Stochastic Lie-Bracket Approximations for Zeroth-Order Global Optimization on Manifolds

Mahmoud Abdelgalil, *Member, IEEE*, Miroslav Krstic, *Fellow, IEEE*, and Jorge I. Poveda, *Member, IEEE*

***Abstract*—We propose an algorithm for zeroth-order global optimization of smooth functions on compact manifolds. The algorithm is expressed as an ordinary differential equation driven by piecewise-smooth McShane-type approximations of the Wiener process, and requires only zeroth-order (measurement) access to the objective. We show that the limit of this ODE is a Riemannian Langevin diffusion whose unique invariant measure is a Gibbs measure that concentrates, as a design gain increases, on the set of global minimizers of the objective. We then develop the invariant measure theory of the algorithm. Specifically, by exploiting the i.i.d. structure of the driving increments, we show that the algorithm gives rise to a time-homogeneous Markov chain. We then prove the existence of an invariant measure for this chain and establish its uniqueness together with uniform geometric ergodicity under standard regularity conditions. Finally, we prove that the invariant measure of the algorithm converge weakly to the Gibbs measure of the limiting Riemannian Langevin diffusion as the discretization parameter vanishes, yielding a global practical probabilistic guarantee.**

***Index Terms*—Extremum seeking, stochastic optimization, Riemannian manifolds, averaging and Lie-bracket approximation, ergodic theory of Markov processes.**

I. Introduction

MODEL-FREE optimization algorithms that use only real-time measurements of an objective function play a central role in adaptive control, source seeking, and learning in dynamical systems. Among such algorithms, extremum seeking (ES) based on periodic dithering and averaging has a long history [1], [2], and its re-interpretation via Lie-bracket approximation [3] has enabled systematic designs with bounded update rates [4] and extensions to systems evolving on Lie groups and manifolds [5]–[7]. Deterministic ES algorithms of this type inherit a fundamental limitation of the averaged dynamics they emulate. Namely, gradient flows converge to local minimizers, and the guarantees are correspondingly local, or at best semi-global.

A classical remedy of this limitation in Euclidean settings is to inject persistent stochasticity so that the averaged/limiting dynamics become a Langevin diffusion, whose invariant Gibbs measure concentrates near global minimizers. This idea underlies simulated annealing in continuous time [8]–[10], stochastic approximation [11], and stochastic extremum seeking [12], [13]. On Riemannian manifolds, Langevin-type sampling and optimization have received substantial attention in the learning literature [14], and first-order optimization on manifolds is by now standard [15]. However, these methods require gradient information which is not available in the zeroth-order measurement-based setting.

In this note, we introduce and analyze a *stochastic Lie-bracket approximation* (SLBA) algorithm that combines the bounded-update-rate design of [4] with Brownian, rather than periodic, dither. The algorithm is an ordinary differential equation (ODE) on a compact manifold $\mathcal{M}$, driven by piecewise-smooth signals built from the increments of a Wiener process via a McShane-type interpolation [16]. The interpolation is designed so that, in the Wong–Zakai limit [17]–[20], the ODE trajectories converge to those of a Stratonovich stochastic differential equation (SDE) whose drift is the gradient of the objective function, even though the algorithm utilizes zeroth-order information only.

Contributions. First, we show that for a canonical choice of drift compensation, the generator of the limiting SDE coincides with that of a Riemannian Langevin diffusion with respect to the metric induced by the actuation vector fields. Consequently, its unique invariant measure is the Gibbs measure which concentrates on the global minimizers of the objective function ϕ. We then quantify this concentration and, when ϕ is a Morse function with a unique local minimum, we obtain a uniform mixing rate for the SDE, leading to a finite-horizon practical guarantee for the algorithm.

Second, we analyze the ergodic properties of the *algorithm itself* at a fixed discretization parameter $\varepsilon > 0$, rather than only of its diffusion limit. The key structural observation enabling our analysis is that the driving signals restart at each sampling instant and depend on the Wiener path only through its i.i.d. increments. As a result, the sampled process turns out to be a time-homogeneous Markov chain with a well-behaved transition kernel. The existence of an invariant measure π_ε then follows from the Krylov–Bogolyubov argument. To establish uniqueness, we prove a uniform Doeblin condition by showing that the map from noise increments to the state after one step is a submersion. Our proof yields uniqueness of π_ε and geometric ergodicity in total variation. Finally, we prove that, in the limit $\varepsilon \to 0$, the sequence of invariant measures π_ε converges weakly to the Gibbs measure of the limiting SDE. The combination produces a global practical optimization probabilistic guarantee that holds for arbitrary

This manuscript is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Grant No. HR00112530225.

M. Abdelgalil is with the Department of Mechanical and Aerospace Engineering, University at Buffalo, State University of New York, Buffalo, NY, 14260, USA (email: maabdelg@buffalo.edu).

Miroslav Krstic is with the Department of Mechanical and Aerospace Engineering, University of California San Diego, La Jolla, CA 92093, USA (email: mkrstic@ucsd.edu).

J. I. Poveda is with the Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093, USA (email: jipoveda@ucsd.edu).

initial conditions and on infinite horizons.

Relation to prior work. With respect to deterministic Lie-bracket ES on manifolds [5], [6], the present algorithm trades periodic dither for Brownian dither and thereby upgrades local practical convergence to a global, measure-theoretic guarantee. With respect to stochastic ES [12], the novelty is threefold: the manifold setting, the bounded update rates, and most importantly the analysis of the invariant measure of the *approximating* dynamics at fixed ε, which to the best of our knowledge has no counterpart in the ES literature. The Markov-chain viewpoint adopted here is closest in spirit to the analysis of numerical schemes for SDEs [24], [25], but the dynamics analyzed are not a discretization scheme: they are a continuous-time ODE with a genuinely non-smooth dependence on the noise increments.

Notation. $\mathcal{M}$ denotes a smooth, connected, compact, orientable n-dimensional manifold without boundary, embedded in $\mathbb{R}^m$; $C^\infty(\mathcal{M})$ and $\mathfrak{X}(\mathcal{M})$ denote smooth functions and vector fields; $\mathcal{P}(\mathcal{M})$ denotes the space of Borel probability measures on $\mathcal{M}$ with the topology of weak convergence; $\rightharpoonup$ denotes weak convergence; $\|\cdot - \cdot\|_{\mathrm{TV}}$ denotes total variation. For a metric $\langle\cdot,\cdot\rangle$ on $\mathcal{M}$, grad, div, Δ, ∇, and vol denote the gradient, divergence, Laplace–Beltrami operator, Levi-Civita connection, and Riemannian density. $B(z,\rho)$ denotes the open metric ball. $\mathcal{N}(0,\Sigma)$ denotes the centered Gaussian law with covariance Σ.

II. Problem Formulation and the SLBA Algorithm

Let $\phi \in C^\infty(\mathcal{M})$ be the objective function, accessible only through evaluations $\phi(x)$ along trajectories, and let $\mathcal{F} = \{f_1,\dots,f_r\} \subset \mathfrak{X}(\mathcal{M})$ be a family of vector fields satisfying the span condition

$$\operatorname{span}_{\mathbb{R}}\{f_1(x),\dots,f_r(x)\} = T_x\mathcal{M}, \qquad \forall x \in \mathcal{M}. \tag{1}$$

Such a family always exists: e.g., the tangential projections of the ambient coordinate vector fields of the embedding $\mathcal{M} \subset \mathbb{R}^m$ provide $r = m$ fields satisfying (1).

Let $\mathcal{W} := \{W \in C([0,\infty);\mathbb{R}^{2r}) : W_0 = 0\}$ be the standard Wienere space, i.e., the space of continuous functions equipped with the Wiener measure, and, for any $W \in \mathcal{W}$, let $\Delta_k^\ell := W^\ell_{(k+1)\varepsilon} - W^\ell_{k\varepsilon}$ denote the increments over the grid of step $\varepsilon > 0$, with $k \in \mathbb{N}_0$, $\ell \in \{1,\dots,2r\}$. For a gain $\kappa > 0$, define the interporation functions

$$U^1_\kappa(s) := s + \tfrac{\sqrt{4\pi\kappa+1}+1}{2\pi}\sin(2\pi s), \tag{2}$$

$$U^2_\kappa(s) := s + \tfrac{\sqrt{4\pi\kappa+1}+1}{\pi}\sin^2(\pi s), \tag{3}$$

and, for $t \in [k\varepsilon,(k+1)\varepsilon)$, $i \in \{1,\dots,r\}$, $j \in \{1,2\}$, the piecewise-smooth process

$$B^{2i+j-2}_\varepsilon(t,W) := W^{2i+j-2}_{k\varepsilon} + \Delta^{2i+j-2}_k\, U^\iota_\kappa\Big(\tfrac{t-k\varepsilon}{\varepsilon}\Big), \tag{4}$$

where $\iota = j$ if $\Delta^{2i-1}_k\Delta^{2i}_k \geq 0$ and $\iota = 3-j$ otherwise, i.e., the two interpolation profiles are swapped within the i-th pair when the corresponding increments have opposite signs. This piecewise smooth interpolation of Brownian motion is known as a McShane-type interpolation [16]. Writing $b^\ell_\varepsilon := \dot{B}^\ell_\varepsilon$, a piecewise continuous function, the SLBA algorithm is defined by the random ODE

$$\dot{x} = \bar{f}_0(x) + \tfrac{1}{\sqrt{\kappa}}\sum_{i=1}^{r}\sum_{j=1}^{2} F_j(\phi(x))\, f_i(x)\, b^{2i+j-2}_\varepsilon(t,W), \tag{5}$$

with $F_1 = \cos$, $F_2 = -\sin$, and a compensation drift $\bar{f}_0 \in \mathfrak{X}(\mathcal{M})$ specified in Theorem 2 in terms of the family of vector fields $\mathcal{F}$ only and independently from the function ϕ. Note that (5) is an ordinary differential equation for each sample path, and the objective enters only through the measured values $\phi(x(t))$, as in [4]. We write $x^\varepsilon_t(x_0)$ for the solution of (5) from an initial condition $x_0 \in \mathcal{M}$. Solutions exist globally and remain on $\mathcal{M}$ since the right-hand side is tangent to $\mathcal{M}$.

By the Wong–Zakai theory for approximations with corrections [16], [18, Ch. 7], the trajectories of (5) approximate, as $\varepsilon \to 0$, the sample paths of the Stratonovich SDE

$$\begin{aligned}\mathrm{d}\bar{x}_t = \Big(\bar{f}_0(\bar{x}) + \sum_{i=1}^{r}\big[F_1(\phi)f_i,\, F_2(\phi)f_i\big](\bar{x})\Big)\mathrm{d}t \\ + \tfrac{1}{\sqrt{\kappa}}\sum_{i=1}^{r}\sum_{j=1}^{2} F_j(\phi(\bar{x}))f_i(\bar{x}) \bullet \mathrm{d}W_t^{2i+j-2},\end{aligned} \tag{6}$$

where the bracket term is the drift correction generated by (2)–(4). We write $\bar{x}_t(x_0)$ for the solution of (6) and $\bar{P}_t$ for its Markov semigroup. The precise convergence statement we use is the following adaptation of [18, Thm. 7.2 and Rem. 7.2].

Theorem 1. *For every $T,\delta > 0$ there exists $\varepsilon^*(\delta,T) > 0$ such that for all $\varepsilon \in (0,\varepsilon^*)$,*

$$\sup_{x_0\in\mathcal{M}} \mathbb{E}\Big[\sup_{t\in[0,T]} \big\|x^\varepsilon_t(x_0) - \bar{x}_t(x_0)\big\|^2\Big] < \delta, \tag{7}$$

when both solutions are driven by the same sample path.

In what remains of this manuscript, we utilize the *stochastic convergence of trajectories* property established by Theorem 1, in conjunction with the properties of the limiting SDE (6), to analyze the behavior of the random ODE (5).

III. The Limiting Diffusion and Its Gibbs Measure

We begin by analyzing the limiting SDE (6). The span condition (1) induces a canonical metric on $\mathcal{M}$. Indeed, since the co-metric

$$\sum_{i=1}^{r} f_i \otimes f_i,$$

is positive definite by virtue of (1), there exists a unique Riemannian metric $\langle\cdot,\cdot\rangle$ on $\mathcal{M}$ such that, $\forall\psi \in C^\infty(\mathcal{M})$,

$$\operatorname{grad}\psi = \sum\nolimits_{i=1}^{r} f_i(\psi)\, f_i. \tag{8}$$

In the sequel, all geometric operators refer to this metric.

Theorem 2. *Let $\langle\cdot,\cdot\rangle$ be the metric* (8) *and define*

$$\bar{f}_0 = -\tfrac{1}{2\kappa}\sum_{i=1}^{r}\nabla_{f_i} f_i.$$

where ∇ is the associated Levi-Civita connection. Then, the infinitesimal generator of (6) *is given by*

$$\mathcal{L}\psi = \tfrac{1}{2\kappa}\Delta\psi - \langle \operatorname{grad}\phi, \operatorname{grad}\psi\rangle, \tag{9}$$

for all $\psi \in C^\infty(\mathcal{M})$. That is, the infinitesimal generator of (6) *coincides with the generator of the Langevin SDE defined by*

$$\mathrm{d}x = -\operatorname{grad}\phi(x)\mathrm{d}t + \kappa^{-1/2}\sum_{i=1}^{n} e_i(x) \bullet \mathrm{d}W^i$$

where $\{e_i\}$ is any local orthonormal frame. Moreover, the SDE (6) *admits the unique invariant measure*

$$\mu_\kappa = c_\kappa^{-1} e^{-2\kappa\phi}\,\mathrm{vol}, \qquad c_\kappa := \int_{\mathcal{M}} e^{-2\kappa\phi}\mathrm{vol}. \tag{10}$$

Proof. To simplify notation, let $\phi_j = F_j \circ \phi$. Utilizing the properties of the Lie Bracket, we compute

$$[\phi_1 f_i, \phi_2 f_i] = (\phi_1 f_i(\phi_2) - \phi_2 f_i(\phi_1))f_i. \tag{11}$$

Consequently, the limiting SDE (6) simplifies to

$$\mathrm{d}\bar{x}_t = f_0(\bar{x})\,\mathrm{d}t + \frac{1}{\sqrt{\kappa}}\sum_{i=1}^{r}\sum_{j=1}^{2}\phi_j f_i(\bar{x}) \bullet dW_t^{2i+j-2}, \tag{12}$$

where the vector field f_0 is defined by

$$f_0 := \bar{f}_0 + \sum_{i=1}^{r}(\phi_1 f_i(\phi_2) - \phi_2 f_i(\phi_1))f_i. \tag{13}$$

The associated infinitesmal generator $\mathcal{L}$ is defined by

$$\mathcal{L}\psi = f_0(\psi) + \frac{1}{2\kappa}\sum_{i=1}^{r}\sum_{j=1}^{2}\phi_j f_i(\phi_j f_i(\psi)), \tag{14}$$

for all $\psi \in C^\infty(\mathcal{M})$. Using the product rule, we compute that

$$\phi_j f_i(\phi_j f_i(\psi)) = \phi_j f_i(\phi_j) f_i(\psi) + \phi_j^2 f_i(f_i(\psi)). \tag{15}$$

Therefore, for any $i \in \{1, \dots, r\}$, we have that

$$\begin{aligned}\sum_{j=1}^{2}\phi_j f_i(\phi_j f_i(\psi)) &= (\phi_1^2 + \phi_2^2) f_i(f_i(\psi)) \\ &\quad + (\phi_1 f_i(\phi_1) + \phi_2 f_i(\phi_2)) f_i(\psi).\end{aligned} \tag{16}$$

From the chain rule and the definition of F_j, we compute that

$$f_i(\phi_1) = \phi_2 f_i(\phi), \quad f_i(\phi_2) = -\phi_1 f_i(\phi), \quad \phi_1^2 + \phi_2^2 = 1.$$

Substituting into (13) and (16), we obtain that

$$f_0 = \bar{f}_0 - \sum_{i=1}^{r} f_i(\phi) f_i, \quad \sum_{j=1}^{2}\phi_j f_i(\phi_j f_i(\psi)) = f_i(f_i(\psi)).$$

Substituting into the infinitesmal generator, we obtain that

$$\mathcal{L}\psi = f_0(\psi) + \frac{1}{2\kappa}\sum_{i=1}^{r} f_i(f_i(\psi)),$$

If ∇ is any affine connection on $\mathcal{M}$, then, using [26, Proposition 4.15], we compute that

$$f_i(f_i(\psi)) = (\nabla_{f_i} d\psi)(f_i) + \nabla_{f_i} f_i(\psi), \tag{17}$$

In particular, if ∇ is the Levi-Civita connection associated with the metric (8), then

$$\sum_{i=1}^{r} f_i(f_i(\psi)) = \Delta\psi + \sum_{i=1}^{r}\nabla_{f_i} f_i(\psi). \tag{18}$$

where Δ is the Laplace-Beltrami operator. Therefore, the infinitesmal generator $\mathcal{L}$ acts on a smooth function ψ as

$$\mathcal{L}\psi = f_0(\psi) + \frac{1}{2\kappa}\sum_{i=1}^{r}\nabla_{f_i} f_i(\psi) + \frac{1}{2\kappa}\Delta\psi \tag{19}$$

In particular, if the vector field $\bar{f}_0$ is taken to be

$$\bar{f}_0 = -\frac{1}{2\kappa}\sum_{i=1}^{r}\nabla_{f_i} f_i, \tag{20}$$

then the infinitesmal generator $\mathcal{L}$ acts is defined by

$$\mathcal{L}\psi = -\langle \operatorname{grad}\phi, \operatorname{grad}\psi\rangle + \frac{1}{2\kappa}\Delta\psi. \tag{21}$$

Due to compactness of $\mathcal{M}$ and the fact that the operator $\mathcal{L}$ is uniformly elliptic for any $\kappa > 0$, it follows that its adjoint $\mathcal{L}^*$, defined by

$$\mathcal{L}^*\psi = \operatorname{div}(\psi \operatorname{grad}\phi) + \frac{1}{2\kappa}\Delta\psi,$$

where div is the divergence operator, has a unique invariant density. Indeed, the Fokker-Planck equation corresponding to the limiting SDE (12) can be succinctly expressed as

$$\begin{aligned}\partial_t\rho = \mathcal{L}^*\rho &= \operatorname{div}(\rho \operatorname{grad}\phi) + \frac{1}{2\kappa}\Delta\rho \\ &= \frac{1}{2\kappa}\operatorname{div}\left(\mathrm{e}^{-2\kappa\phi}\operatorname{grad}\left(\mathrm{e}^{2\kappa\phi}\rho\right)\right).\end{aligned}$$

From the the last equality, we observe that the density

$$\rho_\kappa = c_\kappa^{-1}\,\mathrm{e}^{-2\kappa\phi}, \qquad c_\kappa := \int_{\mathcal{M}}\mathrm{e}^{-2\kappa\phi}\,\mathrm{vol},$$

where vol is the Riemannian density, is a stationary solution of the Fokker-Planck operator, and, therefore, $\mu_\kappa := \rho_\kappa\,\mathrm{vol}$ is the unique invariant measure for (6). □

The explicit form of the invariant measure μ_κ provided by Theorem 2 allows us to study its asymptotic behavior when $\kappa \gg 1$. To that end, we have the following proposition.

Proposition 1. *Let $\phi_* := \min_{\bar{x}\in\mathcal{M}}\phi(\bar{x})$ and $S := \{\bar{x} \in \mathcal{M} : \phi(\bar{x}) = \phi_*\}$. For every open neighborhood U of S,*

$$\mu_\kappa(U) \xrightarrow[\kappa\to\infty]{} 1.$$

Proof. Fix any open $U \supset S$ and set $K := \mathcal{M} \setminus U$, which is compact and disjoint from S. By continuity of ϕ and compactness, the minimum

$$m_K := \min_{\bar{x}\in K}\phi(\bar{x})$$

exists and satisfies $m_K > \phi_*$. Choose $\varepsilon \in (0, m_K - \phi_*)$ and define the nonempty closed set

$$N := \{\bar{x} \in \mathcal{M} :\ \phi(\bar{x}) \le \phi_* + \varepsilon\}.$$

Then, we have that

$$\int_K e^{-2\kappa\phi}\,\mathrm{vol} \;\le\; e^{-2\kappa m_K}\,\mathrm{vol}(K),$$
$$\int_N e^{-2\kappa\phi}\,\mathrm{vol} \;\ge\; e^{-2\kappa(\phi_*+\varepsilon)}\,\mathrm{vol}(N).$$

Since $Z_\kappa \ge \int_N e^{-2\kappa\phi}\,\mathrm{vol}$, we obtain

$$\mu_\kappa(K) \le \frac{e^{-2\kappa m_K}\,\mathrm{vol}(K)}{e^{-2\kappa(\phi_*+\varepsilon)}\,\mathrm{vol}(N)} = C\,e^{-2\kappa\,(m_K-\phi_*-\varepsilon)},$$

where $C := \mathrm{vol}(K)/\mathrm{vol}(N)$ is independent of κ. Because $m_K - \phi_* - \varepsilon > 0$, the right-hand side decays exponentially as $\kappa \to \infty$, hence $\mu_\kappa(K) \to 0$ and therefore $\mu_\kappa(U) \to 1$. □

Although the invariant measure μ_κ concentrates on the set of minimizers S in the limit $\kappa \to \infty$, the convergence of the marginals of the process (6) to the invariant measure μ_κ is, in general, slow for large κ. This is due to the fact that, for a general compact boundary-less manifold $\mathcal{M}$ and a general function ϕ, the spectral gap of the operator $\mathcal{L}$ may decays in the limit $\kappa \to \infty$. Nevertheless, for a fixed $\kappa > 0$, the spectral gap of $\mathcal{L}^*$ is uniformly bounded away from zero and, as a result, convergence is guaranteed. On the other hand, there is an important case wherein the spectral gap of $\mathcal{L}^*$ can be uniformly upper bounded away from 0 in the limit $\kappa \to \infty$.

Proposition 2. *Let $\phi \in C^\infty(\mathcal{M})$ be any morse function with a unique local minimum, i.e., there exists a point $x_\star \in \mathcal{M}$ such that $x_\star$ is the only critical point of ϕ for which* $\mathrm{Hess}\,\phi(x_\star)$ *is positive definite. Let μ_κ be the invariant measure associated with* (6) *and let ρ_t be the solution to the Fokker-Planck equation*

$$\partial_t \rho_t = \mathcal{L}^*\rho_t = \mathrm{div}(\rho_t\,\mathrm{grad}\,\phi) + \frac{1}{2\kappa}\Delta\rho_t, \tag{22}$$

from an initial probability density $\rho_0 \in C^\infty(\mathcal{M})$. Then there exists $\lambda > 0$ such that, for all $\kappa > 0$ and all $t \ge 0$,

$$\|\rho_t - \rho_\kappa\|_{L^1(\mathrm{vol})} \le \mathrm{e}^{-\lambda t}\|\rho_0 - \rho_\kappa\|_{L^2(\rho_\kappa^{-1}\mathrm{vol})}, \tag{23}$$

where ρ_κ is the density of μ_κ with respect to vol.

Proof. The operator $\mathcal{D} = -\mathcal{L}$ belongs to the class of diffusion operators considered in [21, Section 8.2]. Therefore, it follows from [21, Proposition 2.1] and the fact that ϕ is a Morse function with a unique local minimum that the smallest non-trivial eigenvalue of $\mathcal{D}$ is uniformly bounded below by a strictly positive constant $\lambda > 0$. Consequently, the Markov semigroup $P^t = \mathrm{e}^{t\mathcal{L}}$ satisfies

$$\|P^t f\|_{L^2(\mu_\kappa)} \le \mathrm{e}^{-\lambda t}\|f\|_{L^2(\mu_\kappa)},$$

for any $f \in L^2(\mu_\kappa)$ with $\int_{\mathcal{M}} f\,\mu_\kappa = 0$. To proceed, we define the function $u_t := \frac{\rho_t}{\rho_\kappa} - 1$ and compute that

$$\mathcal{L}^*(\rho_\kappa u_t + 1) = \mathcal{L}^*(\rho_\kappa u_t) = \rho_\kappa \mathcal{L} u_t,$$

where the last equality follows from the fact that $\mathcal{L}$ is self-adjoint in $L^2(\mu_\kappa)$. On the other hand,

$$\partial_t(\rho_\kappa u_t + 1) = \rho_\kappa \partial u_t.$$

Since $\rho_t = \rho_\kappa u_t + 1$ and $\partial_t \rho_t = \mathcal{L}^*\rho_t$, it follows that

$$\partial_t u_t = \mathcal{L} u_t.$$

Moreover, direct computation gives that

$$\int_{\mathcal{M}} u_t\,\mu_\kappa = \int_{\mathcal{M}} (\rho_t - \rho_\kappa)\,\mathrm{vol} = 0,$$

and, consequently, we obtain that

$$\|u_t\|_{L^2(\mu_\kappa)} \le \mathrm{e}^{-\lambda t}\|u_0\|_{L^2(\mu_\kappa)},$$

for all $\kappa > 0$ and all $t \ge 0$, where $\|u\|_{L^2(\mu_\kappa)} = (\int u^2\mu_\kappa)^{\frac{1}{2}}$. From the Cauchy-Schwartz inequality and the positivity of ρ_κ for any $\kappa > 0$, we obtain that

$$\begin{aligned}
\|\rho_t - \rho_\kappa\|_{L^1(\mathrm{vol})} &= \int_{\mathcal{M}} |\rho_t - \rho_\kappa|\rho_\kappa^{-\frac{1}{2}}\rho_\kappa^{\frac{1}{2}}\,\mathrm{vol} \\
&\le \left(\int_{\mathcal{M}} (\rho_t - \rho_\kappa)^2\rho_\kappa^{-1}\,\mathrm{vol}\right)^{\frac{1}{2}} \left(\int_{\mathcal{M}} \rho_\kappa\,\mathrm{vol}\right)^{\frac{1}{2}} \\
&= \left(\int_{\mathcal{M}} u_t^2\,\rho_\kappa\,\mathrm{vol}\right)^{\frac{1}{2}} = \|u_t\|_{L^2(\mu_\kappa)} \\
&\le \mathrm{e}^{-\lambda t}\|u_0\|_{L^2(\mu_\kappa)}.
\end{aligned}$$

□

The following Corollary is also immediate.

Corollary 1. *For any $\kappa > 0$ and any smooth probability distribution ρ_0, let $\bar{x}_t$ be a solution of the SDE* (6) *such that the initial condition $\bar{x}_0 \sim \rho_o\,\mathrm{vol}$. Then, for every measurable set $U \subset \mathcal{M}$, we have that*

$$|\mathbb{P}\{\bar{x}_t \in U\} - \mu_\kappa(U)| \le \frac{1}{2}\|\rho_0 - \rho_\kappa\|_{L^2(\rho_\kappa^{-1}\mathrm{vol})}\mathrm{e}^{-\lambda t}, \tag{24}$$

for all $\kappa > 0$ and all $t \ge 0$. In particular, if $\rho_0 = (\int_{\mathcal{M}} \mathrm{vol})^{-1}$, i.e. the uniform probability density, then

$$|\mathbb{P}\{\bar{x}_t \in U\} - \mu_\kappa(U)| \le \frac{1}{2}\mathrm{e}^{\kappa\delta - \lambda t}, \tag{25}$$

where $\delta = \max_{x\in\mathcal{M}}\phi(x) - \min_{x\in\mathcal{M}}\phi(x)$.

Proof. By definition, the total variation distance between the probability laws $\mathbb{P}$ and $\mathbb{Q}$ is

$$\sup_{A \text{ measurable}} |\mathbb{P}(A) - \mathbb{Q}(A)|. \tag{26}$$

When the laws $\mathbb{P}$ and $\mathbb{Q}$ have densities p and q, it follows that

$$\sup_{A \text{ measurable}} |\mathbb{P}(A) - \mathbb{Q}(A)| = \frac{1}{2}\|p - q\|_{L^1(\mathrm{vol})}. \tag{27}$$

Substituting for $\mathbb{P}$ with the law of the random variable $\bar{x}_t$ and for $\mathbb{Q}$ with the invariant measure μ_κ, we obtain that

$$|\mathbb{P}\{\bar{x}_t \in U\} - \mu_\kappa(U)| = \frac{1}{2}\|\rho_t - \rho_\kappa\|_{L^1(\mathrm{vol})}, \tag{28}$$

where ρ_t solve the Fokker-Planck equation from the initial condition ρ_0. Hence, we obtain that

$$|\mathbb{P}\{\bar{x}_t \in U\} - \mu_\kappa(U)| \le \frac{1}{2}\|\rho_0 - \rho_\kappa\|_{L^2(\rho_\kappa^{-1}\mathrm{vol})}\mathrm{e}^{-\lambda t}, \tag{29}$$

for any measurable $U \subset \mathcal{M}$, any κ, and any $t \geq 0$. If, in addition, ρ_0 is the uniform probability density, then

$$\begin{aligned}\|\rho_0 - \rho_\kappa\|^2_{L^2(\rho_\kappa^{-1}\mathrm{vol})} &= \int_{\mathcal{M}} \frac{(\rho_0-\rho_\kappa)^2}{\rho_\kappa}\mathrm{vol} \\ &= \int_{\mathcal{M}} \frac{\rho_0^2 - 2\rho_0\rho_\kappa + \rho_\kappa^2}{\rho_\kappa}\mathrm{vol} \\ &= \int_{\mathcal{M}} \frac{\rho_0^2}{\rho_\kappa}\mathrm{vol} - 2\int_{\mathcal{M}} \rho_0\,\mathrm{vol} + \int_{\mathcal{M}} \rho_\kappa\,\mathrm{vol} \\ &= \frac{1}{\mathrm{Vol}(\mathcal{M})^2}\int_{\mathcal{M}} \frac{1}{\rho_\kappa}\mathrm{vol} - 1,\end{aligned}$$

where $\mathrm{Vol}(\mathcal{M}) = \int_{\mathcal{M}} \mathrm{vol}$, and we used the fact that $\rho_0 = \mathrm{Vol}(\mathcal{M})^{-1}$, $\int_{\mathcal{M}} \rho_\kappa\,\mathrm{vol} = 1$, and $\int_{\mathcal{M}} \rho_0\,\mathrm{vol} = 1$. Recalling the form of the invariant measure, we see that

$$\rho_\kappa(x) = \frac{\mathrm{e}^{-2\kappa\phi(x)}}{\int_{\mathcal{M}} \mathrm{e}^{-2\kappa\phi}\,\mathrm{vol}} \geq \frac{\mathrm{e}^{-2\kappa\delta}}{\mathrm{Vol}(\mathcal{M})}, \tag{30}$$

for all $x \in \mathcal{M}$. Therefore, we obtain the bound

$$\begin{aligned}\|\rho_0 - \rho_\kappa\|^2_{L^2(\rho_\kappa^{-1}\mathrm{vol})} &\leq \frac{\mathrm{e}^{2\kappa\delta}}{\mathrm{Vol}(\mathcal{M})}\int_{\mathcal{M}} \frac{1}{\rho_\kappa}\mathrm{vol} - 1 \\ &= \mathrm{e}^{2\kappa\delta} - 1 \leq \mathrm{e}^{2\kappa\delta},\end{aligned}$$

from which the second inequality follows. □

Based on the properties we have obtained for the limiting SDE (6), we may now utilize the stochastic convergence-of-trajectories property established by Theorem 1. In particular, if x_t^ε and $\bar{x}_t$ are the solutions for the random ODE (5) and the SDE (6), respectively, from the same initial condition $x_0^\varepsilon = \bar{x}_0 \sim \rho_0\,\mathrm{vol}$ and for the same Brownian motion, then the bound in Theorem 1 implies that the law of x_t convergence to the law of $\bar{x}_t$ in the limit $\varepsilon \to 0$, uniformly on any compact interval $[0, T]$. In particular, we have the following proposition.

Proposition 3. *For any $T > 0$, any $\epsilon > 0$, any $\kappa > 0$, and any open set $U \subset \mathcal{M}$ with a smooth boundary, there exists $\varepsilon^* \in (0, \infty)$ such that, for all $\varepsilon \in (0, \varepsilon^*)$,*

$$|\mathbb{P}\{x_t^\varepsilon \in U\} - \mu_\kappa(U)| < \frac{1}{2}\|\rho_0 - \rho_\kappa\|^2_{L^2(\rho_\kappa^{-1}\mathrm{vol})}\mathrm{e}^{-\lambda t} + \epsilon.$$

Proof. Using the triangle inequality, we have that

$$\begin{aligned}|\mathbb{P}\{x_t^\varepsilon \in U\} - \mu_\kappa(U)| &\leq |\mathbb{P}\{x_t^\varepsilon \in U\} - \mathbb{P}\{\bar{x}_t \in U\}| \\ &\quad + |\mathbb{P}\{\bar{x}_t \in U\} - \mu_\kappa(U)| \\ &\leq |\mathbb{P}\{x_t^\varepsilon \in U\} - \mathbb{P}\{\bar{x}_t \in U\}| \\ &\quad + \frac{1}{2}\|\rho_0 - \rho_\kappa\|^2_{L^2(\rho_\kappa^{-1}\mathrm{vol})}\end{aligned}$$

Since $\rho_0 \in C^\infty(\mathcal{M})$, the law of $\bar{x}_t$ admits a density for all $t \geq 0$ and, thus, any open set $U \subset \mathcal{M}$ with a smooth boundary is a continuity set of the law of $\bar{x}_t$. Consequently, from Theorem 1, we obtain that there exists $\varepsilon^* > 0$ such that, for all $\varepsilon \in (0, \varepsilon^*)$ and all $t \in [0, T]$,

$$|\mathbb{P}\{x_t^\varepsilon \in U\} - \mathbb{P}\{\bar{x}_t \in U\}| < \epsilon. \tag{31}$$

□

IV. Invariant Measure Theory of the Algorithm

Proposition 3 is confined to compact horizons since it says nothing about the limit $t \to \infty$. The current section removes that limitation by analyzing the ergodic properties of (5) directly. Throughout, $\varepsilon, \kappa > 0$ are fixed unless stated otherwise, and, for $\ell = 2i + j - 2$, we define

$$g_\ell := F_j(\phi) f_i$$

A. Markovianity and the existence of an invariant measure

Define the sampled process $X_k := x^\varepsilon_{k\varepsilon}$, where x_t^ε is a solution of (5). Then, we have the following result.

Lemma 1. *There exists a measurable map $\Psi_\varepsilon : \mathcal{M} \times \mathbb{R}^{2r} \to \mathcal{M}$ such that the sampled process X_k satisfies*

$$X_{k+1} = \Psi_\varepsilon(X_k, \Delta_k)$$

where $\Delta_k := (\Delta_k^1, \ldots, \Delta_k^{2r})$ are i.i.d., distributed according to $\mathcal{N}(0, \varepsilon I_{2r})$. That is, (X_k) is a time-homogeneous Markov chain with the transition kernel

$$P_\varepsilon(x, A) = \mathbb{P}\{\Psi_\varepsilon(x, \Delta) \in A\}.$$

Moreover, $P_\varepsilon\psi \in C(\mathcal{M})$ for any $\psi \in C(\mathcal{M})$, where

$$P_\varepsilon\psi(x) := \int_{\mathbb{R}^{2r}} \psi(\Psi_\varepsilon(x, \Delta))\,\gamma_\varepsilon(\mathrm{d}\Delta),$$

and $\gamma_\varepsilon := \mathcal{N}(0, \varepsilon I_{2r})$. That is, P_ε maps $C(\mathcal{M})$ into $C(\mathcal{M})$.

Proof. On $[k\varepsilon, (k+1)\varepsilon)$, differentiating (4) in t eliminates the additive constant $W^\ell_{k\varepsilon}$, so $b^\ell_\varepsilon(t, W)$ depends on W only through the increments Δ_k which, by definition, are i.i.d samples from the Gaussian distribution $\mathcal{N}(0, \varepsilon I)$. Therefore, if we define $\Psi_\varepsilon(x, \Delta)$ as the time-1 solution of the ODE

$$y'(s) = \varepsilon\bar{f}_0(y) + \tfrac{1}{\sqrt{\kappa}}\sum\nolimits_\ell g_\ell(y)\,\Delta^\ell\,(U_\kappa^{\iota(\Delta)})'(s), \quad y(0) = x, \tag{32}$$

we see that $\Psi_\varepsilon(x, \Delta)$ is independent of k, and, therefore, the sampled process X_k is time-homogeneous Markov process, where the Markov property follows from the independence of the increments of the Wiener process. To complete the proof, it remains to show that P_ε maps $C(\mathcal{M})$ into $C(\mathcal{M})$. Since $\Delta^{2i-1}\Delta^{2i} = 0$ if and only if one of the two coordinates vanishes, the branch-switching set

$$\mathcal{D}_0 := \bigcup_i^{2r} \{\Delta \in \mathbb{R}^{2r} : \Delta^{2i-1}\Delta^{2i} = 0\}$$

coincides with the union of the coordinate hyperplanes

$$\mathbb{H}_\ell := \{\Delta \in \mathbb{R}^{2r} \mid \Delta^\ell = 0\}.$$

In particular, it is closed and has zero measure with respect to the Lebesgue measure and, consequently, it also has zero with respect to the Gaussian measure $\mathcal{N}(0, \varepsilon I)$. Its complement is the disjoint union of the 4^r open orthants of $\mathbb{R}^{2r}$. On each orthant, the sign pattern of Δ, and therefore the branch $\iota(\Delta)$ of every pair in (32) is constant. Consequently, for Δ ranging over a fixed orthant, the right-hand side of (32) is a single smooth function of (y, Δ, s). By smooth dependence of ODE solutions on initial conditions and parameters, the map Ψ_ε is

jointly smooth on $\mathcal{M} \times O$ for each orthant $O \subset \mathbb{R}^{2r}$. Borel measurability of Ψ_ε on all of $\mathcal{M} \times \mathbb{R}^{2r}$ follows, since its domain partitions into finitely many Borel sets, on each of which it is continuous.

However, continuity of $\Psi_\varepsilon(x, \cdot)$ on all of $\mathbb{R}^{2r}$ fails. Indeed, the map $\Psi_\varepsilon(x, \cdot)$ has a genuine discontinuity across the set $\mathcal{D}_0$. Nevertheless, the discontinuity set $\mathcal{D}_0$ has zero measure, which allows us proceed as follows. Fix a function $\psi \in C(\mathcal{M})$, a sequence $x_n \to x$ in $\mathcal{M}$, and define

$$h_n(\Delta) := \psi(\Psi_\varepsilon(x_n, \Delta)), \qquad h(\Delta) := \psi(\Psi_\varepsilon(x, \Delta)).$$

For every $\Delta \notin \mathcal{D}_0$, the point Δ lies in some orthant O, and $(x_n, \Delta) \to (x, \Delta)$ within $\mathcal{M} \times O$, on which Ψ_ε is continuous. Hence, $h_n(\Delta) \to h(\Delta)$, and $h_n \to h$ pointwise $\mathcal{N}(0, \varepsilon I)$-almost everywhere. Since, by definition, $|h_n| \le \|\psi\|_\infty$, the Dominated Convergence theorem [27, Theorem 3.3.1] yields

$$P_\varepsilon \psi(x_n) = \int h_n \, \mathrm{d}\gamma_\varepsilon \to \int h \, \mathrm{d}\gamma_\varepsilon = P_\varepsilon \psi(x),$$

i.e., $P_\varepsilon \psi$ is sequentially continuous whenever $\psi \in C(\mathcal{M})$. Because $\mathcal{M}$ is metrizable, sequential continuity implies continuity, and we have $P_\varepsilon \psi \in C(\mathcal{M})$. □

Proposition 4. *For every $\varepsilon > 0$, the chain (X_k) admits an invariant probability measure $\pi_\varepsilon \in \mathcal{P}(\mathcal{M})$.*

Proof. From [24, Prop. 6.1.1], the chain (X_k) is weak Feller since its kernel P_ε satisfies $P_\varepsilon \psi \in C(\mathcal{M})$ for any $\psi \in C(\mathcal{M})$. From [24, Prop. 12.0.1], it follows that the chain (X_k) admits at least one invariant measure. □

B. Uniqueness of the invariant measure

Proposition (4) establishes that the sampled process has an invariant measure but it does not ascertain its uniqueness. That is, the chain could possess several invariant measures. The goal of this subsection is exclude this possibility. To that end, we have the following result.

Theorem 3. *There exists $\varepsilon_0 > 0$ such that, for every $\varepsilon \in (0, \varepsilon_0]$, there exist $N \in \mathbb{N}$, $\alpha \in (0, 1)$, and $\nu \in \mathcal{P}(\mathcal{M})$ such that, for all $x \in \mathcal{M}$, we have that*

$$P_\varepsilon^{N+1}(x, \cdot) \ge \alpha \nu(\cdot).$$

Consequently, π_ε is the unique invariant measure of the chain, $\operatorname{supp} \pi_\varepsilon = \mathcal{M}$, and

$$\big\| P_\varepsilon^k(x, \cdot) - \pi_\varepsilon \big\|_{\mathrm{TV}} \le 2(1 - \alpha)^{\lfloor k/(N+1) \rfloor}, \tag{33}$$

for all $x \in \mathcal{M}$ and all $k \in \mathbb{N}$

Proof. Fix z and take $\rho, \mathcal{O}, c$ from Lemma 2 and N, δ_0 from Lemma 3, both in the Appendix. Then $P_\varepsilon^{N+1}(x, A) \ge \int_{B(z,\rho)} P_\varepsilon^N(x, \mathrm{d}y) P_\varepsilon(y, A) \ge \delta_0 c \operatorname{vol}(A \cap \mathcal{O})$, which is the claimed minorization with $\nu := \operatorname{vol}(\cdot \cap \mathcal{O})/\operatorname{vol}(\mathcal{O})$ and $\alpha := \delta_0 c \operatorname{vol}(\mathcal{O})$. Uniqueness and (33) are the standard consequences of a uniform Doeblin condition [24, Ch. 16]. Full support follows from Lemma 3 and invariance. □

C. Convergence to the Gibbs measure

Theorem 4. *$\pi_\varepsilon \rightharpoonup \mu_\kappa$ as $\varepsilon \to 0$.*

Proof. By compactness of $\mathcal{P}(\mathcal{M})$ it suffices to show that every weak limit point π of (π_{ε_n}), $\varepsilon_n \to 0$, is $\bar{P}_t$-invariant; uniqueness in Theorem 2 then gives $\pi = \mu_\kappa$. Fix $t > 0$, $\psi \in C(\mathcal{M})$, and set $k_n := \lceil t/\varepsilon_n \rceil$, $t_n := k_n \varepsilon_n \to t$. By invariance of π_{ε_n},

$$\Big| \int (\bar{P}_t \psi - \psi) \, \mathrm{d}\pi_{\varepsilon_n} \Big| \le \sup_{x \in \mathcal{M}} \big| \mathbb{E}[\psi(\bar{x}_t(x)) - \psi(x_{t_n}^{\varepsilon_n}(x))] \big|.$$

For any $\theta > 0$, the right-hand side is bounded by $\omega_\psi(\theta) + 2\|\psi\|_\infty \theta^{-2} \sup_x \mathbb{E}[\|\bar{x}_t(x) - x_{t_n}^{\varepsilon_n}(x)\|^2]$, where ω_ψ is the modulus of continuity of ψ, and

$$\begin{aligned} \sup_x \mathbb{E}\big[\|\bar{x}_t(x) - x_{t_n}^{\varepsilon_n}(x)\|^2\big] \le 2 \sup_x \mathbb{E}\big[\|\bar{x}_t(x) - \bar{x}_{t_n}(x)\|^2\big] \\ + 2 \sup_x \mathbb{E}\Big[\sup_{s \le t+1} \|\bar{x}_s(x) - x_s^{\varepsilon_n}(x)\|^2 \Big]. \end{aligned}$$

The second term vanishes as $n \to \infty$ thanks to Theorem 1 with $T = t + 1$. The first term vanishes by L^2-continuity of the SDE flow, uniformly by compactness. Since $\bar{P}_t \psi \in C(\mathcal{M})$ ($\bar{P}_t$ is Feller), passing to the limit along $\pi_{\varepsilon_n} \rightharpoonup \pi$ and then $\theta \to 0$ gives $\int \bar{P}_t \psi \, \mathrm{d}\pi = \int \psi \, \mathrm{d}\pi$ for all t, ψ. □

Corollary 2 (Global practical optimization)**.** *Let $U \subset \mathcal{M}$ be open with $S \subset U$, and let $\eta \in (0, 1)$. There exists κ^* such that for every $\kappa \ge \kappa^*$ there exists $\bar{\varepsilon} \in (0, \varepsilon_0]$ such that for every $\varepsilon \in (0, \bar{\varepsilon}]$ there exists $k^* \in \mathbb{N}$ such that*

$$\inf_{x_0 \in \mathcal{M}} \mathbb{P}\big\{ x_{k\varepsilon}^\varepsilon(x_0) \in U \big\} \ \ge \ 1 - \eta, \qquad \forall k \ge k^*. \tag{34}$$

Proof. By Proposition **??**, choose κ^* with $\mu_\kappa(U) > 1 - \eta/2$ for $\kappa \ge \kappa^*$ and fix such κ. By Theorem 4 and the portmanteau theorem for the open set U, $\liminf_{\varepsilon \to 0} \pi_\varepsilon(U) \ge \mu_\kappa(U)$, so there is $\bar{\varepsilon} \le \varepsilon_0$ with $\pi_\varepsilon(U) > 1 - 3\eta/4$ for $\varepsilon \le \bar{\varepsilon}$; fix such ε. By (33), choose k^* with $2(1 - \alpha)^{\lfloor k^*/(N+1) \rfloor} < \eta/4$. Then $\mathbb{P}\{x_{k\varepsilon}^\varepsilon(x_0) \in U\} \ge \pi_\varepsilon(U) - \eta/4 > 1 - \eta$ for all $k \ge k^*$ and all x_0. □

Remark 1 (Intermediate times)**.** *The statements above concern the sampled chain. For $t \in [k\varepsilon, (k+1)\varepsilon)$, the within-step excursion satisfies $\mathbb{E}[\sup_t \|x_t^\varepsilon - X_k\|^2] = O(\varepsilon)$ uniformly in k by (**??**) and Gaussian moment bounds, so the law of x_t^ε for $t \ge k^* \varepsilon$ lies within $o_\varepsilon(1)$ of π_ε in the weak topology.*

V. Numerical Example

Example 1. *Let $\mathcal{M} = \mathbb{S}^2$ and $\mathcal{F} = \{f_1, f_2, f_3\}$ where the vector fields f_i are given by*

$$\begin{aligned} f_1(x) &= x^2 \partial_1 - x^1 \partial_2, \\ f_2(x) &= x^3 \partial_2 - x^2 \partial_3, \\ f_3(x) &= x^1 \partial_3 - x^3 \partial_1 \end{aligned}$$

in the ambient coordinates on $\mathbb{R}^3$. Fix any $x_\star \in \mathbb{S}^2$ and let the function ϕ to be given in coordinates by

$$\phi(x) := 1 - (x_\star^1 x^1 + x_\star^2 x^2 + x_\star^3 x^3).$$

It is not difficult to verify that the function ϕ is a Morse function. Moreover, ϕ has two critical points, namely, $\pm x_\star$,

Fig. 1. Example 1: (left) the function ϕ, (right) sample paths.

and only the critical point $x_\star$ is a local minimum. The Riemannian metric $\langle \cdot, \cdot \rangle$ corresponding to the family $\mathcal{F}$ is defined in the ambient coordinates by

$$\langle \cdot, \cdot \rangle = \sum_{i,j=1}^{3} g_{ij} \mathrm{d}x^i \otimes \mathrm{d}x^j, \tag{35}$$

where the matrix $G = [g_{ij}]$ is

$$G(x) = I - xx^\top. \tag{36}$$

In particular, the metric $\langle \cdot, \cdot \rangle$ coincides with the induced metric inherited from the ambient Euclidean metric on $\mathbb{R}^3$. Consequently, if $D_X Y$ denotes the directional derivative of the vector field Y along the vector field X in the ambient space $\mathbb{R}^3$, then the Levi-Civita connection on $\mathcal{M}$ in the ambient coordinates is given by

$$\nabla_X Y = G(x)^\dagger D_X Y. \tag{37}$$

where $G(x)^\dagger$ is the Moore-Penrose inverse of $G(x)$. Utilizing this expression, we compute that

$$\sum_{i=1}^{3} \nabla_{f_i} f_i = 0, \tag{38}$$

i.e., we may take the vector field $\bar{f}_0 = 0$. We now generate several sample paths of the random ODE (5) *from the initial condition $-x_\star$. For simplicity, we fix $x_\star = (0, 0, 1)$. A plot of the function ϕ and sample paths on $\mathcal{M}$ are shown in Figure 1 for $\varepsilon = 0.001$, and $\kappa = 25$. As expected, the sample paths converge to the vicinity global minimizer of the function ϕ. Moreover, the trajectories remain on $\mathcal{M}$ for all time.*

VI. Conclusions

We analyzed a stochastic Lie-bracket approximation algorithm for global, measurement-based optimization on compact manifolds. Beyond the characterization of the Wong–Zakai limit as a Riemannian Langevin diffusion with Gibbs invariant measure, the main contribution is the ergodic theory of the algorithm itself at fixed discretization: Markov structure, existence, uniqueness via an explicit Doeblin minorization enabled by the submersivity of the increment-to-state map, geometric ergodicity, and weak convergence of the invariant measures to the Gibbs measure, culminating in an infinite-horizon global practical optimization guarantee from arbitrary initial conditions. Future work includes quantitative bounds on the distance between π_ε and μ_κ via rate-explicit Wong–Zakai estimates, mixing estimates uniform in ε that would permit rearranging the limits in Corollary 2, and the extension to manifolds with boundary, where a reflection-symmetric doubling construction reduces the reflected problem to the boundaryless theory developed here under suitable evenness conditions on the metric near the boundary.

Acknowledgement

This manuscript is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Grant No. HR00112530225 (COMPASS program).

References

[1] M. Krstic and H.-H. Wang, "Stability of extremum seeking feedback for general nonlinear dynamic systems," *Automatica*, vol. 36, no. 4, pp. 595–601, 2000.

[2] K. B. Ariyur and M. Krstic, *Real-Time Optimization by Extremum-Seeking Control*. Hoboken, NJ: Wiley, 2003.

[3] H.-B. Dürr, M. S. Stanković, C. Ebenbauer, and K. H. Johansson, "Lie bracket approximation of extremum seeking systems," *Automatica*, vol. 49, no. 6, pp. 1538–1552, 2013.

[4] A. Scheinker and M. Krstic, "Extremum seeking with bounded update rates," *Systems & Control Letters*, vol. 63, pp. 25–31, 2014.

[5] H.-B. Dürr, M. S. Stanković, K. H. Johansson, and C. Ebenbauer, "Extremum seeking on submanifolds in the Euclidean space," *Automatica*, vol. 50, no. 10, pp. 2591–2596, 2014.

[6] V. Grushkovskaya, A. Zuyev, and C. Ebenbauer, "On a class of generating vector fields for the extremum seeking problem: Lie bracket approximation and stability properties," *Automatica*, vol. 94, pp. 151–160, 2018.

[7] R. Suttner and S. Dashkovskiy, "Extremum seeking for nonholonomic systems," *IEEE Trans. Autom. Control*, vol. 65, no. 6, pp. 2611–2618, 2020.

[8] T.-S. Chiang, C.-R. Hwang, and S.-J. Sheu, "Diffusion for global optimization in $\mathbb{R}^n$," *SIAM J. Control Optim.*, vol. 25, no. 3, pp. 737–753, 1987.

[9] S. B. Gelfand and S. K. Mitter, "Recursive stochastic algorithms for global optimization in $\mathbb{R}^d$," *SIAM J. Control Optim.*, vol. 29, no. 5, pp. 999–1018, 1991.

[10] R. Holley, S. Kusuoka, and D. Stroock, "Asymptotics of the spectral gap with applications to the theory of simulated annealing," *J. Funct. Anal.*, vol. 83, no. 2, pp. 333–347, 1989.

[11] V. S. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[12] S.-J. Liu and M. Krstic, *Stochastic Averaging and Stochastic Extremum Seeking*. London, U.K.: Springer, 2012.

[13] C. Manzie and M. Krstic, "Extremum seeking with stochastic perturbations," *IEEE Trans. Autom. Control*, vol. 54, no. 3, pp. 580–585, 2009.

[14] M. Girolami and B. Calderhead, "Riemann manifold Langevin and Hamiltonian Monte Carlo methods," *J. Roy. Statist. Soc. Ser. B*, vol. 73, no. 2, pp. 123–214, 2011.

[15] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton Univ. Press, 2008.

[16] E. J. McShane, "Stochastic differential equations and models of random processes," in *Proc. Sixth Berkeley Symp. Math. Statist. Probab.*, vol. 3, 1972, pp. 263–294.

[17] E. Wong and M. Zakai, "On the convergence of ordinary integrals to stochastic integrals," *Ann. Math. Statist.*, vol. 36, no. 5, pp. 1560–1564, 1965.

[18] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, 2nd ed. Amsterdam, The Netherlands: North-Holland, 1989.

[19] H. J. Sussmann, "Limits of the Wong–Zakai type with a modified drift term," in *Stochastic Analysis*. New York, NY: Academic Press, 1991, pp. 475–493.

[20] P. K. Friz and N. B. Victoir, *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[21] V. N. Kolokoltsov, *Semiclassical Analysis for Diffusions and Stochastic Processes*. Berlin, Germany: Springer, 2000.

[22] B. Helffer, M. Klein, and F. Nier, "Quantitative analysis of metastability in reversible diffusion processes via a Witten complex approach," *Mat. Contemp.*, vol. 26, pp. 41–85, 2004.

[23] J.-M. Coron, *Control and Nonlinearity*. Providence, RI: Amer. Math. Soc., 2007.

[24] S. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2009.
[25] J. C. Mattingly, A. M. Stuart, and D. J. Higham, "Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise," *Stochastic Process. Appl.*, vol. 101, no. 2, pp. 185–232, 2002.
[26] J. M. Lee, "Introduction to Riemannian Manifolds," Springer, 2018
[27] S. J. Axler, "Measure, Integration & Real Analysis". Springer International Publishing, 2020.

APPENDIX

Lemma 2. *There exists $\varepsilon_0 > 0$ such that for every $\varepsilon \in (0, \varepsilon_0]$ and every $z \in \mathcal{M}$ there exist $\rho > 0$, a nonempty open $\mathcal{O} \subset \mathcal{M}$, and $c > 0$ such that, for every Borel set A,*

$$P_\varepsilon(y, A) \;\geq\; c\,\mathrm{vol}(A \cap \mathcal{O}), \quad \forall y \in B(z, \rho). \tag{39}$$

Proof. Let $\widetilde{\Psi}_\varepsilon$ denote the time-1 solution map of (32) with the branch frozen to $\iota = j$. It follows that $\widetilde{\Psi}_\varepsilon$ is jointly smooth on $\mathcal{M} \times \mathbb{R}^{2r}$ and agrees with Ψ_ε on the closed cone

$$\mathcal{C} := \{\Delta : \Delta^{2i-1} \geq 0, \Delta^{2i} \geq 0 \ \forall i\}.$$

First, we prove the following claim.

Claim 1. *There exists a neighborhood $\mathcal{D} \subset \mathbb{R}^{2r}$ containing the origin and $\varepsilon_0 > 0$ such that, for all $\varepsilon \in (0, \varepsilon_0]$, the map $\Delta \mapsto \widetilde{\Psi}_\varepsilon(x, \Delta)$ is a submersion on $\mathcal{D}$, for all $x \in \mathcal{M}$.*

Proof: At $\Delta = 0$, the input in (32) vanishes identically, so $\widetilde{\Psi}_\varepsilon(x, 0) = \Phi_1^{\varepsilon \bar{f}_0}(x)$, where $\Phi_s^{\varepsilon \bar{f}_0}$ is the flow of the vector field $\varepsilon \bar{f}_0$ up to time s. Define $y_0(s) := \Phi_s^{\varepsilon \bar{f}_0}(x)$, and denote by $y(s; \Delta)$ the solution of (32) with frozen branch, so that

$$\widetilde{\Psi}_\varepsilon(x, \Delta) = y(1; \Delta).$$

Define the sensitivity with respect to the ℓ-th component of Δ

$$v_\ell(s) := \partial_{\Delta^\ell} y(s; \Delta)|_{\Delta=0},$$

which exists by smooth dependence of solutions on parameters. Differentiating (32) in Δ^ℓ at $\Delta = 0$, every input term dies except the ℓ-th, since the input is linear in Δ, which yields the *variational equation* for v_ℓ along the curve y_0

$$v_\ell'(s) = \varepsilon\, \mathrm{D}\bar{f}_0(y_0(s))\, v_\ell(s) + \tfrac{1}{\sqrt{\kappa}}\, g_\ell(y_0(s))\, (U_\kappa^j)'(s), \tag{40}$$

with $v_\ell(0) = 0$. Note that (40) is a non-homogeneous linear time-varying equation for the sensitivity vector v_ℓ. The associated state-transition from time s to time 1 is the exactly differential $\mathrm{D}\Phi_{1-s}^{\varepsilon \bar{f}_0}$ of the drift flow along y_0. Hence, the variation-of-constants formula gives

$$\partial_{\Delta^\ell} \widetilde{\Psi}_\varepsilon(x, 0) = v_\ell(1) = \frac{1}{\sqrt{\kappa}} \int_0^1 \mathrm{D}\Phi_{1-s}^{\varepsilon \bar{f}_0}\, g_\ell(y_0(s))\, (U_\kappa^j)'(s)\, \mathrm{d}s. \tag{41}$$

Since $\int_0^1 \mathrm{d}(U_\kappa^j) = 1$, direct computation using (41) gives that

$$\lim_{\varepsilon \searrow 0} \partial_{\Delta^\ell} \widetilde{\Psi}_\varepsilon(x, 0) = \frac{1}{\sqrt{\kappa}} g_\ell(x), \tag{42}$$

and, by compactness of $\mathcal{M}$, the convergence is uniform in x. On the other hand, for any $0 \neq \eta \in T_x^*\mathcal{M}$, we obtain that

$$\begin{aligned} \sum_\ell \langle \eta, g_\ell(x) \rangle^2 &= \sum_i (F_1^2 + F_2^2)(\phi(x)) \langle \eta, f_i(x) \rangle^2 \\ &= \sum_i \langle \eta, f_i(x) \rangle^2 > 0, \end{aligned}$$

from (1), where the inner product is taken with respect to the Riemmanian metric defined by the family $\mathcal{F}$ and we used the fact that $F_1^2 + F_2^2 = 1$. Consequently, the smallest singular value of the Jacobian

$$\lim_{\varepsilon \searrow 0} \partial_\Delta \widetilde{\Psi}_\varepsilon(x, 0)$$

is uniformly bounded below on $\mathcal{M}$ by some constant $\sigma_0 > 0$. Since the frozen-branch equation (32) depends smoothly on (x, Δ, ε) up to and including $\varepsilon = 0$, the map

$$(x, \Delta, \varepsilon) \mapsto \partial_\Delta \widetilde{\Psi}_\varepsilon(x, \Delta)$$

is continuous on the compact set $\mathcal{M} \times \overline{B}_1(0) \times [0, 1]$, hence uniformly continuous, and its value at $(x, 0, 0)$ coincides with the limit computed above. Since the smallest singular value is 1-Lipschitz with respect to the operator norm, there exist $\delta \in (0, 1]$ and $\varepsilon_0 \in (0, 1]$ such that

$$\sigma_{\min}\big(\partial_\Delta \widetilde{\Psi}_\varepsilon(x, \Delta)\big) \;\geq\; \tfrac{1}{2}\, \sigma_0,$$

for all $x \in \mathcal{M}$, $\Delta \in \mathcal{D} := B_\delta(0)$, and $\varepsilon \in (0, \varepsilon_0]$. In particular, the differential of $\widetilde{\Psi}_\varepsilon(x, \cdot)$ is surjective at every point of $\mathcal{D}$, which proves the claim.

Now fix $\varepsilon \in (0, \varepsilon_0]$ and $z \in \mathcal{M}$, and set $n := \dim \mathcal{M}$. Pick $\Delta_* \in \mathcal{D} \cap \mathrm{int}(\mathcal{C})$, which is possible because $\mathcal{D}$ is a neighborhood of the origin, and $(t, \ldots, t) \in \mathrm{int}\, \mathcal{C}$, for all $t > 0$. From Claim 1, $\partial_\Delta \widetilde{\Psi}_\varepsilon(z, \Delta_*)$ is surjective. Let ζ be a chart of $\mathcal{M}$ around $\widetilde{\Psi}_\varepsilon(z, \Delta_*)$ and, after permuting the components of Δ, split $\Delta = (\Delta', \Delta'') \in \mathbb{R}^n \times \mathbb{R}^{2r-n}$ so that

$$\partial_{\Delta'}(\zeta \circ \widetilde{\Psi}_\varepsilon)(z, \Delta_*)$$

is invertible. For (y, Δ) in the domain of $\zeta \circ \widetilde{\Psi}_\varepsilon$, define

$$\Xi_y(\Delta) := \big(\zeta(\widetilde{\Psi}_\varepsilon(y, \Delta)),\, \Delta''\big), \quad \Xi(y, \Delta) := \big(y,\, \Xi_y(\Delta)\big).$$

Then, $\mathrm{D}\Xi$ has identity in the y-block, so $\det \mathrm{D}\Xi(y, \Delta) = \det \mathrm{D}\Xi_y(\Delta)$, while $\mathrm{D}\Xi_y$ has identity in the Δ''-block, so $\det \mathrm{D}\Xi_y(\Delta) = \det \partial_{\Delta'}(\zeta \circ \widetilde{\Psi}_\varepsilon)(y, \Delta)$, which is nonzero at the point (z, Δ_*) by the choice of splitting. Hence, by the inverse function theorem, Ξ restricts to a diffeomorphism of a product neighborhood $B(z, \rho_0) \times V$ of (z, Δ_*), where V is an open ball around Δ_* chosen so that $\overline{V} \subset \mathrm{int}\, \mathcal{C}$, onto an open set W that contains the point

$$(z, v_0, w_0) := \Xi(z, \Delta_*).$$

Since Ξ fixes the first coordinate, the map Ξ_y is injective on V with nonvanishing Jacobian for each $y \in B(z, \rho_0)$, and, therefore, it is a diffeomorphism of V onto the slice $W_y := \{(v, w) : (y, v, w) \in W\}$, and the first component of Ξ_y gives

$$\widetilde{\Psi}_\varepsilon\big(y, \Xi_y^{-1}(v, w)\big) = \zeta^{-1}(v), \qquad (v, w) \in W_y. \tag{43}$$

Since W is open, we let $\rho \in (0, \rho_0]$ and $\beta > 0$ such that

$$\overline{B(z, \rho)} \times \overline{Q} \subset W,$$

where $Q := B(v_0, \beta) \times B(w_0, \beta)$. In particular, we have that $Q \subset W_y$, for every $y \in B(z, \rho)$. We now define the set

$$\mathcal{O} := \zeta^{-1}\big(B(v_0, \beta)\big),$$

which is a nonempty open subset of $\mathcal{M}$, and fix $y \in B(z, \rho)$ and a Borel set A. Since $\overline{V} \subset \mathcal{C}$, the maps $\Psi_\varepsilon(y, \cdot)$ and $\widetilde{\Psi}_\varepsilon(y, \cdot)$

coincide on V, and the Gaussian density of Δ is bounded below on the bounded set V by some $\gamma > 0$, so

$$P_\varepsilon(y, A) \;\geq\; \gamma \operatorname{Leb}\big(\{\Delta \in V : \widetilde{\Psi}_\varepsilon(y, \Delta) \in A\}\big),$$

where $\operatorname{Leb}(K)$ denotes the Lebesgue measure of a measurable set K. Changing variables via the diffeomorphism $\Xi_y^{-1} \colon Q \to \Xi_y^{-1}(Q) \subset V$ and using (43), we compute that

$$\begin{aligned}\operatorname{Leb}\big\{\Delta \in V :& \widetilde{\Psi}_\varepsilon(y, \Delta) \in A\big\} \\ &\geq \int_Q \mathbf{1}_{\zeta(A \cap \mathcal{O})}(v)\, \big|\det \mathrm{D}\Xi_y^{-1}(v, w)\big|\, \mathrm{d}v\, \mathrm{d}w \\ &\geq c_1 \operatorname{Leb}\big(B(w_0, \beta)\big) \operatorname{Leb}\big(\zeta(A \cap \mathcal{O})\big),\end{aligned}$$

where $c_1 > 0$ is a lower bound for the continuous positive function $(y, v, w) \mapsto |\det \mathrm{D}\Xi_y^{-1}(v, w)|$ on the compact set $\overline{B(z, \rho)} \times \overline{Q}$. Finally, since ζ is a fixed chart, there is $c_2 > 0$ with $\operatorname{Leb}\big(\zeta(A \cap \mathcal{O})\big) \geq c_2 \operatorname{vol}(A \cap \mathcal{O})$. It follows that the inequality (39) holds with $c := \gamma\, c_1\, c_2 \operatorname{Leb}\big(B(w_0, \beta)\big)$.

□

The localization in Lemma 2 can be made uniform, i.e., there exist $\bar\rho, \bar\beta > 0$ such that, for every $\varepsilon \in (0, \varepsilon_0]$ and $z \in \mathcal{M}$, the conclusion (39) holds with $\rho := \bar\rho$ and with $\mathcal{O}$ containing the ball $B\big(\widetilde{\Psi}_\varepsilon(z, \Delta_*), \bar\beta\big)$. In particular, only c depends on z and ε. Indeed, the neighborhoods produced by the inverse function theorem in the proof are controlled by a bound on the second derivatives of Ξ and a lower bound on $|\det \partial_{\Delta'}(\zeta \circ \widetilde{\Psi}_\varepsilon)|$ at the base point. The former of these two is uniform on the compact set $\mathcal{M} \times \overline{V} \times [0, \varepsilon_0]$ for a fixed finite atlas, and the latter follows from the Claim via the Cauchy–Binet formula, which produces, for each z, a splitting $\Delta = (\Delta', \Delta'')$ with $|\det \partial_{\Delta'}(\zeta \circ \widetilde{\Psi}_\varepsilon)(z, \Delta_*)|$ bounded below by a constant depending only on σ_0, n, r, and the atlas. A quantitative form of the inverse function theorem [**?**, Ch. 1] then yields ρ_0, V, ρ, β of a size depending only on these bounds.

Lemma 3. *For every $\varepsilon > 0$, $z \in \mathcal{M}$, and $\rho > 0$, there exist $N \in \mathbb{N}$ and $\delta_0 > 0$ such that*

$$\inf_{x \in \mathcal{M}} P_\varepsilon^N(x, B(z, \rho)) \geq \delta_0.$$

Proof. For each fixed $\Delta \in \mathbb{R}^{2r}$, the branch assignment in (4) is constant on the step, so (32) is a smooth ODE whose data are bounded uniformly over the finitely many branch patterns. Hence, a Grönwall estimate gives

$$|y(s) - x| \leq C(\varepsilon + \|\Delta\|)$$

for $\|\Delta\| \leq 1$, with C independent of x, Δ, and the branch. Integrating (32), freezing the coefficients at x, and using that $\int_0^1 (U_\kappa^\iota)' = 1$ for both $\iota \in \{1, 2\}$ gives, in ambient coordinates, (**??**) uniformly in $x \in \mathcal{M}$, $\|\Delta\| \leq 1$, and the branch. We stress that (**??**) asserts no continuity of Ψ_ε in Δ. Since the g_ℓ span each tangent space uniformly, for each x and each unit direction there is an increment with $\|\Delta\| \leq M\varepsilon$ (M fixed, large enough to dominate the drift) realizing a displacement of order ε in that direction. Hence any two points can be joined by a chain of $N \asymp \operatorname{diam}(\mathcal{M})/\varepsilon$ steps tracking a smooth path inside a tube of radius $\rho/2$; this is a fully actuated (bracket-free) instance of the Chow–Rashevskii construction [23]. Each step succeeds on an open set of increments of Gaussian mass bounded below, tracking is robust to perturbations within these sets, and shorter chains can be padded to a common length since the chain can idle near a point with per-step probability bounded below (choose Δ to approximately cancel the drift, possible by (1)). Independence gives $P_\varepsilon^N(x, B(z, \rho)) \geq \delta_x > 0$; the map $x \mapsto P_\varepsilon^N(x, B(z, \rho))$ is lower semicontinuous by the Feller property and the portmanteau theorem, so compactness upgrades the bound to a uniform $\delta_0 > 0$. □