying hidden Markov chain is in state e_i. $W = (W_t)_{t\geq 0}$ is a one-dimensional Brownian motion. We assume knowledge of the following constants: $(q_{ij})_{i,j=1}^{d}$ are the entries of the intensity matrix of the Markov chain, β is the diffusion parameter of the surplus and α_i, $i = 1, \dots, d$, is the drift of the surplus, if the Markov chain is in state e_i.

The application of filtering theory leads to an equivalent optimization problem:

$$\sup_u \mathbb{E}_{x,\pi_1,\dots,\pi_{d-1}}\left(\int_0^\tau e^{-\delta s} u_s\,ds\right) \tag{13}$$

with discount rate $\delta > 0$. This is studied in [20] and the candidate for the optimal dividend policy is of the form $u_t^* = \bar{u}\mathbf{1}_{[b(\bar{\alpha}_t),\infty)}(R_t)$ with *threshold level* b, leading to a discontinuous drift of the surplus process from which the dividends are paid. Due to the application of filtering theory, the diffusion coefficient is not uniformly elliptic. In order to verify the admissibility of the candidate for the optimal control policy, existence and uniqueness of the underlying state process has to be proven. This can be done by applying the result presented herein and we can also simulate the optimally controlled surplus (e.g., to calculate the expected time of ruin).

And our results are even further applicable: in [20] the optimization problem (13) is solved for $d = 2$ by policy iteration in combination with solving an associated partial differential equation. Doing the same for dimension 4 or higher

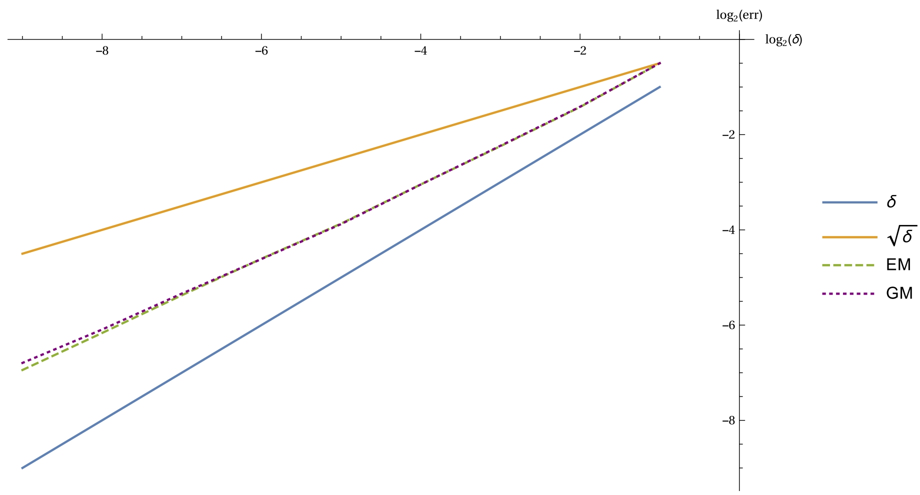

FIG. 5. *The estimated L^2-error for the example of dividend maximization.*

would not be numerically tractable. So, in higher dimension, one needs to solve the problem by combining policy iteration with simulation.

Figure 5 shows the estimated L^2-error of the approximation of the solution of (12) in dimension 5 with a linear initial threshold level. In [20] for $d = 2$, a threshold level which is a linear interpolation of the constant optimal threshold levels of the problem under full-information was used as an initial policy for policy iteration. However, we need not restrict ourselves to linear threshold levels.

Note that for our example checking whether the nonparallelity condition, Assumption 3.3 holds (in dependence on the parameter choice) is straight-forward.

We see that in this practical example the convergence order is again roughly 1/2.

Further examples from stochastic control theory, where SDEs with discontinuous (and unbounded) drift and degenerate diffusion coefficient appear are, for example, [14, 18, 19]. The SDEs appearing there can now be shown to have a unique global strong solution under conditions significantly weaker than known so far, and this solution can be approximated with a numerical method that converges with strong order 1/2. As elaborated above our method can be used for approximating solutions to these optimization problems in dimensions greater than 4, where PDE methods become practically infeasible.

Concluding remarks. In this paper, we have presented an existence and uniqueness result of strong solutions for a very general class of SDEs with discontinuous drift and degenerate diffusion coefficient; a class of SDEs that frequently appears in applications when studying stochastic optimal control problems. This is the most general result for such SDEs. Furthermore, we have derived a numerical algorithm that – under the same conditions as for the existence and uniqueness result – is proven to converge and we have established a strong convergence order of $1/2$. We have applied our algorithm to two examples: one of theoretical interest and one coming from a concrete optimal control problem in actuarial mathematics.

APPENDIX: SUPPLEMENTARY PROOFS

Proof of Lemma 3.10. Let $\xi \in \Theta$. W.l.o.g. $\xi = 0$ and $n(\xi) = e_d$, where e_d is the dth canonical basis vector of the $\mathbb{R}^d$. Thus, Θ can locally be parametrized by $\psi : R \longrightarrow \mathbb{R}^d$ of the form $\psi(y_1, \dots, y_{d-1}) = (y_1, \dots, y_{d-1}, \phi(y_1, \dots, y_{d-1}))^\top$, where $\phi : R \longrightarrow \mathbb{R}$ is a C^3-function with $\phi(0) = 0$ and $\phi'(0) = 0$. Hence, for all $y \in R$,

$$\begin{aligned}
\lambda(y) n\bigl(\psi(y)\bigr) &= \frac{\partial \psi}{\partial y_1}(y) \times \cdots \times \frac{\partial \psi}{\partial y_{d-1}}(y) \\
&= \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ \frac{\partial \phi}{\partial y_1}(y) \end{pmatrix} \times \cdots \times \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \frac{\partial \phi}{\partial y_{d-1}}(y) \end{pmatrix} = \begin{pmatrix} -\frac{\partial \phi}{\partial y_1}(y) \\ -\frac{\partial \phi}{\partial y_2}(y) \\ \vdots \\ -\frac{\partial \phi}{\partial y_{d-1}}(y) \\ 1 \end{pmatrix},
\end{aligned}$$

with $\lambda(y) = \| \frac{\partial \psi}{\partial y_1}(y) \times \cdots \times \frac{\partial \psi}{\partial y_{d-1}}(y) \|$. Note that λ is a C^2-function satisfying $\lambda'(0) = 0$ and w.l.o.g. the parametrization is chosen such that $\lambda(0) = 1$. Hence,

$$\begin{aligned}
(\lambda n \circ \psi)'(y) &= - \begin{pmatrix} \frac{\partial^2 \phi}{\partial y_1^2}(y) & \frac{\partial^2 \phi}{\partial y_2 \, \partial y_1}(y) & \cdots & \frac{\partial^2 \phi}{\partial y_{d-1} \, \partial y_1}(y) \\ \frac{\partial^2 \phi}{\partial y_1 \, \partial y_2}(y) & \frac{\partial^2 \phi}{\partial y_2^2}(y) & \cdots & \frac{\partial^2 \phi}{\partial y_{d-1} \, \partial y_2}(y) \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 \phi}{\partial y_1 \, \partial y_{d-1}}(y) & \frac{\partial^2 \phi}{\partial y_2 \, \partial y_{d-1}}(y) & \cdots & \frac{\partial^2 \phi}{\partial y_{d-1}^2}(y) \\ 0 & 0 & \cdots & 0 \end{pmatrix} \\
&= - \begin{pmatrix} H_\phi(y) \\ 0 \end{pmatrix},
\end{aligned}$$

where $H_\phi(y)$ denotes the Hessian of ϕ in $y \in R$. On the other hand, $(\lambda n \circ \psi)' = n \circ \psi \lambda' + \lambda (n \circ \psi)'$. In particular, $(n \circ \psi)'(0) = -(H_\phi(0), 0)^\top$.

Now choose any $\varepsilon > 0$ that is smaller than the reach of Θ. Then $\psi(0)$ is the unique closest point on Θ both to $\psi(0) + \varepsilon e_d$ and $\psi(0) - \varepsilon e_d$. In other words, the open balls with centers $\psi(0) + \varepsilon e_d$ and $\psi(0) - \varepsilon e_d$ contain no point of Θ. Therefore,

$$-\varepsilon\left(1 - \sqrt{1 - (\|y\|/\varepsilon)^2}\right) \le \phi(y) \le \varepsilon\left(1 - \sqrt{1 - (\|y\|/\varepsilon)^2}\right)$$

for all $y \in R$ with $\|y\| \le \varepsilon$, from which we conclude that $-\frac{\|y\|^2}{\varepsilon} \le \phi(y) \le \frac{\|y\|^2}{\varepsilon}$, for $\|y\|$ sufficiently small. In particular, we have for $j \ne k$ and sufficiently small $|h|$ that

$$-\frac{2}{\varepsilon} \le \frac{\phi(h(e_j + e_k)) - \phi(h(e_j - e_k))}{2h^2} \le \frac{2}{\varepsilon}.$$

By letting $h \to 0$ and applying de l'Hospital's rule twice, we see that

$$-\frac{2}{\varepsilon} \le \frac{\partial^2 \phi}{\partial y_j\, \partial y_k}(0) \le \frac{2}{\varepsilon}.$$

In the same way, we conclude from

$$-\frac{1}{\varepsilon} \le \frac{\phi(he_j) - \phi(0) + \phi(-he_j)}{2h^2} \le \frac{1}{\varepsilon}$$

that $-\frac{1}{\varepsilon} \le \frac{\partial^2 \phi}{\partial y_j^2}(0) \le \frac{1}{\varepsilon}$. Thus,

$$\|n'(\xi)\|^2 = \|H_\phi(0)\|^2 \le \sum_{j=1}^{d-1}\sum_{k=1}^{d-1}\left|\frac{\partial^2 \phi}{\partial y_j\, \partial y_k}(0)\right|^2 \le \sum_{j=1}^{d-1}\sum_{k=1}^{d-1}\frac{4}{\varepsilon^2} = 4\left(\frac{d-1}{\varepsilon}\right)^2,$$

that is, $\|n'\|$ is bounded by $2\frac{d-1}{\varepsilon}$. Since this holds for all $0 < \varepsilon < \operatorname{reach}(\Theta)$, we have $\|n'\| \le 2\frac{d-1}{\operatorname{reach}(\Theta)}$.

Proof of Lemma 3.11. We prove the claim that a hypersurface that satisfies Assumption 3.2 has the property that every line segment from x to y can be replaced by a continuous curve γ from x to y with $\ell(\gamma) < \|x - y\| + \eta$ where $\eta > 0$ is a given constant.

Let from now on $\varepsilon < \varepsilon_0$, where ε_0 is as in Assumption 3.2, so that in particular for every $x \in \mathbb{R}^d$ with $d(x, \Theta) \le \varepsilon$ there is a unique closest point $p(x)$ on Θ.

Denote by s the line segment from x to y and identify it with it's parameter representation $s(t) = x + t(y - x)\|y - x\|^{-1}$. Let $A := \{t \in [0, \|y - x\|] : s(t) \in \Theta\}$. For any set $S \subseteq \mathbb{R}$, denote by $H(S)$ the set of accumulation points of S.

PROPOSITION A.1. *Let $t \in H(A)$. Then $n(s(t)) \perp s'(t)$.*

PROOF. Suppose this was not the case, that is, $n(s(t)) \cdot s'(t) \ne 0$. W.l.o.g. $n(s(t)) \cdot s'(t) = C > 0$. Let $(t_j)_{j \in \mathbb{N}}$ be a sequence in A with $t_j \ne t$, $\lim_j t_j = t$. W.l.o.g, $t_j > t$ for all j, or $t_j < t$ for all j.

By Assumption 3.2, we have $(B_\varepsilon(s(t)-\varepsilon n(s(t))) \cup B_\varepsilon(s(t)+\varepsilon n(s(t)))) \cap \Theta = \varnothing$, where $B_r(z)$ denotes the open ball with midpoint z and radius r.

Suppose $t_j > t$ for all j. Then

$$\begin{aligned}
&\|s(t)+\varepsilon n(s(t))-s(t_j)\|^2 \\
&\quad = \|s(t)-s(t_j)\|^2 + 2\varepsilon(s(t)-s(t_j))\cdot n(s(t)) + \varepsilon^2\|n(s(t))\|^2 \\
&\quad = \|s(t)-s(t_j)\|^2 + 2\varepsilon(t-t_j)s'(t)\cdot n(s(t)) + \varepsilon^2 \\
&\quad = |t-t_j|^2 - 2\varepsilon|t-t_j|C + \varepsilon^2 \\
&\quad = |t-t_j|(|t-t_j| - 2\varepsilon C) + \varepsilon^2,
\end{aligned}$$

and the last expression is smaller than ε^2 for j large enough. Thus, we have found a point ξ on Θ, namely $\xi = s(t_j)$, with $\|\xi - (s(t)+\varepsilon n(s(t)))\| < \|s(t) - (s(t) + \varepsilon n(s(t)))\| = \varepsilon$. But this contradicts the fact that $s(t)$ is the point on Θ closest to $s(t)+\varepsilon n(s(t))$.

If $t_j < t$ for all j, then the same argument carries through with $s(t)+\varepsilon n(s(t))$ replaced by $s(t)-\varepsilon n(s(t))$. $\square$

Denote the tangent hyperplane on Θ in the point ξ by $\vartheta(\xi)$, that is, $\vartheta(\xi) = \xi + \tau(\xi) = \{\xi + b : b \in \tau(\xi)\}$.

PROPOSITION A.2. *For any $\xi \in \Theta$, we can find $r > 0$ such that for any $x \in \vartheta(\xi)$ with $\|x-\xi\| < r$ we have that the line segment $\overline{x-\varepsilon n(\xi), x+\varepsilon n(\xi)}$ has precisely one intersection with Θ.*

PROOF. We can locally parametrize Θ by a function on an open environment V of ξ in the tangent hyperplane $\vartheta(\xi)$. That is, there is an open interval $I \subseteq \mathbb{R}$ and a C^2-function $\hat{\psi} : V \longrightarrow I$ such that every point $z \in \{\xi + b + yn(\xi) : b \in V, y \in I\}$ can be uniquely written as $z = \xi + b + \hat{\psi}(b)n(\xi)$. Since $\xi \in \vartheta(\xi)$, and thus $\hat{\psi}(\xi) = 0$, we may assume that $I = (-\zeta, \zeta)$ for some $0 < \zeta < \varepsilon$. Choose some r such that $0 < r < \sqrt{\varepsilon^2 - (\varepsilon - \zeta)^2}$ and such that for all $x \in \vartheta(\xi)$ we have $x \in V$ whenever $\|x - \xi\| < r$.

Now if $x \in \vartheta(\xi)$ with $\|x-\xi\| < r$, then precisely one point of Θ lies on the line segment $\overline{x-\zeta n(\xi), x+\zeta n(\xi)}$. But there is no point of Θ on the line segment $\overline{x+\zeta n(\xi), x+\varepsilon n(\xi)}$, since this is entirely contained in the open ball $B_\varepsilon(\xi + \varepsilon n(\xi))$, which by the unique closest point property for $\xi + \varepsilon n(\xi)$ does not contain any point of Θ.

By the same reasoning, $\overline{x-\zeta n(\xi), x-\varepsilon n(\xi)} \cap \Theta = \varnothing$. $\square$

PROPOSITION A.3. *Let $\varepsilon_1 < \varepsilon$. Then for any $y \in \mathbb{R}^d$ there exists a point $\hat{y} \in \mathbb{R}^d$ with $d(\hat{y}, \Theta) \geq \varepsilon_1$ and $\|y - \hat{y}\| \leq \varepsilon_1$.*

PROOF. If $d(y, \Theta) \geq \varepsilon_1$, then set $\hat{y} = y$. Otherwise, there is a unique closest point $p(y) \in \Theta$. Set

$$\hat{y} = \begin{cases} p(y) + \varepsilon_1 n(p(y)), & \text{if } n(p(y)) \cdot (y - p(y)) > 0, \\ p(y) - \varepsilon_1 n(p(y)), & \text{if } n(p(y)) \cdot (y - p(y)) < 0. \end{cases}$$

Then $\|y - \hat{y}\| \leq \varepsilon_1$ is obvious, and $d(\hat{y}, \Theta) \geq \varepsilon_1$ by the unique closest point property. $\square$

We can now modify the straight line from x to y to get a continuous curve, which is not much longer than $\|y - x\|$, but has only finitely many intersections with Θ.

For what follows, let $\alpha \in (0, 1)$ and for $0 < \delta < \varepsilon$ set $\varepsilon_1 = \varepsilon - \sqrt{\varepsilon^2 - \delta^2}$.

We construct a sequence $(\gamma_k)_{k \in \mathbb{N}_0}$ of continuous curves of finite length which becomes stationary after finitely many steps, that is, there exists k_0 such that $\gamma_k = \gamma_{k_0}$ for all $k \geq k_0$.

Furthermore, γ_{k_0} will have only finitely many intersections with Θ and it will be only slightly longer than $\|x - y\|$; see (14).

Set $\gamma_0 = s$.

Step 1: *If* $H(s \cap \Theta) = \varnothing$, then set $\gamma_1 = \gamma_0$.

Otherwise, proceed as follows: According to Proposition A.3 there exists a point $\hat{y}$ with $d(\hat{y}, \Theta) \geq \varepsilon_1$ and $\|y - \hat{y}\| \leq \varepsilon_1$. Define γ_1 as the concatenation of the lines $\overline{x, \hat{y}}$ and $\overline{\hat{y}, y}$. We have $\ell(\gamma_1) \leq \|y - x\| + 2\varepsilon_1$, and there is at most one intersection of $\overline{\hat{y}, y}$, the second line segment, with Θ, due to Assumption 3.2. Set $x_1 = x$.

After step 1, we have constructed a polygonal curve γ_1 such that $\ell(\gamma_1) \leq \|y - x\| + 2\varepsilon_1$. If γ_1 has infinitely many intersections with Θ, then all but finitely many are contained in a single line segment, $s_1 = \overline{x_1, \hat{y}}$, which satisfies $\ell(s_1) = \|\hat{y} - x_1\| = \|\hat{y} - x\| = \|(y - x) + (\hat{y} - y)\| \leq \|y - x\| + \varepsilon_1$.

Now we enter an iteration procedure. Suppose that after $k \geq 1$ steps we have constructed a polygonal curve γ_k, with the properties that $\ell(\gamma_k) \leq \|y - x\| + 2k\varepsilon_1$, and such that either γ_k has finitely many intersections with Θ, or all intersections are contained in a single line segment, $s_k = \overline{x_k, \hat{y}}$, which satisfies $\ell(s_k) \leq \|y - x\| - (k - 1)(\alpha\delta - \varepsilon_1) + \varepsilon_1$.

We construct γ_{k+1} from γ_k as follows:

Step $k + 1$: *If* $H(\gamma_k \cap \Theta) = \varnothing$, then set $\gamma_{k+1} = \gamma_k$.

Otherwise, $H(\gamma_k \cap \Theta)$ is contained in the line segment $\overline{x_k, \hat{y}}$. Parametrize this segment by $s_k(t) = x_k + t\|\hat{y} - x_k\|^{-1}(\hat{y} - x_k)$, $t \in [0, \|\hat{y} - x_k\|]$ and let $H_k = H(\{t : s_k(t) \in \Theta\})$.

Set $t_k = \min H_k$, and let $n_k = n(s_k(t_k))$. If t_k is isolated from the left, or if $t_k = 0$, then set $r_k = 0$. Now consider the case where t_k is not isolated from the left. By Proposition A.1, s_k lies in the tangent hyperplane $\vartheta(s_k(t_k)) = s_k(t_k) + \tau(s_k(t_k))$ and we can find a small ball with radius $r_k > 0$ such that, for any t with $|t - t_k| < r_k$, the line segment $\overline{s_k(t) - \varepsilon_1 n_k, s_k(t) + \varepsilon_1 n_k}$ has at most one intersection with Θ, by Proposition A.2.

Consider the line segment $\overline{s_k(t_k - r_k) + \varepsilon_1 n_k, s_k(t_k + \alpha\delta) + \varepsilon_1 n_k}$.

If the intersection of this with the plane through $\hat{y}$, which is orthogonal to the line segment, is nonempty, denote the unique intersection point by z_k.

Then we construct γ_{k+1} as the concatenation of the following line segments:

- $\overline{s_k(0), s_k(t_k - r_k)}$, which by definition of t_k and r_k has only finitely many intersections with Θ;
- $\overline{s_k(t_k - r_k), s_k(t_k - r_k) + \varepsilon_1 n_k}$, which has at most one intersection with Θ by the construction of r_k and Proposition A.2;
- $\overline{s_k(t_k - r_k) + \varepsilon_1 n_k, z_k}$, which is completely contained in $B_\varepsilon(s_k(t_k) + \varepsilon n_k)$, which does not contain any point of Θ by the unique closest point property for $s_k(t_k) + \varepsilon n_k$;
- $\overline{z_k, \hat{y}}$, which has no intersection with Θ, because as $\|z_k - \hat{y}\| = \varepsilon_1$, there is no intersection strictly between z_k and $\hat{y}$, and z_k lies in the closure of $B_{\varepsilon_1}(\hat{y})$ (this is where we need Step 1);
- $\overline{\hat{y}, y}$.

In this case, the curve γ_{k+1} has only finitely many intersections with Θ and $\ell(\gamma_{k+1}) = \ell(\gamma_k) + 2\varepsilon_1 \le \|y - x\| + 2(k+1)\varepsilon_1$.

Otherwise, set $x_{k+1} = s_k(t_k + \alpha\delta) + \varepsilon_1 n_k$, and construct γ_{k+1} as the concatenation of the following line segments:

- $\overline{s_k(0), s_k(t_k - r_k)}$, which by definition of t_k and r_k has only finitely intersections with Θ;
- $\overline{s_k(t_k - r_k), s_k(t_k - r_k) + \varepsilon_1 n_k}$, which has at most one intersection with Θ by the construction of r_k and Proposition A.2;
- $\overline{s_k(t_k - r_k) + \varepsilon_1 n_k, x_{k+1}}$, which is completely contained in $B_\varepsilon(s_k(t_k) + \varepsilon n_k)$, which does not contain any point of Θ by the unique closest point property for $s_k(t_k) + \varepsilon n_k$;
- $s_{k+1} := \overline{x_{k+1}, \hat{y}}$, which still may have infinitely many intersections with Θ;
- $\overline{\hat{y}, y}$.

Again we have that $\ell(\gamma_{k+1}) \le \ell(\gamma_k) + 2\varepsilon_1 \le \|y - x\| + 2(k+1)\varepsilon_1$. Note that

$$\begin{aligned}\ell(s_{k+1})^2 &= \|x_{k+1} - \hat{y}\|^2 = \|s_k(t_k + \alpha\delta) + \varepsilon_1 n_k - \hat{y}\|^2 \\ &= \|s_k(t_k + \alpha\delta) - \hat{y}\|^2 + \varepsilon_1^2 = (\|s_k(t_k) - \hat{y}\| - \alpha\delta)^2 + \varepsilon_1^2.\end{aligned}$$

In particular, $\|x_{k+1} - \hat{y}\| \le |\|s_k(t_k) - \hat{y}\| - \alpha\delta| + \varepsilon_1 = \|s_k(t_k) - \hat{y}\| - \alpha\delta + \varepsilon_1$. Note that $\|s_k(t_k) - \hat{y}\| - \alpha\delta \ge 0$, since otherwise the line segment

$$\overline{s_k(t_k - r_k) + \varepsilon_1 n_k, s_k(t_k + \alpha\delta) + \varepsilon_1 n_k}$$

would intersect the hyperplane orthogonal to s_k and passing through $\hat{y}$.

Thus, $\|x_{k+1} - \hat{y}\| \le \|x_k - \hat{y}\| - \alpha\delta + \varepsilon_1 \le \|x - y\| - k(\alpha\delta - \varepsilon_1) + \varepsilon_1$.

After step $k+1$, we have constructed a polygonal curve γ_{k+1} such that $\ell(\gamma_{k+1}) \le \|y - x\| + 2(k+1)\varepsilon_1$. If γ_{k+1} has infinitely many intersections with Θ, then all but finitely many are contained in a single line segment, $s_{k+1} = \overline{x_{k+1}, \hat{y}}$, and $\ell(s_{k+1}) \leq \|x-y\| - k(\alpha\delta - \varepsilon_1) + \varepsilon_1$.

So finally we have constructed a sequence $(\gamma_k)_{k\in\mathbb{N}_0}$ with:

- $\ell(\gamma_k) \leq \|x-y\| + 2k\varepsilon_1$;
- γ_k either has only finitely many intersections with Θ, or all but finitely many intersections are contained in a segment of length at most $\|x-y\| - (k-1)(\alpha\delta - \varepsilon_1) + \varepsilon_1$.

Since $\delta < \varepsilon$, we have that $\varepsilon_1 = \varepsilon - \sqrt{\varepsilon^2 - \delta^2} = \varepsilon(1 - \sqrt{1 - (\frac{\delta}{\varepsilon})^2}) < \frac{\delta^2}{\varepsilon}$, such that

$$\alpha\delta - \varepsilon_1 > \delta\left(\alpha - \frac{\delta}{\varepsilon}\right) > 0.$$

With this, and since $\|x-y\| - (k-1)(\alpha\delta - \varepsilon_1) + \varepsilon_1 \geq \ell(s_k) \geq 0$, the iteration can have at most

$$k \leq 1 + \frac{\|x-y\| + \varepsilon_1}{2(\alpha\delta - \varepsilon_1)} < 1 + \frac{\|x-y\| + \varepsilon_1}{2\delta(\alpha - \frac{\delta}{\varepsilon})} < 1 + \frac{\|x-y\| + \varepsilon}{2\delta(\alpha - \frac{\delta}{\varepsilon})}$$

steps before the sequence becomes stationary, and thus there exists a k_0 such that γ_{k_0} has at most finitely many intersections with Θ.

For the length of γ_k for $k \geq k_0$, we have

$$\ell(\gamma_k) \leq \|x-y\| + 2k\varepsilon_1 \leq \|x-y\| + \left(2\delta + \frac{\|x-y\| + \varepsilon}{\alpha - \frac{\delta}{\varepsilon}}\right)\frac{\delta}{\varepsilon}. \tag{14}$$

This can be made as close to $\|x-y\|$ as we desire by making δ small. Thus, the proof is complete.

Acknowledgement. The authors thank Bert Jüttler for valuable discussions regarding questions related to differential geometry.

REFERENCES

[1] BERKAOUI, A. (2004). Euler scheme for solutions of stochastic differential equations with non-Lipschitz coefficients. *Port. Math.* (*N.S.*) **61** 461–478. MR2113559

[2] ÉTORÉ, P. and MARTINEZ, M. (2013). Exact simulation of one-dimensional stochastic differential equations involving the local time at zero of the unknown process. *Monte Carlo Methods Appl.* **19** 41–71. MR3039402

[3] ÉTORÉ, P. and MARTINEZ, M. (2014). Exact simulation for solutions of one-dimensional stochastic differential equations with discontinuous drift. *ESAIM Probab. Stat.* **18** 686–702. MR3334009

[4] FOOTE, R. L. (1984). Regularity of the distance function. *Proc. Amer. Math. Soc.* **92** 153–155. MR0749908

[5] GYÖNGY, I. (1998). A note on Euler's approximations. *Potential Anal.* **8** 205–216. MR1625576

[6] HALIDIAS, N. and KLOEDEN, P. E. (2008). A note on the Euler–Maruyama scheme for stochastic differential equations with a discontinuous monotone drift coefficient. *BIT* **48** 51–59. MR2386114

[7] HUTZENTHALER, M., JENTZEN, A. and KLOEDEN, P. E. (2012). Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. *Ann. Appl. Probab.* **22** 1611–1641. MR2985171

[8] ITÔ, K. (1951). On stochastic differential equations. *Mem. Amer. Math. Soc.* **4** 1–57.

[9] KARATZAS, I. and SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. *Graduate Texts in Mathematics* **113**. Springer, New York. MR1121940

[10] KLOEDEN, P. E. and PLATEN, E. (1992). *Numerical Solutions of Stochastic Differential Equations*. Springer, Berlin–Heidelberg.

[11] KOHATSU-HIGA, A., LEJAY, A. and YASUDA, K. (2013). Weak approximation errors for stochastic differential equations with non-regular drift. Preprint, Inria, hal-00840211.

[12] KRANTZ, S. G. and PARKS, H. R. (1981). Distance to C^k hypersurfaces. *J. Differential Equations* **40** 116–120. MR0614221

[13] LEOBACHER, G. and SZÖLGYENYI, M. (2016). A numerical method for SDEs with discontinuous drift. *BIT* **56** 151–162. MR3486457

[14] LEOBACHER, G., SZÖLGYENYI, M. and THONHAUSER, S. (2014). Bayesian dividend optimization and finite time ruin probabilities. *Stoch. Models* **30** 216–249. MR3202121

[15] LEOBACHER, G., SZÖLGYENYI, M. and THONHAUSER, S. (2015). On the existence of solutions of a class of SDEs with discontinuous drift and singular diffusion. *Electron. Commun. Probab.* **20** no. 6, 1–14.

[16] RUZHANSKY, M. and SUGIMOTO, M. (2015). On global inversion of homogeneous maps. *Bull. Math. Sci.* **5** 13–18. MR3319979

[17] SABANIS, S. (2013). A note on tamed Euler approximations. *Electron. Commun. Probab.* **18** no. 47, 1–10. MR3070913

[18] SHARDIN, A. A. and SZÖLGYENYI, M. (2016). Optimal control of an energy storage facility under a changing economic environment and partial information. *Int. J. Theor. Appl. Finance* **19** 1650026, 1–27. MR3505500

[19] SHARDIN, A. A. and WUNDERLICH, R. (2017). Partially observable stochastic optimal control problems for an energy storage. *Stochastics* **89** 280–310. MR3574704

[20] SZÖLGYENYI, M. (2016). Dividend maximization in a hidden Markov switching model. *Stat. Risk Model.* **32** 143–158.

[21] VERETENNIKOV, A. YU. (1983). Criteria for the existence of a strong solution of a stochastic equation. *Theory Probab. Appl.* **27** 441–449.

[22] VERETENNIKOV, A. Y. U. (1981). On strong solutions and explicit formulas for solutions of stochastic integral equations. *Math. USSR, Sb.* **39** 387–403.

[23] VERETENNIKOV, A. Y. U. (1984). On stochastic equations with degenerate diffusion with respect to some of the variables. *Math. USSR, Izv.* **22** 173–180.

[24] ZVONKIN, A. K. (1974). A transformation of the phase space of a diffusion process that removes the drift. *Math. USSR, Sb.* **22** 129–149.

DEPARTMENT OF FINANCIAL MATHEMATICS
AND APPLIED NUMBER THEORY
JOHANNES KEPLER UNIVERSITY LINZ
ALTENBERGERSTRASSE 69
4040 LINZ
AUSTRIA
E-MAIL: gunther.leobacher@jku.at

INSTITUTE OF STATISTICS AND MATHEMATICS
VIENNA UNIVERSITY OF ECONOMICS
AND BUSINESS
WELTHANDELSPLATZ 1
1020 VIENNA
AUSTRIA
E-MAIL: michaela.szoelgyenyi@wu.ac.at