Author/s:
TARINGOO, F;Dower, PM;Nesic, D;Tan, Y

Title:
Optimization Methods on Riemannian Manifolds via Extremum Seeking Algorithms

Date:
2018

Citation:
TARINGOO, F., Dower, P. M., Nesic, D. & Tan, Y. (2018). Optimization Methods on Riemannian Manifolds via Extremum Seeking Algorithms. SIAM Journal on Control and Optimization, 56 (5), pp.3867-3892. https://doi.org/10.1137/15M1018022.

OPTIMIZATION METHODS ON RIEMANNIAN MANIFOLDS VIA EXTREMUM SEEKING ALGORITHMS *

FARZIN TARINGOO[†], PETER M. DOWER[†], DRAGAN NEŠIĆ[†] AND YING TAN [†]

Abstract. This paper formulates the problem of Extremum Seeking for optimization of cost functions defined on Riemannian manifolds. We extend the conventional extremum seeking algorithms for optimization problems in Euclidean spaces to optimization of cost functions defined on smooth Riemannian manifolds. This problem falls within the category of online optimization methods. We introduce the notion of geodesic dithers which is a perturbation of the optimizing trajectory in the tangent bundle of the ambient state manifolds and obtain the extremum seeking closed loop as a perturbation of the averaged gradient system. The main results are obtained by applying closeness of solutions and averaging theory on Riemannian manifolds. The main results are further extended for optimization on Lie groups. Numerical examples on the Stiefel manifold $V_{3,2}$ and the Lie group $SE(3)$ are presented at the end of the paper.

Key words. Extremum Seeking Control, Riemannian Manifolds.

AMS subject classifications. 49M99, 49M30, 90C56, 65K10

1. Introduction.

Optimization on manifolds is an important research area in optimization theory, see [2, 38, 44]. In this class of problems, the underlying optimization space is a manifold and consequently the analysis differs from the standard optimization algorithms in Euclidean spaces. Numerical techniques and methods for optimization on manifolds should guarantee that in each step an optimizer is an element of the search space which is a manifold. Hence, optimization methods on manifolds are closely related to geometry of manifolds, see [2, 13, 26, 44].

As known, smooth manifolds can be embedded in high dimensional Euclidean spaces (Whitney Theorem) [3]. This means that optimization on manifolds can be carried out as constrained optimization problems in Euclidean spaces with sufficiently large dimensions. However, the corresponding embeddings for each particular manifold may not be available and algorithms developed on manifolds may be more efficient in terms of convergence speed and computational burden [34]. The algorithms investigated in this field range from simple line search methods to more sophisticated algorithms such as trust region methods, see for example [5, 46].

Optimization on manifolds can arise in a wide range of applications where the search space is constrained. Its applications may appear in signal processing [27], robotics [16], and statistics [11]. The main underlying assumption in most of the numerical algorithms presented for optimization on manifolds is that the cost function is available. This simplifies the implementation of numerical algorithms since various numerical methods can be applied to calculate the sensitivity of cost functions to obtain numerical optimization trajectories. However, in many problems, cost functions may not be given in a well defined closed form. This necessitates generating a class of numerical methods which do not explicitly depend on the closed form of cost functions derivatives [37, 39]. This is the main motivation of this paper and one of the major contributions of this paper is presenting algorithms to address this class of problems on Riemannian manifolds.

---

[*] This work was supported by the Australian Research Council Discovery Project DP120101144.

[†] Department of Electrical and Electronic Engineering, The University of Melbourne, {ftaringoo, pdower, dnesic, yingt}@unimelb.edu.au.

Extremum seeking is a class of on-line or real-time optimization methods for optimization of the steady-state behavior of dynamical systems [22]. This method is applied for optimization of both static functions and dynamical systems where the optimizer does not have a model of the cost/utility function or dynamical equations. In other words, either cost functions or dynamical systems are unknown for the optimization procedure and the optimization algorithm should be able to converge to a vicinity of a local optimizer, see [8, 14, 15, 19, 22, 29–31, 40, 43]. In this paper we only consider the extremum seeking algorithms for optimization of static cost functions and consequently we assume that cost functions are not available for the optimization procedure.

Extremum seeking finds application in a vast area of dynamical systems including robotics and mechanical systems. As is known, mechanical systems are mathematically modeled on manifolds which do not necessarily possess vector space properties, see [1, 6, 7]. Traditionally, extremum seeking systems have been analyzed in the class of unconstrained optimization methods on $\mathbb{R}^n$ where the vector space properties of Euclidean spaces simplify the analysis.

In a more general framework, the underlying Euclidean spaces can be replaced by Riemannian manifolds. To this end, we define a class of online optimization methods where the optimization trajectories lie on manifolds. As an example, the standard gradient descent and Newton methods are extended to their geometric versions by employing the notion of geodesics on Riemannian manifolds, see [26, 38]. In this paper this step is done for extremum seeking algorithms by introducing the so-called *geodesic dithers* which are the geometric versions of dither signals in standard extremum seeking framework, see [22, 40]. This enables us to relax the metric constraint of extremum seeking algorithms in Euclidean spaces and generalize the extremum seeking algorithms for a large class of optimization problems on generic Riemannian manifolds. By employing the geodesic dithers, we guarantee that during the optimization phase optimizing trajectories always lie on state manifolds. To analyze the behavior of the closed loop system, we employ averaging techniques developed for dynamical systems on Riemannian manifolds, [7, 28], and apply results of closeness of solutions to obtain closeness of optimizing trajectories to local optimizers.

A recent version of extremum seeking algorithms for optimization of cost functions on submanifolds of Euclidean spaces appeared in [9]. In [9], the authors analyzed an extremum seeking algorithm based on a Lie bracket approach. The method presented in [9] is based upon embeddings of manifolds in Euclidean spaces and employs the Riemannian metric induced by the ambient Euclidean space. In this paper we present an algorithm which can be applied to cost functions defined on manifolds with generic Riemannian metrics (not necessarily Euclidean). This distinction is clarified via the first example by optimization of a cost function on a Stiefel manifold with a non-Euclidean metric. The algorithm presented in this paper does not require specific embeddings of manifolds in Euclidean spaces and, hence can be used for a large class of optimization problems on Riemannian manifolds.

In terms of exposition, Section 2 presents some mathematical preliminaries needed for the analysis of the paper. Section 3 presents the extremum seeking problems on Riemannian manifolds and in Section 4 we extend the extremum seeking algorithm for optimization on Lie groups. In Sections 5 and 6 we present simple optimization examples on $V_{3,2}$ and $SE(3)$ by applying the extremum seeking methods developed in Section 3.

2. Preliminaries. In this section we provide the differential geometric material which is necessary for the analysis presented in the rest of the paper. We define some of the frequently used symbols in Table 2. For the complete definitions see [25] and [42].

Table 2.1
Symbols and Their Descriptions

Symbol	Description
(M, g^M)	Riemannian manifold with metric g^M
G	Lie Group
$\star$	Lie group operation
$\mathfrak{X}(M)$	space of smooth time invariant vector fields on M
$\mathfrak{X}(M \times \mathbb{R})$	space of smooth time varying vector fields on M
$\mathfrak{X}(\mathbb{R} \times M)$	space of smooth parameter varying vector fields on M
∇	Levi-Civita connection
$\exp_x$	restricted exponential map
T_xM	tangent space at $x \in M$
TM	tangent bundle of M
T_x^*M	cotangent space at $x \in M$
T^*M	cotangent bundle of M
$\frac{\partial}{\partial x_i}$	basis tangent vectors at $x \in M$
$\frac{\partial}{\partial g_i}$	representation of basis tangent vectors in the Lie algebra $\mathcal{L}$
dx_i	basis cotangent vectors at $x \in M$
$f(x,t)$	time varying vector fields on M
$\|\|\cdot\|\|_{g^M}$	Riemannian norm
$g^M(\cdot,\cdot)$	Riemannian metric on M
$d(\cdot,\cdot)$	Riemannian distance on M
$C^\infty(M)$	Space of smooth functions on M
Φ_f	flow associated with f
TF	pushforward of F
T_xF	pushforward of F at x
$\mathbb{R}_{>0}$	$(0,\infty)$
$\mathbb{R}_{\geqslant 0}$	$[0,\infty)$

2.1. Geodesic Curves. Geodesics are defined [17] as length minimizing curves on Riemannian manifolds which satisfy

$$\nabla_{\dot{\gamma}(t)}\dot{\gamma}(t) = 0,$$

where $\gamma(\cdot)$ is a geodesic curve on (M, g^M) and $\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ is the *Levi-Civita* connection on M, see [23].

Definition 2.1 ([23], P. 72). *The restricted exponential map is defined by*

$$\exp_x : T_xM \to M, \;\; \exp_x(v) = \gamma_v(1), v \in T_xM,$$

where $\gamma_v(1)$ *is the unique maximal geodesic [23, P. 59] initiating from* x *with the velocity* v *up to one.*

Throughout, *restricted exponential maps* are referred to as *exponential maps.* An open ball of radius $\delta > 0$ and centered at $0 \in T_xM$ in the tangent space at x is denoted by $B_\delta(0) \doteq \{v \in T_xM \mid \ ||v||_{g^M} < \delta\}$. Similarly, the corresponding closed ball is denoted by $\overline{B}_\delta(0)$. Using the local diffeomorphic property of exponential maps, the corresponding geodesic ball centred at x is defined as follows.

Definition 2.2 ([23], p. 76). *A normal neighborhood around* $x \in M$ *is any open neighborhood of* x *which is a diffeomorphic image of a star shaped neighborhood, see [25], of* $0 \in T_xM$ *under the exponential map* $\exp_x$.

Lemma 2.3 ([23], Lemma 5.10). *For any* $x \in M$*, there exists a neighborhood* $B_\delta(0)$ *in* T_xM *on which* $\exp_x$ *is a diffeomorphism onto* $\exp_x(B_\delta(0)) \subset M$.

Definition 2.4 ([23], p. 76). *In a neighborhood of* $x \in M$*, where* $\exp_x$ *is a local diffeomorphism (this neighborhood always exists by Lemma 2.3), a geodesic ball of radius* $\delta > 0$ *is denoted by* $\exp_x(B_\delta(0)) \subset M$*. The corresponding closed geodesic ball is* $\exp_x(\overline{B}_\delta(0))$.

Definition 2.5. *The injectivity radius of* M *is*

$$\iota(M) \doteq \inf_{x \in M} \iota(x),$$

where $\iota(x) \doteq \sup\{r \in \mathbb{R}_{\geqslant 0} | \exp_x \textit{ is diffeomorphic onto} \exp_x(B_r(0))\}$.

Definition 2.6. *The metric ball with respect to* d *on* (M, g^M) *is defined by*

$$B(x, r) \doteq \{y \in M \mid \ d(x, y) < r\}.$$

The following lemma reveals a relationship between normal neighborhoods and metric balls on (M, g^M).

Lemma 2.7 ([33], p. 122). *Given any* $\epsilon \in \mathbb{R}_{>0}$ *and* $x \in M$*, suppose that* $\exp_x$ *is a diffeomorphism on* $B_\epsilon(0) \subset T_xM$*. If* $B(x, r) \subset \exp_x B_\epsilon(0)$ *for some* $r \in \mathbb{R}_{>0}$*, then*

$$\exp_x B_r(0) = B(x, r).$$

This paper focuses on dynamical systems governed by differential equations on a connected n dimensional Riemannian manifold (M, g^M). Locally these differential equations are defined by (see [25])

$$\begin{aligned} \dot{x}(t) &= f(x(t), t), \quad f \in \mathfrak{X}(M \times \mathbb{R}), \\ x(0) &= x_0 \in M, t \in [t_0, t_f) \subset \mathbb{R}, \end{aligned}$$

where t_f can be infinite (complete vector fields). The time dependent flow associated with a differentiable time dependent vector field f is a map Φ_f satisfying

$$\begin{aligned} &\Phi_f : [t_0, t_f) \times [t_0, t_f) \times M \to M, \\ &(s_f, s_0, x) \mapsto \Phi_f(s_f, s_0, x) \in M, \end{aligned} \tag{2.1}$$

and $\frac{d\Phi_f(s,s_0,x)}{ds}|_{s=t} = f(\Phi_f(t, s_0, x), t)$. One may show, for a smooth vector field f, the integral flow $\Phi_f(s, t_0, \cdot) : M \to M$ is a local diffeomorphism, see [25].

2.2. Lie groups. As is well known, a Lie group $(G, \star)$ is a Riemannian manifold equipped with operations $g_1 \star g_2$ and g^{-1} where $g_1, g_2, g \in G$ which are smooth in their topologies ($\star$ is the group operation of G), see [21, 45]. We recall that the Lie algebra $\mathcal{L}$ of a Lie group G (see [7], [45]) is the tangent space at the identity element e with the associated Lie bracket defined on the tangent space $\mathcal{L} \doteq T_eG$ of G. A vector field X on G is called *left invariant* if $\forall g_1, g_2 \in G,\quad X(g_1 \star g_2) = T_{g_2}g_1 \star X(g_2)$, where $g\star : G \to G,\ \ g \star (h) = g \star h,\ \ T_hg\star : T_hG \to T_{g\star h}G$ which immediately implies $X(g \star e) = X(g) = T_eg \star X(e)$. Without further confusion, we simply use T_hg as the pushforward of $g\star$ at $h \in G$ instead of $T_hg\star$ in the rest of the paper. For a left invariant vector field X, we define the *exponential map* on Lie groups as follows:

$$\exp : \mathcal{L} \to G, \quad \exp(tX(e)) \doteq \Phi(t, X), t \in \mathbb{R}, \tag{2.2}$$

where $\Phi(t, X)$ is the solution of $\dot{g}(t) = X(g(t)) = T_eg(t)X(e)$ with the boundary condition $g(0) = e$. It may be shown that the solution of $\dot{g}(t) = X(g(t))$ with initial condition $g_0 \in G$ is given by $g_0 \star \exp(tX(e))$ where $\star$ is the group operation of G, see [7]. A Riemannian metric g^G on a Lie group G is left invariant if

$$g^G_{g_2}(X(g_2), Y(g_2)) = g^G_{g_1 \star g_2}(T_{g_2}g_1X(g_2), T_{g_2}g_1Y(g_2)), \quad X, Y \in \mathfrak{X}(G).$$

Analogous to left invariant metrics, right invariant Riemannian metrics on G are defined. A Riemannian metric which is both left and right invariant is called *bi-invariant.* The Levi-Civita connection corresponding to a left invariant Riemannian metric g^G is denoted by $\nabla^{\mathbb{G}}$. It may be shown that the Levi-Civita connection of a left invariant metric is left invariant i.e. (see [7]) $Tg\left(\nabla^{\mathbb{G}}_X Y\right) = \nabla^{\mathbb{G}}_{TgX} TgY$. The following lemma gives a relationship between the exponential maps (2.2) and geodesics on Lie groups.

Lemma 2.8 ([32]). *Assume G is a Lie group which admits a bi-invariant Riemannian metric. Then, for a left invariant vector field X on G we have* $\exp(tX(e)) = \exp_e(tX(e))$, *where* $\exp(tX(e))$ *is the exponential map (2.2) and* $\exp_e(tX(e))$ *is the geodesic emanating from e by velocity $X(e)$.*

3. Optimization on Manifolds and Extremum Seeking. Let us consider the optimization of a smooth function $J : M \to \mathbb{R}_{\geqslant 0}$, where (M, g^M) is an n dimensional smooth Riemannian manifold. Extremum seeking algorithms are a class of online optimization methods developed for minimizing/maximizing smooth functions defined on Euclidean spaces. These methods can be applied to both static and dynamic functions. In this paper we restrict our analysis to static functions defined on Riemannian manifolds. The simplest form of the extremum seeking algorithm is given by minimizing/maximizing a scalar function $J : \mathbb{R} \to \mathbb{R}_{\geqslant 0}$, see [40]. The dither signal $a\sin(\omega t)$ provides a variation of the searching signal $\hat{x}(t)$ in the one dimensional space $\mathbb{R}$. The output $x(t) \in \mathbb{R}$ of the extremum seeking controller at time t is $x(t) = \hat{x}(t) + a\,\sin(\omega\, t)$, where $\hat{x}(t) \in \mathbb{R}$ is the corresponding output of the integrator. The closed loop dynamics in $\hat{x}$ coordinates are described by

$$\dot{\hat{x}}(t) = ka\sin(\omega t)J(\hat{x}(t) + a\sin(\omega t)). \tag{3.1}$$

The next lemma shows that, on average, the extremum seeking is a perturbation of the gradient algorithm.

Lemma 3.1 ([40]). *Consider the extremum seeking scheme given in (3.1). Then, the average vector field of the extremum seeking algorithm is in a perturbation form of the gradient algorithm* $\dot{\hat{x}}(t) = \frac{ka^2}{2}\frac{\partial J}{\partial \hat{x}}|_{\hat{x}=\hat{x}(t)}$ *in $\hat{x}$ coordinates.*

The proof is based on the Taylor expansion of the cost function J in $\hat{x}$ coordinates. Here, we extend the framework above for optimization of cost functions defined on finite dimensional Riemannian manifolds. The main challenge is to introduce a class of dither signals which perturb the optimizer $\hat{x}$ without violating the restrictions imposed by the ambient manifolds. This is done by employing the so-called *geodesic dithers* as follows.

Consider an n dimensional Riemannian manifold (M, g^M). For any $x \in M$, we consider the following local time-varying perturbation

$$x_p(t) = \exp_x\left(\sum_{i=1}^{n} a_i \sin(\omega_i t)\frac{\partial}{\partial x_i}\right), \quad 0 < \omega_i, a_i, \tag{3.2}$$

where $\{\frac{\partial}{\partial x_i}\}, \;\; i = 1, \cdots, n$, are the base tangent vectors in the tangent space at x. As formalized in Definition 2.1, $\exp_x v$ is a geodesic emanating from $x \in M$ with velocity $v \in T_xM$. In this case we perturb the optimizing trajectory in n different coordinates on M with different frequencies $\omega_i, \;\; i = 1, \cdots, n$. Motivated by the classical extremum seeking, we present a time-varying *extremum seeking vector field* $f(\hat{x}, t) \in T_{\hat{x}}M$ for optimization on (M, g^M) which is locally given by

$$f(\hat{x}, t) \doteq k\sum_{i=1}^{n} a_i \sin(\omega_i t) J\left(\exp_{\hat{x}} \sum_{j=1}^{n} a_j \sin(\omega_j t)\frac{\partial}{\partial x_j}\right)\frac{\partial}{\partial x_i} \in T_{\hat{x}}M. \tag{3.3}$$

In this paper we assume that the optimization problem is to minimize a cost function, hence without loss of generality assume $k = -1$. Finally, the optimizing trajectory $\hat{x}(\cdot)$ is a solution of the time dependent differential equation

$$\dot{\hat{x}}(t) = f(\hat{x}(t), t) \in T_{\hat{x}(t)}M. \tag{3.4}$$

The closed loop system (3.4) is called the *extremum seeking system* on (M, g^M). Note that t appears as a parameter in $\exp_{\hat{x}}\left(\sum_{i=1}^{n} a_i \sin(\omega_i t)\frac{\partial}{\partial x_i}\right)$. That is to say

$$\exp_{\hat{x}}\left(\sum_{i=1}^{n} a_i \sin(\omega_i t)\frac{\partial}{\partial x_i}\right) = \exp_{\hat{x}}\left(\eta\sum_{i=1}^{n} a_i \sin(\omega_i t)\frac{\partial}{\partial x_i}\right)|_{\eta=1},$$

where η and t are independent. Also note that the vector field (3.3) does not require any information about the gradient of J. However, cost function J should be observable for the optimizing algorithm at time t.

The next lemma proves that on compact Riemannian manifolds, for all $x \in M$, one may choose parameters $a_i > 0$ sufficiently small such that $\exp_x\left(\sum_{i=1}^{n} a_i \sin(\omega_i t)\frac{\partial}{\partial x_i}\right) \in \exp_x\left(B_{\iota(x)}(0)\right)$. These results will be employed to obtain the Taylor expansion of cost functions on Riemannian manifolds along geodesics.

Lemma 3.2. *Consider the geodesic dither introduced by (3.2) on a smooth n dimensional compact Riemannian manifold (M, g^M). Then for all $x \in M$, we may select $a_i > 0, \;\; i = 1, \cdots, n$, such that for all $t \in \mathbb{R}$, $\exp_x\left(\sum_{i=1}^{n} a_i \sin(\omega_i t)\frac{\partial}{\partial x_i}\right) \in \exp_x\left(B_{\iota(x)}(0)\right)$.*

Proof. Since M is compact and smooth then $\iota(M)$ is strictly positively bounded from below, see [20]. Hence, there exists $\kappa \in \mathbb{R}_{>0}$ such that for all $x \in M$, $\kappa < \iota(x)$. The Riemannian norm of the dither signal is given by

$$\left\|\sum_{i=1}^{n} a_i \sin(\omega_i t)\frac{\partial}{\partial x_i}\right\|^2_{g^M} = \sum_{i,j=1}^{n} a_i a_j \sin(\omega_i t)\sin(\omega_j t) g^M\left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}\right).$$

Since M is compact, $\sum_{i,j=1}^{n} g^M(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j})$ attains its maximum on M. Hence, $a_i > 0,\ \ i = 1, \cdots, n$, may be selected sufficiently small such that $|| \sum_{i=1}^{n} a_i \sin(\omega_i t) \frac{\partial}{\partial x_i} ||_{g^M}^2 \leqslant \kappa^2,\ \ \forall t \in \mathbb{R}$, which completes the proof. □

We adopt the following assumption for the cost function J on (M, g^M) and the dither frequencies $\omega_i,\ \ i = 1, \cdots, n$. This assumption is compatible with the main assumption on dither frequencies in [12] for multi-agent extremum seeking algorithms.

ASSUMPTION 1. *Cost function $J : M \to \mathbb{R}_{\geqslant 0}$ is smooth and locally positive definite in a neighborhood of a unique local minimum $x^* \in M$, where $J(x^*) = 0$. The dither frequencies are $\omega_i = \omega\bar{\omega}_i$, where $\bar{\omega}_i$ is rational, $\bar{\omega}_i \neq \bar{\omega}_j$, $2\bar{\omega}_i \neq \bar{\omega}_j, j \neq i$ and $\bar{\omega}_i \neq \bar{\omega}_j + \bar{\omega}_k$ for distinct i, j, k, where $\omega_i, \omega \in \mathbb{R}_{>0}, i, j, k \in 1, \cdots, n$.*

Here we introduce the *gradient and average systems* which correspond to the extremum seeking vector field (3.3) on (M, g^M). For the smooth function $J : M \to \mathbb{R}_{\geqslant 0}$, in this paper, the gradient system is defined by

$$\dot{x}(t) = -\sum_{i=1}^{n} \nabla_{\frac{\partial}{\partial x_i}} J(x(t)) \frac{\partial}{\partial x_i}, \tag{3.5}$$

where the set $\{\frac{\partial}{\partial x_i}, i = 1, \cdots, n\}$ is a basis of T_xM. We note that the formal definition of the gradient *gradJ* of J is given by [17] as

$$dJ(X) = g^M(\mathrm{grad}J, X), \quad X \in \mathfrak{X}(M), \tag{3.6}$$

where dJ is the one form differential of J locally given by $dJ = \sum_{i=1}^{n} \frac{\partial J}{\partial x_i} dx_i \in T_x^*M$. Note that the existence of grad$J$ in (3.6) is implied by an application of Riesz representation theorem since dJ belongs to the dual space T_x^*M and $g^M(\cdot, \cdot)$ defines an inner product on T_xM, see [35]. In this case,

$$\mathrm{grad}J(x) = \sum_{i,j=1}^{n} g^{ij}(x) \nabla_{\frac{\partial}{\partial x_i}} J(x) \frac{\partial}{\partial x_j},$$

where $[g^{ij}] = [g_{ij}]^{-1}$. Hence, the formal gradient system corresponding to J is $\dot{x}(t) = -\sum_{i,j=1}^{n} g^{ij}(x(t)) \nabla_{\frac{\partial}{\partial x_i}} J(x(t)) \frac{\partial}{\partial x_j}$. However, in this paper, we adopt the terminology that the gradient system of J refers to (3.5). The *scaled* version of the gradient system (3.5) is locally given as

$$\dot{x}(t) = \acute{f}(x(t)) \doteq -\sum_{i=1}^{n} \frac{a_i^2}{2} \nabla_{\frac{\partial}{\partial x_i}} J(x(t)) \frac{\partial}{\partial x_i}, a_i > 0. \tag{3.7}$$

With no further confusion we refer to the scaled gradient system as the gradient system. For a periodic time varying vector field $f(x, t)$, the averaged dynamical system is defined as follows

$$\dot{x}(t) = \hat{f}(x(t)) \doteq \frac{1}{T} \int_0^T f(x(t), \tau) d\tau, \tag{3.8}$$

where T is the period of f, i.e. $f(x, t) = f(x, t + T)$. The following lemma proves that, on average, the extremum seeking system (3.4) is a perturbation of the scaled gradient system of J.

LEMMA 3.3. *Consider the extremum seeking system in (3.4) on a compact Riemannian manifold (M, g^M) where the optimizing trajectory is perturbed by the geodesic* *dither presented in (3.2). Then, subject to Assumption 1, the averaged dynamical system of (3.3) is a perturbation of the scaled gradient system (3.7) of J.*

Proof. See Appendix C. □

The results of Lemma 3.3 imply that the state trajectories of the averaged dynamical system (3.8) can be estimated by the state trajectories of the scaled gradient system (3.7). In the case $a_i = a_j,\ \ i,j = 1,\cdots,n$ then $\dot{x}(t) = -\sum_{i=1}^{n} \frac{a_i^2}{2}\nabla_{\frac{\partial}{\partial x_i}} J(x(t))\frac{\partial}{\partial x_i}$ is identical to (3.5).

Remark 1. We analyze properties of the state trajectory of (3.4) based on the state trajectory of the average system (3.8). Also stability properties of the scaled gradient system $\dot{x}(t) = -\sum_{i=1}^{n} \frac{a_i^2}{2}\nabla_{\frac{\partial}{\partial x_i}} J(x(t))\frac{\partial}{\partial x_i}$ facilitate closeness of solutions between the time varying dynamical system (3.3) and the scaled gradient system (3.7). The same results on Euclidean spaces are presented in [36]. The following lemma gives the uniform local asymptotic stability of the scaled gradient system obtained in Lemma 3.3.

Lemma 3.4. *Consider the gradient dynamical system (3.7) on a compact connected n dimensional Riemannian manifold (M, g^M). Then, subject to Assumption 1, J is a Lyapunov function and x^* is locally asymptotically stable (see [18]) for the scaled gradient system (3.7) on (M, g^M)*

Proof. See Appendix C.1. □

The following theorem is the main result of this paper which gives a local convergence of the geodesic extremum seeking system to a unique local minimum/maximum of J on an n dimensional compact connected Rimeannian manifold (M, g^M).

Theorem 3.5. *Consider the geodesic extremum seeking system (3.4) on a compact connected n dimensional Riemannian manifold (M, g^M), where $\omega_i = \omega\bar{\omega}_i, i = 1,\cdots,n$ satisfy Assumption 1. Assume $x^* \in M$ is a unique local optimizer of $J : M \to \mathbb{R}_{\geqslant 0}$. Then for any neighborhood $U_{x^*} \subset M$ of x^* on M, there exist sufficiently small parameters $a_i > 0,\ \ i = 1,\cdots,n$, sufficiently large frequency ω and a neighborhood of x^* denoted by $\hat{U}_{x^*} \subset M$, such that for any $x_0 \in \hat{U}_{x^*}$, the state trajectory of the closed loop system (3.4) initiating from x_0 ultimately enters and remains in U_{x^*}.*

Proof. See Appendix C.2. □

Remark 2. *The compactness assumption of M in Theorem 3.5 can be relaxed when the proof is carried out in a compact neighborhood of x^*. This type of relaxation is illustrated in the proof of Theorem 4.6 in Appendix D for extremum seeking systems on non-compact Lie groups. The presented technique can be directly applied to non-compact Riemannian manifolds.*

Definition 3.6. *Let $X, Y \in \mathfrak{X}(M)$ be smooth vector fields on M, where it may be shown that $\Phi_Y(t, t_0, .) : M \to M$ is a local diffeomorphism (see [1]). Let us denote $T_x\Phi_Y^{(t,t_0)^{-1}}$ as the pushforward of $\Phi_Y^{-1}(t, t_0, \cdot) : M \to M$ at $x \in M$. Define the pull back of $\Phi_Y^{(t,t_0)} \doteq \Phi_Y(t, t_0, \cdot)$ by $\Phi_Y^{(t,t_0)^*}$ denoted as follows.*

$$\begin{aligned}
&\Phi_Y^{(t,t_0)^*} : \mathfrak{X}(M) \to \mathfrak{X}(M),\\
&\left(\Phi_Y^{(t,t_0)^*} X\right)(x) \doteq T_{\Phi_Y(t,t_0,x)}\Phi_Y^{(t,t_0)^{-1}} X(\Phi_Y(t, t_0, x)),\\
&X \in \mathfrak{X}(M), x \in M, t \in \mathbb{R}. \qquad (3.9)
\end{aligned}$$

In $\mathbb{R}^n$, for a diffeomorphism $\phi : \mathbb{R}^n \to \mathbb{R}^n$ and $X \in \mathfrak{X}(\mathbb{R}^n)$, we have

$$(\phi^* X)(x) = \left(\frac{\partial \phi^{-1}}{\partial x} \circ X \circ \phi\right)(x) = \frac{\partial \phi^{-1}}{\partial x}(X(\phi(x))).$$

Lemma 3.7. *Consider a* T *periodic time varying dynamical system* $\dot{x} = \epsilon f(x,t), \epsilon \in \mathbb{R}_{\geqslant 0}$, *where* $f(x,t+T) = f(x,t)$, *on an* n *dimensional Riemannian manifold* (M, g^M). *The averaged dynamical system is given by* $\dot{x} = \epsilon \hat{f}(x)$, *where* $\hat{f}(x) \doteq \frac{1}{T}\int_0^T f(x,s)ds$. *Consider the combination of state flows* $\zeta(t) \doteq \Phi^{(1,0)}_{\epsilon Z} \circ \Phi_{\epsilon f}(t, t_0, x_0) \doteq \Phi_{\epsilon Z}(1, 0, \Phi_{\epsilon f}(t, t_0, x_0)) \in M$ *where* $Z(t,x) \doteq \int_0^t \big(\hat{f}(x) - f(x,s)\big)ds$. *Then* $\zeta(\cdot)$ *satisfies*

$$\dot{\zeta}(t) = \epsilon\Big[(\Phi^{-1})^{(1,0)^*}_{\epsilon Z} f + \int_0^1 (\Phi^{-1})^{(1,s)^*}_{\epsilon Z}\big(\hat{f} - f\big)ds\Big] \circ \zeta(t),$$

where $(\Phi^{-1})^{(1,s)^*}_{\epsilon Z}$ *is the pullback (see [7]) of the local diffeomorphism* $(\Phi^{-1})^{(1,s)}_{\epsilon Z} \doteq \Phi^{-1}_{\epsilon Z}(1, s, \cdot)$.

Proof. The proof parallels the results of [7], Chapter 9 and [28] on $\mathbb{R}^n$ and is omitted. □

Lemma 3.7 is employed to obtain closeness of solutions for the dynamical systems (3.3) and its averaged system as per Lemma 3.3. The full proof of Theorem 3.5 is given in Appendix C.2.

4. Extremum seeking on Lie groups. The extremum seeking system (3.4) is modified for optimization on Lie groups. In this case we employ the group structure of the ambient manifold and define the extremum seeking vector field along the exponential maps on Lie groups. This makes the computation of the geodesic dithers defined before particularly simple for matrix Lie groups. The following lemma characterizes geodesics on Lie groups which admit bi-invariant Riemannian metrics.

Lemma 4.1 ([32]). *Assume* G *is a Lie group which admits a bi-invariant Riemannian metric. Then for a left invariant vector field* X *on* G, $g \star \exp(tX)$ *is a geodesic emanating from* $g \in G$.

The geodesic dither (3.2) is given along exp on a Lie group G by

$$g_p(t) = g \star \exp\left(\sum_{i=1}^n a_i \sin(\omega_i t)\frac{\partial}{\partial g_i}\right), \quad g, g_p \in G, \tag{4.1}$$

where $\frac{\partial}{\partial g_i}$, $i = 1, \cdots, n$ are the base elements of the Lie algebra $\mathcal{L}$. Note that $\exp\left(\sum_{i=1}^n a_i \sin(\omega_i t)\frac{\partial}{\partial g_i}\right)$ is given by the solution of the left invariant time dependent vector field
$X(t,g) = T_e g\left(\sum_{i=1}^n a_i \sin(\omega_i t)\frac{\partial}{\partial g_i}\right) \in \mathfrak{X}(\mathbb{R} \times G)$ initiating from $e \in G$, where $\left(\sum_{i=1}^n a_i \sin(\omega_i t)\frac{\partial}{\partial g_i}\right) \in \mathcal{L}$. The extremum seeking vector field on G is then defined by

$$f(g,t) \doteq -\sum_{i=1}^n a_i \sin(\omega_i t) J\left(g \star \exp\left(\sum_{j=1}^n a_j \sin(\omega_j t)\frac{\partial}{\partial g_j}\right)\right) T_e g\left(\frac{\partial}{\partial g_i}\right) \in T_g G, \tag{4.2}$$

where $T_e g\left(\frac{\partial}{\partial g_i}\right)$, $i = 1, \cdots, n$ are tangent vectors at $T_g G$. Similar to (3.4), the extremum seeking system on $(G, \star)$ is defined as

$$\dot{g}(t) = f(g,t), \quad g(t) \in G. \tag{4.3}$$

One may easily verify that $T_e(g_1 \star g_2)\left(\frac{\partial}{\partial g_i}\right) = T_{g_2} g_1\left(T_{g_2}\left(\frac{\partial}{\partial g_i}\right)\right)$ which shows that $T_e g\left(\frac{\partial}{\partial g_i}\right)$ is a left invariant vector field. Note that $f(g,t)$ is not necessarily left invariant since $J(g_1 \star g_2 \star \exp(\sum_{i=1}^n a_i \sin(\omega_i t))) = J(g_2 \star \exp(\sum_{i=1}^n a_i \sin(\omega_i t)))$ does not hold in general. The following theorem gives the stability of the geodesic extremum seeking algorithm (4.2) on compact Lie groups.

Theorem 4.2. *Consider the extremum seeking system (4.3) on a compact connected n dimensional Lie group G, where $\omega_i = \omega\bar{\omega}_i$, $i = 1, \cdots, n$ satisfy Assumption 1. Assume $g^* \in G$ is a unique local optimizer of $J : G \to \mathbb{R}_{\geqslant 0}$. Then for any neighborhood $U_{g^*} \subset G$ of g^* on G, there exist sufficiently small parameters $a_i > 0$, $i = 1, \cdots, n$, sufficiently large ω and a neighborhood $\hat{U}_{g^*} \subset G$ of g^* such that for any $g_0 \in \hat{U}_{g^*}$, the state trajectory of (4.2) initiating from g_0 ultimately enters and remains in U_{g^*}.*

Proof. Since G is compact then it possesses a bi-invariant Riemannian metric, see [32]. Hence, Lemma 4.1 implies that $g \star \exp(\eta \sum_{i=1}^n a_i \sin(\omega_i t)\frac{\partial}{\partial g_i})$, $\eta \in \mathbb{R}$ is a geodesic through $g \in G$ and the proof is identical to the proof of Theorem 3.5. □

In the case that G is not compact or does not possess a bi-invariant Riemannian metric we employ the Taylor expansion of smooth functions on G, given in [21] in order to prove Theorem 4.2. This alternative case is codified via the following Lemma.

Lemma 4.3 ([21]). *Consider a left invariant vector field $X \in \mathfrak{X}(G)$ which is identified by $X(e) \in \mathcal{L}$. Then for a smooth function $J : G \to \mathbb{R}$ we have*

$$J(g \star \exp(X(e))) = \sum_{j=1}^{m} \frac{1}{j!}(X^j J)(g) + \frac{1}{m!}\int_0^1 (1-s)^m (X^{m+1}J)(g \star \exp(sX(e)))ds,$$

where $X^j J(g) = \frac{d^j}{dt^j} J(g \star \exp(tX(e)))\,|_{t=0}$.

Lemma 4.4. *Consider the extremum seeking system in (4.3) on a connected Lie group G where the optimizing trajectory is perturbed by the exponential dither presented in (4.1). Then, subject to Assumption 1, the averaged dynamical system of (4.2) is a perturbation of the first order variation $X_a(g)J$, where $X_a(g)$ is the left invariant vector field identified by $X_a(e) = \sum_{i=1}^n a_i^2 \frac{\partial}{\partial g_i}$.*

Proof. The proof follows the results of Lemma 4.3 and parallels the proof of Lemma 3.3 and hence is omitted. □

Lemma 4.5. *Consider the gradient dynamical system*
$\dot{g} = -\sum_{i=1}^n \frac{a_i^2}{2}\left(T_e g \frac{\partial}{\partial g_i} J\right)(g) T_e g \frac{\partial}{\partial g_i}$, $g \in G$ *on an n dimensional Lie group $(G, \star)$. Then subject to Assumption 1, g^* is locally asymptotically stable on $(G, \star)$ for the gradient dynamical system.*

Proof. The proof parallels the proof of Lemma 3.4 and hence is omitted. □

The following theorem parallels Theorem 4.2 for optimization on connected Lie groups which do not admit bi-invariant Riemannian metrics.

Theorem 4.6. *Consider the extremum seeking system (4.3) on a connected n dimensional Lie group G which admits no bi-invariant Riemannian metric. Assume $g^* \in G$ is a unique local optimizer of $J : G \to \mathbb{R}_{\geqslant 0}$ and $\omega_i = \omega\bar{\omega}_i$, $i = 1, \cdots, n$, where J and ω_i satisfy Assumption 1. Then for any neighborhood $U_{g^*} \subset G$ of g^* on G, there exist sufficiently small parameters a_i, $i = 1, \cdots, n$, sufficiently large ω and a neighborhood $\hat{U}_{g^*} \subset G$ of g^* such that for any $g_0 \in \hat{U}_{g^*}$, the state trajectory of (4.2) initiating from g_0 ultimately enters and remains in U_{g^*}.*

Proof. See Appendix D. □

5. Example on Stiefel manifolds $V_{n,p}$. In this section we consider an optimization problem on a Stiefel manifold $V_{3,2}$. As is known $V_{n,p}$ consists of n by p

($p < n$) orthogonal matrices such that (see [2])

$$V_{n,p} = \{X \in \mathbb{R}^{n\times p} | X^T X = I_{p,p}\}. \tag{5.1}$$

The Stifle manifold $V_{n,p}$ can be considered as the quotient space O_n/O_{n-p} where O_n is the orthogonal matrix group. The equivalence classes are given by right multiplication, see [10]. In [10] the authors have shown that the tangent vectors at $X \in V_{n,p}$ are given by

$$\Delta = XA + X_\perp B \in T_X V_{n,p}, \tag{5.2}$$

where $X_\perp$ is the n by $n-p$ orthogonal compliment of X such that $XX^T + X_\perp X_\perp^T = I_{n,n}$ and A and B are arbitrary p by p skew symmetric and $n-p$ by p matrices. For $V_{n,p}$ we consider a canonical Riemannian metric which is not the same as the Euclidean metric of $V_{n,p}$ as a submanifold of $\mathbb{R}^{n\times p}$. This metric is given by

$$g^{V_{n,p}}(\Delta_1, \Delta_2) = \frac{1}{2} tr(A_1^T A_2) + tr(B_1^T B_2), \tag{5.3}$$

where $\Delta_1, \Delta_2 \in T_X V_{n,p}$ and A_1, A_2, B_1, B_2 are their representations at X. It is important to note that the settings of this problem do not satisfy the hypotheses of the method presented in [9] (we do not consider the submanifold Euclidean metric on $V_{n,p}$) and therefore cannot be tackled by their algorithm. By employing the approach presented in this paper, we are able to calculate the geodesic curves with respect to the canonical Riemannian metric and implement the optimization algorithm formulated in Section 3. By employing the results of [10], the geodesics emanating from $X(0) = X$ with the initial velocity $\dot{X}(0) = H$ are given by

$$X(t) = XM(t) + (I_{n,n} - XX^T)H \int_0^t M(t)dt, \tag{5.4}$$

where $\ddot{M} - X^T\dot{M} + H^T(I_{n,n} - XX^T)HM = 0$, and $M(0) = I_{p,p}, \dot{M}(0) = X^T H$. For our example we select $n = 3$ and $p = 2$. As $V_{n,p}$ is a $\frac{p(p-1)}{2} + p(n-p)$ dimensional manifold then $V_{3,2}$ is a three dimensional manifold and hence the tangent space at $X \in V_{3,2}$ is a three dimensional vector space. Hence, we consider the following geodesic dither to construct the extremum seeking vector field.

$$X_p = \exp_X \left(a_1 \sin(\omega_1 t) X \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} + X_\perp \begin{pmatrix} a_2 \sin(\omega_2 t) & a_3 \sin(\omega_3 t) \end{pmatrix} \right). \tag{5.5}$$

Following [10], the cost function to be minimized is considered as

$$J(X) = \frac{1}{2} tr(X^T Y X N), \tag{5.6}$$

where Y and N are 3 by 3 and 2 by 2 symmetric matrices respectively. In this example $Y = \begin{pmatrix} 1 & 3 & 5 \\ 3 & 5 & 7 \\ 5 & 7 & 9 \end{pmatrix}$ and $N = \begin{pmatrix} 1 & 5 \\ 5 & 4 \end{pmatrix}$. The algorithm starts from $X(0) = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 \end{pmatrix}$. The minimization of the cost function and convergence of $X(t)$ are shown in Figure 5.

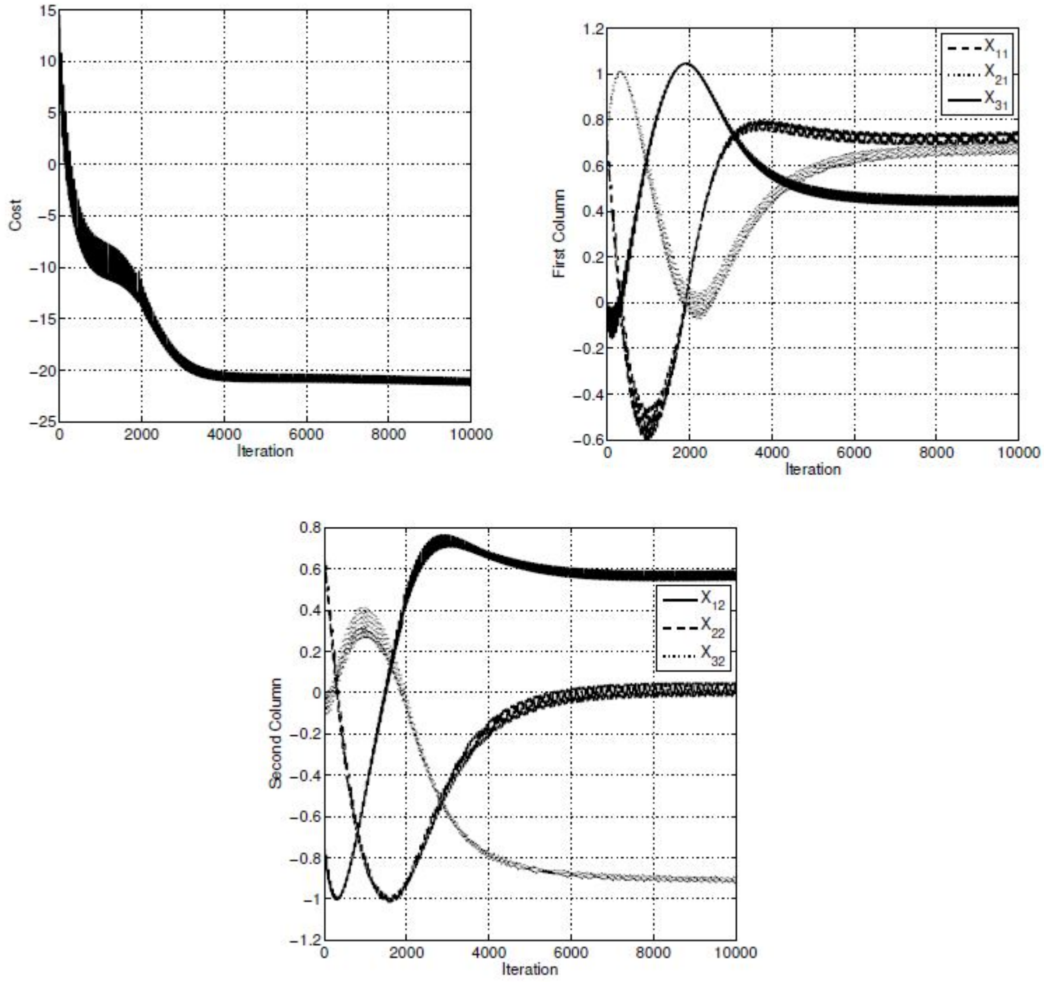

FIG. 5.1. *Cost and state convergence on* $V_{3,2}$.

6. Example on $SE(3)$**.** In this section we give another conceptual example for an orientation control on $SE(3)$, which is not compact.

As is known, $SE(3)$ is the space of rotation and translation which is used for robotic modelling. We have

$$SE(3) = \left\{ \begin{pmatrix} g_{SO(3)} & g_{\mathbb{R}} \\ 0_{1\times 3} & 1 \end{pmatrix} \in \mathbb{R}^{4\times 4} | \quad g_{SO(3)} \in SO(3), g_{\mathbb{R}} \in \mathbb{R}^{3\times 1} \right\},$$

where $g_{SO(3)}$ models rotation and $g_{\mathbb{R}}$ models translation in $\mathbb{R}^3$. The Lie algebra of $SE(3)$ which is denoted by $se(3)$ is given by (see [45])

$$se(3) = \left\{ \begin{pmatrix} S & v \\ 0_{1\times 3} & 0 \end{pmatrix} \in \mathbb{R}^{4\times 4} | \quad S \in so(3), v \in \mathbb{R}^3 \right\}.$$

Let us consider the cost function $J : SE(3) \to \mathbb{R}$ as

$$J(g) = \frac{1}{2}tr((g_{SO(3)} - g^*_{SO(3)})^T(g_{SO(3)} - g^*_{SO(3)})) + \frac{1}{2}||g_{\mathbb{R}} - r^*||^2_{\mathbb{R}^3},$$

where $g^*_{SO(3)}$ is the optimal orientation matrix in $SO(3)$ and r^* is the optimal distance from the origin in $\mathbb{R}^3$. The optimal solution for the optimization problem above is $\begin{pmatrix} g^*_{SO(3)} & r^* \\ 0_{1\times 3} & 1 \end{pmatrix} \in SE(3)$. The cost function above minimizes the distance from the orientation $g^*_{SO(3)}$ and distance from r^*. Without loss of generality, we assume $g^*_{SO(3)} = I_{3\times 3} \in SO(3)$ and $r^* = (0,0,0) \in \mathbb{R}^3$.

The Lie algebra $se(3)$ is spanned by $\frac{\partial}{\partial g_1} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$, $\frac{\partial}{\partial g_2} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$, $\frac{\partial}{\partial g_3} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$, $\frac{\partial}{\partial g_4} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$, $\frac{\partial}{\partial g_5} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ and $\frac{\partial}{\partial g_6} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$. For this example the dither vector $X(t,e)$ at the Lie algebra $se(3)$ is given by

$$X(t,e) = \sum_{i=1}^{6} a_i \sin(\omega_i t)\frac{\partial}{\partial g_i} = \begin{pmatrix} 0 & a_1\sin(\omega_1 t) & a_3\sin(\omega_3 t) & a_4\sin(\omega_4 t) \\ -a_1\sin(\omega_1 t) & 0 & a_2\sin(\omega_2 t) & a_5\sin(\omega_5 t) \\ -a_3\sin(\omega_3 t) & -a_2\sin(\omega_2 t) & 0 & a_6\sin(\omega_6 t) \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{6.1}$$

hence, the dither vector field is given by $X(t,g) = g \cdot X(t,e)$, where $g \in SE(3)$.

Similar to the example on $SO(3)$, the extremum seeking vector field on $SE(3)$ is given by the following vector field

$$f(g,t) \doteq -\sum_{i=1}^{6} a_i \sin(\omega_i t) J(g \exp \sum_{j=1}^{6} a_j \sin(\omega_j t)\frac{\partial}{\partial g_j})g\frac{\partial}{\partial g_i},$$

where exp is the exponential operator defined on $SE(3)$. On $SE(3)$ the exponential map does not coincide with geodesics since $SE(3)$ does not admit a bi-invariant Riemannian metric, see [32]. Hence, we employ the results of Theorem 4.6 to guarantee a local convergence of the algorithm.

In this case, the exp operator is not the same as the exp operator on $SO(3)$. For a tangent vector $\begin{pmatrix} S & v \\ 0_{1\times 3} & 0 \end{pmatrix} \in se(3)$, where $S = \begin{pmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{pmatrix}$, we have $\exp\left(\begin{pmatrix} S & v \\ 0_{1\times 3} & 0 \end{pmatrix}\right) = \begin{pmatrix} \exp(S) & Av \\ 0_{1\times 3} & 1 \end{pmatrix}$, where $A = I_{3\times 3} + \frac{(1-\cos(\theta))}{\theta^2}S + \frac{(\theta - \sin(\theta))}{\theta^3}S^2$, and $\theta = \sqrt{a^2+b^2+c^2}$. In the case that $\theta = 0$, we have $\exp\left(\begin{pmatrix} S & v \\ 0_{1\times 3} & 0 \end{pmatrix}\right) = \begin{pmatrix} \exp(S) & v \\ 0_{1\times 3} & 1 \end{pmatrix}$.

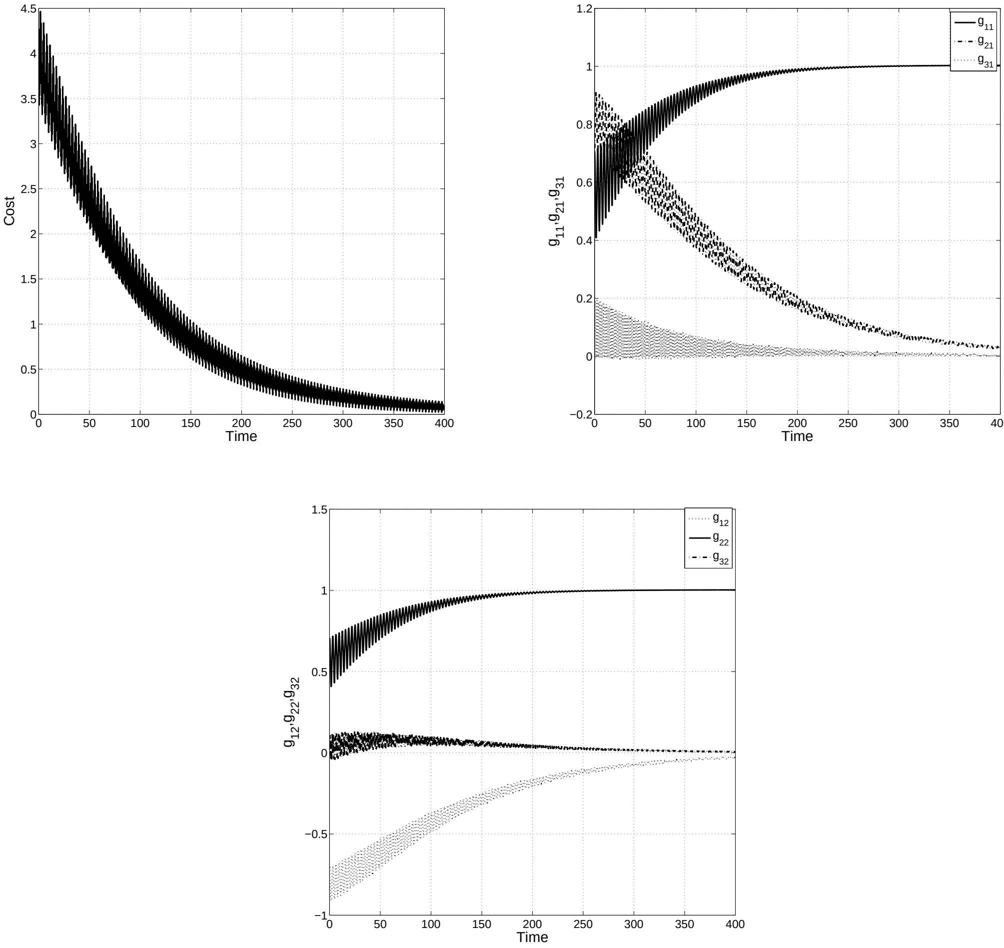

FIG. 6.1. *Cost and state convergence on $SE(3)$.*

The optimizing trajectory $g(\cdot)$ is a solution of the time dependent differential equation

$$\dot{g}(t) = f(g(t), t) \in T_g SE(3).$$

The algorithm initiates from the initial orientation at $g_{SO(3)}(0) = \begin{pmatrix} \cos(\frac{\pi}{4}) & -\sin(\frac{\pi}{4}) & 0 \\ \sin(\frac{\pi}{4}) & \cos(\frac{\pi}{4}) & 0 \\ 0 & 0 & 1 \end{pmatrix} \in SO(3)$ and $g_{\mathbb{R}}(0) = (1, -1, 2) \in \mathbb{R}^3$. The amplitudes and frequencies are set at $a_1 = \cdots = a_6 = .1$ and $\omega_i = 2i + .1, \; i = 1, \cdots, 6$.

The results for the convergence of the cost function and the state trajectory $g(t) = \begin{pmatrix} g_{11}(t) & g_{12}(t) & g_{13}(t) & g_{14}(t) \\ g_{21}(t) & g_{22}(t) & g_{23}(t) & g_{24}(t) \\ g_{31}(t) & g_{32}(t) & g_{33}(t) & g_{34}(t) \\ 0 & 0 & 0 & 1 \end{pmatrix} \in SE(3)$ are shown in Figures 6.1 and 6.2.

Appendix A. Averaging on Riemannian manifolds.

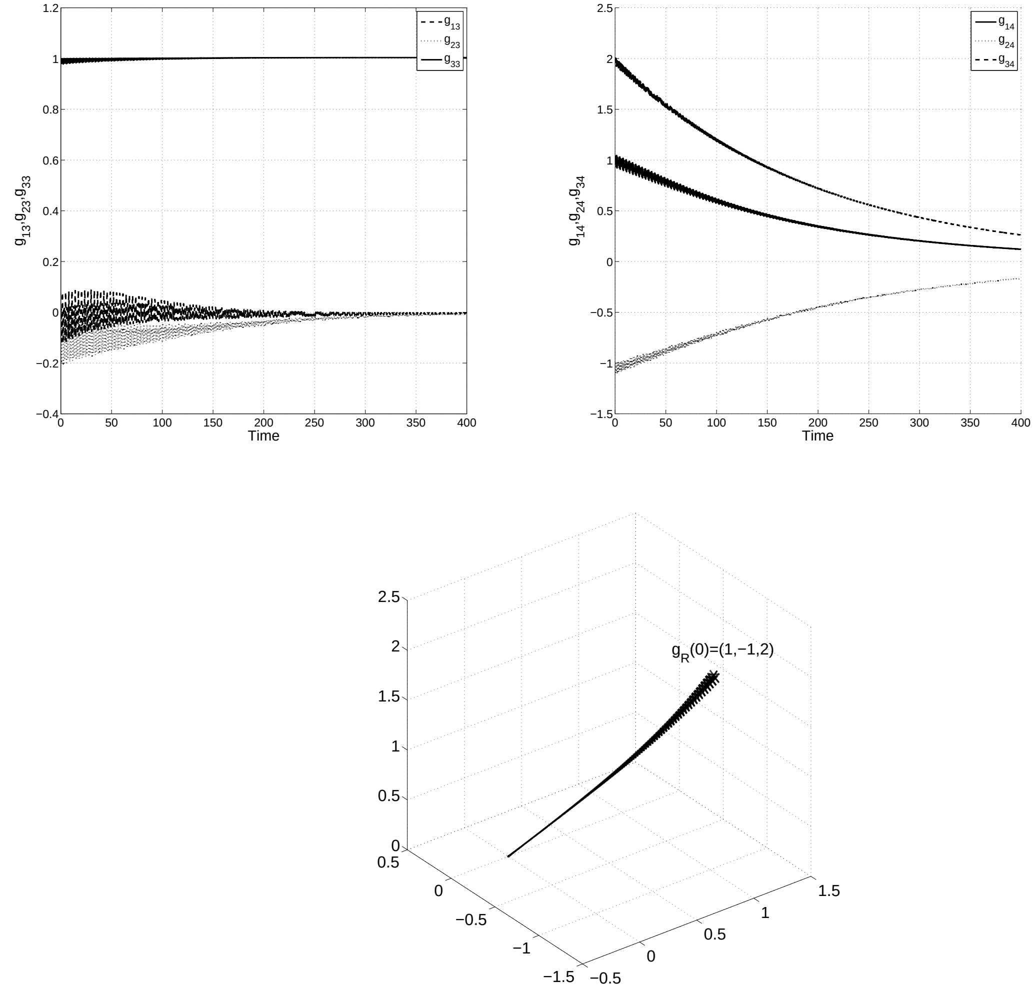

FIG. 6.2. *State convergence on $SE(3)$.*

Let us consider a time varying perturbed dynamical system as

$$\dot{x}(t) = \epsilon f(x(t), t), \quad f \in \mathfrak{X}(M \times \mathbb{R}), x_0 \in M, \quad \epsilon \geqslant 0,$$

where f is periodic in t with the period T, i.e. $f(x,t) = f(x, t+T)$. Such a system is referred to as *T-periodic.* The averaged vector field $\hat{f}$ is given by

$$\hat{f}(x) \doteq \frac{1}{T} \int_0^T f(x,s) ds, \tag{A.1}$$

where the average dynamical system is locally given by $\dot{x}(t) = \epsilon \hat{f}(x(t))$.

In order to obtain closeness of solutions for dynamical systems we employ the notion of pullbacks of vector fields along diffeomorphisms on M as per Definition 3.6.

We have the following lemma for the variation of smooth parameter varying vector fields.

Lemma A.1 ([4], Page 40, [7], Page 451). *Consider a smooth parameter varying vector field* $Y(\lambda, x)$*, where* $Y \in \mathfrak{X}(\mathbb{R} \times M)$ *with the associated flow* $\Phi_Y(t, t_0, \cdot) : M \to M$*. Then,*

$$\frac{\partial}{\partial \lambda}\Phi_Y(t, t_0, x_0) = T_{x_0}\Phi_Y^{(t,t_0)}\int_{t_0}^{t}\left(\Phi_Y^{(s,t_0)*}\frac{\partial}{\partial \lambda}Y(\lambda, \cdot)\right)(x_0)ds =$$

$$\int_{t_0}^{t}(\Phi^{-1})_Y^{(t,s)*}\frac{\partial}{\partial \lambda}Y(\lambda, \cdot)ds \circ \Phi_Y(t, t_0, x_0) \in T_{\Phi_Y(t,t_0,x_0)}M, \tag{A.2}$$

where $T_{x_0}\Phi_Y^{(t,t_0)}$ *is the pushforward of* $\Phi_Y(t, t_0, \cdot)$ *at* x_0 *and* $(\Phi^{-1})_Y^{(t,s)*}$ *is the pull back of* $\Phi^{-1}(t, s, \cdot)_Y$ *as per Definition 3.6.*

Appendix B. Stability of perturbed systems on Riemannian manifolds. Consider the following perturbed dynamical system on (M, g^M).

$$\dot{x}(t) = f(x, t) + h(x, t), f, h \in \mathfrak{X}(M \times \mathbb{R}). \tag{B.1}$$

The term h is considered as a perturbation of the nominal vector field f. The next lemma gives the existence of Lyapunov functions for dynamical systems on Riemannian manifold which satisfy specific local properties.

Lemma B.1 ([41]). *Let* x^* *be an equilibrium for the smooth dynamical system* $\dot{x} = f(x, t)$ *which is uniformly asymptotically stable (see [18]) on an open set* $\mathcal{N}_{x^*} \subset \mathcal{U}^n_{x^*}$ *(*$\mathcal{U}^n_{x^*}$ *is a normal neighborhood around* x^**). Assume* $||T_x f(\cdot, t)||$ *is uniformly bounded with respect to* t *on* $\mathcal{N}_{x^*}$*, where* $||\cdot||$ *is the norm of the bounded linear operator* $Tf : TM \to TTM$*. Then, for some* $\mathcal{U}_{x^*} \subset \mathcal{U}^n_{x^*}$*, there exist a differentiable function* $w : M \times \mathbb{R} \to \mathbb{R}_{\geqslant 0}$ *and* $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in \mathcal{K}$ *(family of continuous, strictly increasing and zero at zero functions, see [18]), such that for all* $x \in \mathcal{U}_{x^*}$ *and* $t \in [t_0, \infty)$*,*

$$\begin{aligned} &(i): \ \alpha_1\left(d(x, x^*)\right) \leqslant w(x, t) \leqslant \alpha_2\left(d(x, x^*)\right), \\ &(ii): \ \mathcal{L}_{f(x,t)}w \leqslant -\alpha_3\left(d(x, x^*)\right), \\ &(iii): \ ||T_x w|| \leqslant \alpha_4\left(d(x, x^*)\right), \end{aligned} \tag{B.2}$$

where $d(\cdot, \cdot)$ *is the Riemannian metric,* $\mathcal{L}$ *is the Lie derivative and* $Tw : TM \to T\mathbb{R} \simeq \mathbb{R} \times \mathbb{R}$ *is the pushforward of* w*.*

Note that the lemma above holds for the time invariant gradient system $\dot{x} = -\sum_{i=1}^{n}\frac{a_i^2}{2}\nabla_{\frac{\partial}{\partial x_j}}J(x)\frac{\partial}{\partial x_i}$ since by Lemma 3.4 the gradient system is locally asymptotically stable. In this case the Lyapunov function w is time invariant. By the results of Lemma 3.4, it has been shown that J can be considered as a Lyapunov function. However, generally, w may not be necessary identical to J. One may show that items (i)-(iii) in Lemma B.1 locally hold for J around x^* when Assumption 1 is satisfied. Also note that for a compact manifold M, $||T_x f||$ is a bounded operator and the hypotheses of Lemma B.1 are satisfied for the extremum seeking algorithm (3.3) on compact Riemannian manifolds.

The following theorem gives the stability of (B.1), where the nominal system is locally uniformly asymptotically stable.

Theorem B.2. *Let* x^* *be an equilibrium of dynamical system* $\dot{x} = f(x, t)$*, which is locally uniformly asymptotically stable (see [18]) on a neighborhood* $\mathcal{N}_{x^*} \subset \mathcal{U}^n_{x^*}$ *(*$\mathcal{U}^n_{x^*}$ *is a normal neighborhood around* x^**). Assume the perturbed dynamical system (B.1) is complete and the Riemannian norm of the perturbation* $h \in \mathfrak{X}(M \times \mathbb{R})$ *is* *bounded on $\mathcal{N}_{x^*}$, i.e. $||h(x,t)||_{g^M} \leqslant \delta, x \in \mathcal{N}_{x^*}, t \in [t_0, \infty)$. Then, for sufficiently small δ, there exists a neighborhood U_{x^*} and a function $\rho \in \mathcal{K}$, such that*

$$\limsup_{t\to\infty} d(\Phi_{f+h}(t,t_0,x_0), x^*) \leqslant \rho(\delta), \quad x_0 \in U_{x^*}.$$

Proof. See [42] fo the detailed proof. ☐

Appendix C. Proof of Lemma 3.3 . The cost function J may be expanded along geodesics by using the Taylor expansion on Riemannian manifolds in normal neighborhods, see [38]. We employ Lemma 3.1 to guarantee that there exist $a_i, \;\; i = 1, \cdots, n$, such that $\exp_x \sum_{i=1}^n a_i \sin(\omega_i t)\frac{\partial}{\partial x_i} \in \exp_x\big(B_{\iota(x)}(0)\big)$. Then the Taylor expansion of J along the geodesic $\exp_x \theta X, \;\; 0 < \theta$, where $X \in T_xM$, is given by (see [38])

$$\begin{aligned} &J(\exp_x \theta X) = J(x) + \theta(\nabla_X J)(x) + ... + \frac{\theta^{m-1}}{(m-1)!} \times \\ &(\nabla_X^{m-1} J)(x) + \frac{\theta^m}{(m-1)!}\int_0^1 (1-s)^{m-1}\nabla_X^m J(\exp_x s\theta X)ds, 0 < \theta < \theta^*, \end{aligned} \tag{C.1}$$

which is equivalent to

$$\begin{aligned} &J(\exp_x \theta X) = J(x) + \theta(dJ(X))|_x + ... + \frac{\theta^{m-1}}{(m-1)!} \times (\nabla_X^{m-2} dJ)(X)|_x + \\ &\frac{\theta^m}{(m-1)!}\int_0^1 (1-s)^{m-1}(\nabla_X^{m-1} dJ)(X)(\exp_x s\theta X)ds, \quad 0 < \theta < \theta^*, \end{aligned} \tag{C.2}$$

where $dJ : TM \to \mathbb{R}$ is a differential form of J, θ^* is the upper existence limit for geodesics on M and ∇ is the *Levi-Civita* connection, see [23]. Note that for compact manifolds $\theta^* = \infty$. The expansion above along the geodesic dithers in (3.2) is

$$\begin{aligned} &J(\exp_x \sum_{i=1}^n a_i \sin(\omega_i t)\frac{\partial}{\partial x_i}) = J(x) + (\nabla_{\sum_{i=1}^n a_i \sin(\omega_i t)\frac{\partial}{\partial x_i}} J)(x) + \cdots + \\ &\frac{1}{(m-1)!} \times (\nabla^{m-1}_{\sum_{i=1}^n a_i \sin(\omega_i t)\frac{\partial}{\partial x_i}} J)(x) + \frac{1}{(m-1)!} \times \int_0^1 (1-s)^{m-1}\nabla^m_{\sum_{i=1}^n a_i \sin(\omega_i t)\frac{\partial}{\partial x_i}} \\ &J(\exp_x s\sum_{i=1}^n a_i \sin(\omega_i t)\frac{\partial}{\partial x_i})ds. \end{aligned}$$

Linear properties of ∇ imply that (see [23]) $\nabla_{\sum_{i=1}^n a_i \sin(\omega_i t)\frac{\partial}{\partial x_i}} J(x) = \sum_{i=1}^n a_i \sin(\omega_i t)\nabla_{\frac{\partial}{\partial x_i}} J(x)$, and iteratively we have

$$\nabla^m_{\sum_{i=1}^n a_i \sin(\omega_i t)\frac{\partial}{\partial x_i}} J(x) = \sum_{i=1}^n a_i \sin(\omega_i t)\nabla_{\frac{\partial}{\partial x_i}}\big(\nabla^{m-1}_{\sum_{j=1}^n a_j \sin(\omega_j t)\frac{\partial}{\partial x_j}} J\big)(x).$$

Dropping the $\hat{}$ notation for the state trajectory in (3.3), the dynamical equations for the extremum seeking feedback loop are given in x coordinates as follows:

$$\dot{x}(t) = -\Big(\sum_{i=1}^n a_i \sin(\omega_i t) J(x)\frac{\partial}{\partial x_i} + \sum_{i,j=1}^n a_i a_j \sin(\omega_i t)\sin(\omega_j t)\nabla_{\frac{\partial}{\partial x_j}} J(x)\frac{\partial}{\partial x_i} + \cdots +$$

$$\frac{1}{(m-1)!}\sum_{i,j=1}^{n} a_i a_j \sin(\omega_i t)\sin(\omega_j t)\times \nabla_{\frac{\partial}{\partial x_i}}\big(\nabla^{m-2}_{\sum_{i=1}^{n} a_i \sin(\omega_i t)\frac{\partial}{\partial x_i}} J\big)(x)\frac{\partial}{\partial x_i} +$$
$$\frac{1}{(m-1)!}\sum_{i=1}^{n} a_i \sin(\omega_i t)\Big(\int_0^1 (1-s)^{m-1}\nabla^{m}_{\sum_{i=1}^{n} a_i \sin(\omega_i t)\frac{\partial}{\partial x_i}}$$
$$J(\exp_x s\sum_{i=1}^{n} a_i \sin(\omega_i t)\frac{\partial}{\partial x_i})ds\Big)\frac{\partial}{\partial x_i}\Big).$$
(C.3)

Denoting a new time scale by $\tau \doteq \omega t$, then (C.3) is a τ varying vector field on (M, g^M) which is periodic with respect to τ. The dynamical system in this new time scale is given by

$$\frac{dx}{d\tau} = -\frac{1}{\omega}\Big(\sum_{i=1}^{n} a_i \sin(\bar{\omega}_i\tau)J(x)\frac{\partial}{\partial x_i} + \sum_{i,j=1}^{n} a_i a_j \sin(\bar{\omega}_i\tau)\sin(\bar{\omega}_j\tau)\nabla_{\frac{\partial}{\partial x_j}} J(x)\frac{\partial}{\partial x_i} + \cdots +$$
$$\frac{1}{(m-1)!}\sum_{i,j=1}^{n} a_i a_j \sin(\bar{\omega}_i\tau)\sin(\bar{\omega}_j\tau)\times \nabla_{\frac{\partial}{\partial x_i}}\big(\nabla^{m-2}_{\sum_{i=1}^{n} a_i \sin(\bar{\omega}_i\tau)\frac{\partial}{\partial x_i}} J\big)(x)\frac{\partial}{\partial x_i} +$$
$$\frac{1}{(m-1)!}\sum_{i=1}^{n} a_i \sin(\bar{\omega}_i\tau)\Big(\int_0^1 (1-s)^{m-1}\nabla^{m}_{\sum_{i=1}^{n} a_i \sin(\bar{\omega}_i\tau)\frac{\partial}{\partial x_i}}$$
$$J(\exp_x s\sum_{i=1}^{n} a_i \sin(\bar{\omega}_i\tau)\frac{\partial}{\partial x_i})ds\Big)\frac{\partial}{\partial x_i}\Big).$$

Let T denote the least common multiplier of the periods of $\sin(\bar{\omega}_i\tau),\ \ i = 1,\cdots,n$ which its existence is guaranteed by Assumption 1 since $\bar{\omega}_i,\ \ i = 1,\cdots,n$, are rational. The average dynamical system is then given by

$$\frac{dx}{d\tau} = \frac{1}{T}\int_0^T -\frac{1}{\omega}\Big(\sum_{i=1}^{n} a_i \sin(\bar{\omega}_i\tau)J(x)\frac{\partial}{\partial x_i} + \sum_{i,j=1}^{n} a_i a_j \sin(\bar{\omega}_i\tau)\sin(\bar{\omega}_j\tau)\nabla_{\frac{\partial}{\partial x_j}} J(x)\frac{\partial}{\partial x_i} + \cdots +$$
$$\frac{1}{(m-1)!}\sum_{i,j=1}^{n} a_i a_j \sin(\bar{\omega}_i\tau)\sin(\bar{\omega}_j\tau)\times \nabla_{\frac{\partial}{\partial x_i}}\big(\nabla^{m-2}_{\sum_{i=1}^{n} a_i \sin(\bar{\omega}_i\tau)\frac{\partial}{\partial x_i}} J\big)(x)\frac{\partial}{\partial x_i} +$$
$$\frac{1}{(m-1)!}\sum_{i=1}^{n} a_i \sin(\bar{\omega}_i\tau)\Big(\int_0^1 (1-s)^{m-1}\nabla^{m}_{\sum_{i=1}^{n} a_i \sin(\bar{\omega}_i\tau)\frac{\partial}{\partial x_i}}$$
$$J(\exp_x s\sum_{i=1}^{n} a_i \sin(\bar{\omega}_i\tau)\frac{\partial}{\partial x_i})ds\Big)\frac{\partial}{\partial x_i}\Big)d\tau.$$
(C.4)

Since M is compact and J is smooth then the higher derivatives of J are all bounded above on M and (C.4) is written as

$$\frac{dx}{d\tau} =-\frac{1}{\omega}\sum_{i=1}^{n}\frac{a_i^2}{2}\nabla_{\frac{\partial}{\partial x_i}} J(x)\frac{\partial}{\partial x_i} + \frac{1}{\omega}\sum_{i=1}^{n} O((\max_{i\in 1,\cdots,n} a_i)^4)\frac{\partial}{\partial x_i}. \tag{C.5}$$

The vector field $-\frac{1}{\omega}\sum_{i=1}^{n}\frac{a_i^2}{2}\nabla_{\frac{\partial}{\partial x_i}} J(x)\frac{\partial}{\partial x_i} +$
$\frac{1}{\omega}\sum_{i=1}^{n} O((\max_{i\in 1,\cdots,n} a_i)^4)\frac{\partial}{\partial x_i}$ is a perturbed version of the time invariant vector field

$-\frac{1}{\omega}\sum_{i=1}^{n}\frac{a_i^2}{2}\nabla_{\frac{\partial}{\partial x_i}}J(x)\frac{\partial}{\partial x_i}$ on (M, g^M). Following (3.5), we note that $-\frac{1}{\omega}\sum_{i=1}^{n}\frac{a_i^2}{2}\nabla_{\frac{\partial}{\partial x_i}}J(x)\frac{\partial}{\partial x_i}$ is a scaled version of the gradient system presented in (3.5).

C.1. Proof of Lemma 3.4. Consider J as the candidate Lyapunov function on (M, g^M). The variation of J along $-\sum_{i=1}^{n}\frac{a_i^2}{2}\nabla_{\frac{\partial}{\partial x_i}}J(x)\frac{\partial}{\partial x_i}$ is

$$\begin{aligned}\mathcal{L}_{-\sum_{i=1}^{n}\frac{a_i^2}{2}\nabla_{\frac{\partial}{\partial x_i}}J(x)\frac{\partial}{\partial x_i}}J &= dJ\Big(-\sum_{i=1}^{n}\frac{a_i^2}{2}\nabla_{\frac{\partial}{\partial x_j}}J(x)\frac{\partial}{\partial x_i}\Big)\\ &= -\sum_{i=1}^{n}\frac{a_i^2}{2}dJ\Big(\nabla_{\frac{\partial}{\partial x_i}}J(x)\frac{\partial}{\partial x_i}\Big).\end{aligned}$$

where $\mathcal{L}$ is the Lie derivative of J along vector fields on (M, g^M). Locally $dJ = \sum_{i=1}^{n}\frac{\partial J}{\partial x_i}dx_i$, where $dx_i(\frac{\partial}{\partial x_j}) = \delta_{i,j}$. Hence,

$$\mathcal{L}_{-\sum_{i=1}^{n}\frac{a_i^2}{2}\nabla_{\frac{\partial}{\partial x_i}}J(x)\frac{\partial}{\partial x_i}}J = -\sum_{i=1}^{n}\frac{a_i^2}{2}\Big(\frac{\partial J}{\partial x_i}\Big)^2 \leqslant 0. \tag{C.6}$$

Note that Assumption 1 guarantees that, locally, $\sum_{i=1}^{n}\frac{a_i^2}{2}(\frac{\partial J}{\partial x_i})^2 \neq 0$ for $x \neq x^*$, i.e. $\mathcal{L}_f J$ is locally negative-definite. By Assumption 1 and (C.6) the cost function $J : M \to \mathbb{R}_{\geqslant 0}$ is locally positive definite, its derivative with respect to time is negative definite and $J(x^*) = 0$. Hence, J is a Lyapunov function on (M, g^M), see [7]. Therefore, applying the results of [7], Theorem 6.14, implies that x^* is locally asymptotically stable.

C.2. Proof of Theorem 3.5. We analyze closeness of solutions between state trajectories of $\frac{dx}{d\tau} = \frac{1}{\omega}f(x,\tau)$ and state trajectories of $\frac{dx}{d\tau} = \frac{1}{\omega}\hat{f}(x)$, where $f(x,\tau) = -\sum_{i=1}^{n}\sin(\bar{\omega}_i\tau)J(\exp_{\hat{x}}\sum_{i=1}^{n}a\sin(\bar{\omega}_i\tau)\frac{\partial}{\partial x_i})\frac{\partial}{\partial x_i}$ and $\hat{f}(x) = \frac{1}{T}\int_0^T f(x,\tau)d\tau$. Note that $\frac{dx}{d\tau} = \frac{1}{\omega}f(x,\tau)$ is the extremum seeking dynamical system (C.3) in the time scale $\tau = \omega t$ as per the proof of Lemma 3.3 in Appendix C. Also note that $\frac{1}{\omega}$ plays the role of ϵ which appeared in perturbed dynamical systems in (A.1) in Appendix A. In this case increasing ω guarantees that the magnitude of $\frac{1}{\omega}f(x,\tau)$ in τ scale decreases.

We consider the periodic vector field Z defined in Lemma 3.7, $Z(\tau,x) \doteq \int_0^\tau(\hat{f}(x) - f(x,s))ds,\ \ x \in M, \tau \in \mathbb{R}_{\geqslant 0}$, where $Z(\tau,x) = Z(\tau+T,x)$ and T is the period of the extremum seeking system f as per the proof of Lemma 3.3. Now consider a composition of flows on M given by $\zeta(\tau) = \Phi^{(1,0)}_{\frac{1}{\omega}Z}\circ\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0)$. By Lemma 3.7, the tangent vector of ζ is computed by

$$\begin{aligned}\dot{\zeta}(\tau) &= T_{\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0)}\Phi^{(1,0)}_{\frac{1}{\omega}Z}\Big(\frac{1}{\omega}f(\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0),\tau)\Big) + \frac{\partial}{\partial\tau}\big(\Phi^{(1,0)}_{\frac{1}{\omega}Z}\circ\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0)\big)\\ &= (\Phi^{-1})^{(1,0)^*}_{\frac{1}{\omega}Z}\Big(\frac{1}{\omega}f(\cdot,\tau)\Big)(\zeta(\tau)) + \frac{1}{\omega}\int_0^1(\Phi^{-1})^{(1,s)^*}_{\frac{1}{\omega}Z}\big(\hat{f}(\cdot) - f(\cdot,\tau)\big)ds\circ\zeta(\tau),\end{aligned}\tag{C.7}$$

where $(\Phi^{-1})^{(1,s)^*}_{\frac{1}{\omega}Z}$ is the pullback of the state flow $\Phi^{-1}_{\frac{1}{\omega}Z}$ and $\epsilon = \frac{1}{\omega}$, see Definition 3.6 of pullbacks along diffeomorphisms. Equivalently, in a compact form, we have

$$\dot{\zeta}(\tau) = \frac{1}{\omega}\Big[(\Phi^{-1})^{(1,0)^*}_{\frac{1}{\omega}Z}f + \int_0^1(\Phi^{-1})^{(1,s)^*}_{\frac{1}{\omega}Z}\big(\hat{f} - f\big)ds\Big]\circ\zeta(\tau)$$

$$\doteq \frac{1}{\omega} H\left(\frac{1}{\omega}, \tau, \zeta(\tau)\right). \tag{C.8}$$

One may see that $H(0,\tau,x) = \hat{f}(x)$ where by the construction above, H is smooth with respect to $\frac{1}{\omega}$. By applying the Taylor expansion with remainder we have

$$H\left(\frac{1}{\omega}, \tau, x\right) = \hat{f}(x) + \frac{1}{\omega} h(x,\xi,\tau),$$

where $h(x,\xi,\tau) = \frac{\partial}{\partial \epsilon} H(\epsilon,\tau,x)\,|_{\epsilon=\frac{1}{\omega}=\xi}$ and $\xi \in [0, \frac{1}{\omega}]$. We note that $H\left(\frac{1}{\omega},\tau,x\right)$ is periodic with respect to τ since $f(x,\tau)$ and $Z(\tau,x)$ are both T-periodic. Hence, $h(x,\xi,\tau)$ is a T-periodic vector field on M.

The metric triangle inequality on (M, g^M) implies

$$\begin{aligned}
&d(\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0), \Phi_{\frac{1}{\omega}\hat{f}}(\tau,\tau_0,x_0)) \leqslant d(\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0), \Phi^{(1,0)}_{\frac{1}{\omega}Z} \circ \Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0)) + \\
&d(\Phi^{(1,0)}_{\frac{1}{\omega}Z} \circ \Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0), \Phi_{\frac{1}{\omega}\hat{f}}(\tau,\tau_0,x_0)) \leqslant d(\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0), \Phi^{(1,0)}_{\frac{1}{\omega}Z} \circ \Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0)) + \\
&d(\Phi^{(1,0)}_{\frac{1}{\omega}Z} \circ \Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0), x^*) + d(\Phi_{\frac{1}{\omega}\hat{f}}(\tau,\tau_0,x_0), x^*).
\end{aligned} \tag{C.9}$$

Based on (C.9), We analyze closeness of solutions for the following dynamics.

$$\begin{aligned}
\dot{x}(\tau) &= \frac{1}{\omega}\hat{f}(x(\tau)), && x(\tau_0) = x_0, \\
\dot{y}(\tau) &= \frac{1}{\omega}\acute{f}(y(\tau)), && y(\tau_0) = x_0, \\
\dot{z}(\tau) &= \frac{1}{\omega}\hat{f}(z(\tau)) + \frac{1}{\omega^2} h(z,\xi,\tau), && z(\tau_0) = x_0,
\end{aligned} \tag{C.10}$$

where $\acute{f}(x) = -\sum_{i=1}^{n} \frac{a_i^2}{2} \nabla_{\frac{\partial}{\partial x_j}} J(x) \frac{\partial}{\partial x_i}$ is the gradient system and $\frac{1}{\omega}\hat{f}(z(\tau)) + \frac{1}{\omega^2} h(z,\xi,\tau)$ is the vector field corresponding to the state trajectory $\Phi^{(1,0)}_{\frac{1}{\omega}Z} \circ \Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0)$. Rescaling time back to t via $t = \frac{1}{\omega}\tau$, yields

$$\begin{aligned}
\frac{dx}{dt} &= \hat{f}\left(x(t)\right), && x(t_0) = x_0, \\
\frac{dy}{dt} &= \acute{f}\left(y(t)\right), && y(t_0) = x_0, \\
\frac{dz}{dt} &= \hat{f}\left(z(t)\right) + \frac{1}{\omega} h(z,\xi,t), && z(t_0) = x_0,
\end{aligned}$$

or equivalently by Lemma 3.3,

$$\begin{aligned}
\frac{dx}{dt} &= \acute{f}\left(x(t)\right) + \sum_{i=1}^{n} O((\max_{i \in 1,\cdots,n} a_i)^4) \frac{\partial}{\partial x_i}, \\
\frac{dy}{dt} &= \acute{f}\left(y(t)\right), \\
\frac{dz}{dt} &= \acute{f}\left(z(t)\right) + \sum_{i=1}^{n} O((\max_{i \in 1,\cdots,n} a_i)^4) \frac{\partial}{\partial z_i} + \frac{1}{\omega} h(z,\xi,t),
\end{aligned} \tag{C.11}$$

where $x(t_0) = y(t_0) = z(t_0) = x_0$. The variation of the cost function J along $\hat{f}(\cdot)$ is given by

$$\begin{aligned}\mathcal{L}_{\hat{f}}J &= \mathcal{L}_{\acute{f}+\sum_{i=1}^{n} O((\max_{i\in 1,\cdots,n} a_i)^4)\frac{\partial}{\partial x_i}}J \\ &= \mathcal{L}_{\acute{f}}J + \mathcal{L}_{\sum_{i=1}^{n} O((\max_{i\in 1,\cdots,n} a_i)^4)\frac{\partial}{\partial x_i}}J \\ &= \mathcal{L}_{\acute{f}}J + \sum_{i=1}^{n} O((\max_{i\in 1,\cdots,n} a_i)^4)\mathcal{L}_{\frac{\partial}{\partial x_i}}J.\end{aligned}$$

As shown by the proof of Lemma 3.4, locally, we have $\mathcal{L}_{\acute{f}}J \leqslant 0$. Without loss of generality, assume positive definiteness and negative definiteness of J and $\mathcal{L}_{\acute{f}}J = \mathcal{L}_{-\frac{1}{\omega}\sum_{i=1}^{n}\frac{a_i^2}{2}\nabla_{\frac{\partial}{\partial x_j}}J(x)\frac{\partial}{\partial x_i}}J$ on $U_{x^*} \subset M$ of x^*. Otherwise we apply the intersection of the corresponding neighborhoods to perform the analysis above. Define the sublevel set $\mathcal{N}_b, b \in \mathbb{R}_{>0}$ of the positive definite function $J : M \to \mathbb{R}_{\geqslant 0}$ as $\mathcal{N}_b \doteq \{x \in M,\ J(x) \leqslant b\}$. By $\mathcal{N}_b(x^*)$ we denote a connected sublevel set of M containing $x^* \in M$. By Lemma 6.12 in [7], there exists a sublevel set $\mathcal{N}_b(x^*) \subset U_{x^*}$, such that $\mathcal{N}_b(x^*)$ is compact. Consider a neighborhood of x^* denoted by W_{x^*} such that $W_{x^*} \subset int(\mathcal{N}_b(x^*)) \subset U_{x^*} \subset M$, where $int()$ gives the interior set. Compactness of M implies that $M - W_{x^*}$ is closed and compact. Hence, $\mathcal{L}_{\acute{f}}J < 0$ for all $x \in (M - W_{x^*})\bigcap\mathcal{N}_b(x^*)$. Note that $\mathcal{L}_{\acute{f}}J < 0$ on $U_{x^*} - \{x^*\}$.

Smoothness of J and compactness of $M - W_{x^*}$ together imply that $\mathcal{L}_{\acute{f}}J$ attains its bounded maximum value in $M - W_{x^*}$. Hence, by selecting $a_i,\ \ i = 1, \cdots, n$ sufficiently small we have $\mathcal{L}_{\hat{f}}J < 0$ on $(M - W_{x^*})\bigcap\mathcal{N}_b(x^*)$. This implies that the state trajectory $\Phi_{\hat{f}}(t, t_0, x_0)$ remains in $\mathcal{N}_b(x^*)$ for $x_0 \in int(\mathcal{N}_b(x^*))$.

The variation of J along $\hat{f}\,(z(t)) + \frac{1}{\omega}h(z, \xi, t)$ is then given by

$$\begin{aligned}\mathcal{L}_{\hat{f}+\frac{1}{\omega}h}J &= \mathcal{L}_{\hat{f}}J + \frac{1}{\omega}\mathcal{L}_{h}J \\ &= \mathcal{L}_{\acute{f}}J + \sum_{i=1}^{n} O((\max_{i\in 1,\cdots,n} a_i)^4)\mathcal{L}_{\frac{\partial}{\partial x_i}}J + \frac{1}{\omega}\mathcal{L}_{h}J.\end{aligned}$$

The same argument applies to the variation of J along $\hat{f} + \frac{1}{\omega}h(z, \xi, t)$ and we obtain that $\mathcal{L}_{\hat{f}+\frac{1}{\omega}h}J < 0$ on $(M - W_{x^*})\bigcap\mathcal{N}_b(x^*)$ for sufficiently small a_i and sufficiently large ω. Note that h is periodic with respect to t, $\xi \in [0, \frac{1}{\omega}]$ and M is compact. Hence, $\mathcal{L}_{h}J$ is bounded and this implies that by choosing a_i sufficiently small and ω sufficiently large the state trajectory $z(\cdot)$ remains in $\mathcal{N}_b(x^*)$ for all initial states $z_0 \in int(\mathcal{N}_b(x^*))$. Denote the uniform normal neighborhood of $x^* \in M$ with respect to U_{x^*} by $U^n_{x^*}$ (its existence is guaranteed by Lemma 5.12 in [23]). Consider a geodesic ball of radius δ where $U^n_{x^*} \subset \exp_{x^*}(B_\delta(0)) \subset U_{x^*}$. By definition, $\exp_{x^*}(B_\delta(0))$ is an open set containing x^* in the topology of M. Therefore one may shrink b to $\acute{b}, 0 < \acute{b} \leqslant b$, such that $\mathcal{N}_{\acute{b}}(x^*) \subset \exp_{x^*}(B_\delta(0))$. Hence, by the argument above, we select the set of initial states such that $\Phi_{\hat{f}+\frac{1}{\omega}h}(\cdot, t_0, x_0)$ stays in a normal neighborhood of x^*. For the economy of notation we replace $\acute{b}$ with b and assume $\mathcal{N}_b(x^*) \subset \exp_{x^*}(B_\delta(0)) \subset U_{x^*}$.

We analyze the distance between the state trajectory $y(\cdot)$ of the asymptotically stable system and the averaged and full systems. As is obvious from (C.11), the averaged system and the dynamical system corresponding to $z(\cdot)$ are perturbations of the gradient system $\dot{y} = \acute{f}(y)$, where $\dot{y} = \acute{f}(y)$ is locally asymptotically stable and the magnitude of the perturbations vector fields is arbitrarily shrunken by adjusting a_i and ω in (C.11). Since the initial state set is chosen such that the state trajectory $z(\cdot)$ remains in a normal neighborhood of x^*, then the conditions of Theorems B.2 in Appendix B are satisfied. Hence, there exist a neighborhood $U^1_{x^*} \subset int(\mathcal{N}_b(x^*))$ and a continuous function ρ, such that for all $x_0 \in U^1_{x^*}$

$$\limsup_{t\to\infty} d\left(\Phi_{\hat{f}+\frac{1}{\omega}h}(t,t_0,x_0),x^*\right) \leqslant \rho\left(\sup_{z\in M, t\in[t_0,t_0+T]} ||\sum_{i=1}^{n} O((\max_{i\in 1,\cdots,n} a_i)^4)\frac{\partial}{\partial z_i}+\frac{1}{\omega}h(z,\xi,t)||_{g^M}\right), \tag{C.12}$$

where ρ is a continuous function which is zero at zero. We note that since M is compact, h is periodic with respect to t and $\xi \in [0,\frac{1}{\omega}]$, then $\sup_{z\in M,t\in[t_0,t_0+T]} ||\sum_{i=1}^{n} O((\max_{i\in 1,\cdots,n} a_i)^4)\frac{\partial}{\partial z_i}+\frac{1}{\omega}h(z,\xi,t)||_{g^M}$ is bounded. Also note that (C.12) does not guarantee the convergence of the perturbed state trajectory to x^*. However, it gives a local closeness of solutions in terms of the Riemannian distance function d to x^* after elapsing enough time. The closeness estimation provided in (C.12) is used to bound the distance between $y(\cdot)$, $x(\cdot)$ and $z(\cdot)$ as follows. By employing the triangle inequality we have

$$d\left(\Phi_f(t,t_0,x_0),\Phi_{\acute{f}}(t,t_0,x_0)\right) \leqslant d\left(\Phi_f(t,t_0,x_0),\Phi_{\hat{f}+\frac{1}{\omega}h}(t,t_0,x_0)\right) + d\left(\Phi_{\hat{f}+\frac{1}{\omega}h}(t,t_0,x_0),\Phi_{\acute{f}}(t,t_0,x_0)\right), \tag{C.13}$$

and

$$d\left(\Phi_{\hat{f}+\frac{1}{\omega}h}(t,t_0,x_0),\Phi_{\acute{f}}(t,t_0,x_0)\right) \leqslant d\left(\Phi_{\hat{f}+\frac{1}{\omega}h}(t,t_0,x_0),x^*\right) + d\left(x^*,\Phi_{\acute{f}}(t,t_0,x_0)\right), \tag{C.14}$$

where in (C.14), $d(x^*,\Phi_{\acute{f}}(t,t_0,x_0))$ converges to zero and $d(\Phi_{\hat{f}+\frac{1}{\omega}h}(t,t_0,x_0),x^*)$ can be chosen arbitrarily small by (C.12). Note that $||h(\cdot,\cdot,\cdot)||_{g^M}$ is bounded since M is compact, $\xi \in [0,\frac{1}{\omega}]$ and h is periodic with respect to t. In order to show the boundedness of $d(\Phi_f(t,t_0,x_0),\Phi_{\hat{f}+\frac{1}{\omega}h}(t,t_0,x_0))$ in (C.13), we switch back to the time scale τ. Now we prove $d(\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0),\Phi^{(1,0)}_{\frac{1}{\omega}Z}\circ\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0)) = O\left(\frac{1}{\omega}\right)$. By the definition of the distance function given in [25], we have $d(\Phi_{\frac{1}{\omega}Z}(s,0,x),x) \leqslant \ell(\Phi_{\frac{1}{\omega}Z}(s,0,x))$, where $\ell(\Phi_{\frac{1}{\omega}Z}(s,0,x))$ is the length of the curve connecting x to $\Phi_{\frac{1}{\omega}Z}(s,0,x)$ on M. Therefore,

$$d\left(\Phi_{\frac{1}{\omega}Z}(1,0,x),x\right) \leqslant \ell(\Phi_{\frac{1}{\omega}Z}(1,0,x)) = \frac{1}{\omega}\int_0^1 ||Z(\tau,\Phi_{\frac{1}{\omega}Z}(s,0,x))||_{g^M}ds. \tag{C.15}$$

Periodicity of Z with respect to τ, boundedness of $\Phi_{\frac{1}{\omega}Z}(\cdot,0,x)$ in the sense of compactness of M and smoothness of Z with respect to x together yield $d(\Phi_{\frac{1}{\omega}Z}(1,0,x),x) = O\left(\frac{1}{\omega}\right)$. Since x is a generic element of M we have

$$d\left(\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0),\Phi^{(1,0)}_{\frac{1}{\omega}Z}\circ\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0)\right) = O\left(\frac{1}{\omega}\right), \forall\tau\in[\tau_0,\infty),$$

where x is replaced by $\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0)\in M$. Hence, by using (C.9), for any $x_0\in U^1_{x*}$, there exists a time $\acute{T}_{x_0}$, such that

$$d\left(\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0),\Phi_{\frac{1}{\omega}\acute{f}}(\tau,\tau_0,x_0)\right)\leqslant d\left(\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0),\Phi^{(1,0)}_{\frac{1}{\omega}Z}\circ\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0)\right)+$$
$$d\left(\Phi^{(1,0)}_{\frac{1}{\omega}Z}\circ\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0),\Phi_{\frac{1}{\omega}\acute{f}}(\tau,\tau_0,x_0)\right)\leqslant d\left(\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0),\Phi^{(1,0)}_{\frac{1}{\omega}Z}\circ\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0)\right)+$$
$$d\left(\Phi^{(1,0)}_{\frac{1}{\omega}Z}\circ\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0),x^*\right)+d\left(x^*,\Phi_{\frac{1}{\omega}\acute{f}}(\tau,\tau_0,x_0)\right)\leqslant$$
$$O\left(\frac{1}{\omega}\right)+\rho\left(\sup_{z\in M,t\in[t_0,t_0+T]}||\sum_{i=1}^{n}O((\max_{i\in1,\cdots,n}a_i)^4)\frac{\partial}{\partial zi}+\frac{1}{\omega}h(z,\xi,t)||_{g^M}\right)+$$
$$d\left(x^*,\Phi_{\frac{1}{\omega}\acute{f}}(\tau,\tau_0,x_0)\right),\quad\forall\tau\in[\omega\acute{T}_{x_0},\infty).$$

Note that $\Phi^{(1,0)}_{\frac{1}{\omega}Z}\circ\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0)=\Phi_{f+\frac{1}{\omega}h}(t,t_0,x_0)$, for $\tau=\omega t$ and $\tau_0=\omega t_0$. Finally we have

$$d\left(\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0),x^*\right)\leqslant d\left(\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0),\Phi_{\frac{1}{\omega}\acute{f}}(\tau,\tau_0,x_0)\right)+d\left(x^*,\Phi_{\frac{1}{\omega}\acute{f}}(\tau,\tau_0,x_0)\right)\leqslant$$
$$O\left(\frac{1}{\omega}\right)+\rho\left(\sup_{z\in M,t\in[t_0.t_0+T]}||\sum_{i=1}^{n}O((\max_{i\in1,\cdots,n}a_i)^4)\frac{\partial}{\partial zi}+\frac{1}{\omega}h(z,\xi,t)||_{g^M}\right)+$$
$$2d\left(x^*,\Phi_{\frac{1}{\omega}\acute{f}}(\tau,\tau_0,x_0)\right),\quad\forall\tau\in[\omega\acute{T}_{x_0},\infty),x_0\in U^1_{x*}.$$
(C.16)

As (C.16) indicates $d\left(\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,x_0),x^*\right)$ can be ultimately bounded by shrinking $a_i,\ i=1,\cdots,n$ and increasing ω such that the state trajectory $\Phi_f(t,t_0,x_0)$ enters U_{x*} and remains there.

Appendix D. Proof of Theorem 4.6.

The proof parallels the proof of Theorem 3.5 by employing the results of Lemmas 4.4 and 4.5. However, (C.15) does not necessarily hold since G is not compact. The same as (C.15) we observe that

$$d(\Phi_f(t,t_0,g_0),\Phi_{\acute{f}}(t,t_0,g_0))\leqslant d(\Phi_f(t,t_0,g_0),\Phi_{\hat{f}+\frac{1}{\omega}h}(t,t_0,g_0))+$$
$$d(\Phi_{\hat{f}+\frac{1}{\omega}h}(t,t_0,g_0),\Phi_{\acute{f}}(t,t_0,g_0)),$$

and

$$d(\Phi_{\hat{f}+\frac{1}{\omega}h}(t,t_0,g_0),\Phi_{\acute{f}}(t,t_0,g_0))\leqslant d(\Phi_{\hat{f}+\frac{1}{\omega}h}(t,t_0,g_0),g^*)+d(g^*,\Phi_{\acute{f}}(t,t_0,g_0)),$$

where $f,\ \hat{f}$ and $\acute{f}$ are the extremum seeking, averaged and gradient vector fields induced by (4.2). By Lemma 4.5 and Theorem B.2, $d(g^*,\Phi_{\acute{f}}(t,t_0,g_0))$ converges to zero and $d(\Phi_{\hat{f}+\frac{1}{\omega}h}(t,t_0,g_0),g^*)$ is ultimately bounded by a continuous function ρ where the bound is shrunken by adjusting $a_i, i=1,\cdots,n$ and ω. Note that in the proof of Theorem B.2 it is shown that asymptotic stability of the gradient vector field $\acute{f}$ guarantees that the state trajectories of $\hat{f}$ and $\hat{f}+\frac{1}{\omega}h$ remains in a compact subset containing the equilibrium of $\acute{f}$. It remains to show $d(\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,g_0),\Phi^{(1,0)}_{\frac{1}{\omega}Z}\circ\Phi_{\frac{1}{\omega}f}(\tau,\tau_0,g_0))=O\left(\frac{1}{\omega}\right)$ in the time scale $\tau=\omega t$ for $\tau\in[\tau_0,\infty)$. Following the proof of

Theorem B.2 one may show $\Phi_{\hat{f}+\epsilon h}(\omega, \omega_0, g_0) \in \mathcal{N}_b(g^*),\ \ \omega \in [\omega_0, \infty), g_0 \in int(\mathcal{N}_b(g^*))$ for some $b \in \mathbb{R} > 0$, where $\mathcal{N}_b(g^*)$ is a compact connected sublevel set of a Lyapunov function containing $g^* \in G$, see the proof of Theorem B.2 and [7] Lemma 6.12. Hence, $\Phi^{(1,0)}_{\frac{1}{\omega}Z} \circ \Phi_{\frac{1}{\omega}f}(t, t_0, g_0) \in \mathcal{N}_b(g^*),\ \ \tau \in [\tau_0, \infty), g_0 \in int(\mathcal{N}_b(g^*))$. Therefore,

$$\begin{aligned}
\bigcup_{\tau\in[\tau_0,\infty)} \Phi_{\frac{1}{\omega}f}(\tau, \tau_0, g_0) \subset & \bigcup_{\tau\in[\tau_0,\infty)} (\Phi^{-1})^{(1,0)}_{\frac{1}{\omega}Z} \circ \mathcal{N}_b(g^*) \\
= & \bigcup_{\tau\in[\tau_0,\tau_0+\omega T]} (\Phi^{-1})^{(1,0)}_{\frac{1}{\omega}Z} \circ \mathcal{N}_b(g^*) \\
\subset & \bigcup_{\tau\in[\tau_0,\tau_0+\omega T],\, \frac{1}{\omega}\in[0,\hat{\epsilon}]} (\Phi^{-1})^{(1,0)}_{\frac{1}{\omega}Z} \circ \mathcal{N}_b(g^*),
\end{aligned}$$

where the equality is due the periodicity of Z and $\hat{\epsilon}$ is a limit for the minimum frequency $(0 < \omega < \infty)$. Compactness of $\mathcal{N}_b(g^*)$, $[\tau_0, \tau_0 + \omega T]$ and $[0, \hat{\epsilon}]$ together with the smoothness of Φ^{-1} gives compactness of $\bigcup_{\tau\in[\tau_0,\tau_0+\omega T],\, \frac{1}{\omega}\in[0,\hat{\epsilon}]} (\Phi^{-1})^{(1,0)}_{\frac{1}{\omega}Z} \circ \mathcal{N}_b(g^*)$ in G. Hence, we show that

$$\text{(D.1)}\quad d(\Phi_{\frac{1}{\omega}Z}(1, 0, z), z) = O\left(\frac{1}{\omega}\right),\ \ z \in \bigcup_{\tau\in[\tau_0,\tau_0+\omega T],\, \frac{1}{\omega}\in[0,\hat{\epsilon}]} (\Phi^{-1})^{(1,0)}_{\frac{1}{\omega}Z} \circ \mathcal{N}_b(g^*),$$

which proves $d(\Phi_{\frac{1}{\omega}f}(\tau, \tau_0, g_0), \Phi^{(1,0)}_{\frac{1}{\omega}Z} \circ \Phi_{\frac{1}{\omega}f}(\tau, \tau_0, g_0)) = O\left(\frac{1}{\omega}\right)$, where z is replaced by $\Phi_{\frac{1}{\omega}f}(\tau, \tau_0, g_0)$ in (D.1). The rest of the proof is identical to the proof of Theorem 3.5.

REFERENCES

[1] R. Abraham, J. E. Marsden, and T. S. Ratiu. *Manifolds, Tensor Analysis, and Applications.* Springer, 1988.

[2] P.A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds.* Princeton University Press, 2007.

[3] M. Adachi. *Embeddings and Immersions.* American Mathematical Soc., 1993.

[4] A. Agrachev and Y. Sachkov. *Control Theory from the Geometric Viewpoint.* Springer, 2004.

[5] C. G. Baker, P.-A. Absil, and K. A. Gallivan. An implicit trust-region method on Riemannian manifolds. *IMA J. Numer. Anal.*, 28:665–689, 2008.

[6] A. M. Bloch. *Nonholonomic Mechanics and Control.* Springer, 2000.

[7] F. Bullo and A.D. Lewis. *Geometric Control of Mechanical Systems: Modeling, Analysis, and Design for Mechanical Control Systems.* Springer, 2005.

[8] H. B. Dürr, M. S. Stanković, C. Ebenbauera, and K. H. Johansson. Lie bracket approximation of extremum seeking systems. *Automatica*, (49):1538–1552, 2013.

[9] H. B. Dürr, M. S. Stanković, K. H. Johansson, and C. Ebenbauera. Extremum seeking on submanifolds in Euclidean spaces. *Automatica*, Provisionally accepted for Publication, 2014.

[10] A. Edelman, T. A. Arias, and S. T. Smith. Maximum principle for a hybrid system via singular perturbations. *SIAM J. on Matrix Anal. Appl.*, 20(2):303–353, 1998.

[11] O. Freifeld. Statistics on manifolds with applications to modeling shape deformations. *PhD thesis, The Division of Applied Mathematics at Brown University*, 2014.

[12] P. Frihauf, M. Krstić, and T. Başar. Nash equilibrium seeking in non cooperative games. *IEEE Trans. Automatic Control*, 57(5):1192–1207, 2012.

[13] D. Gabay. Minimizing a differentiable function over a differentiable manifold. *Journal of Optimization Theory and Applications*, 37:177–219, 1982.

[14] M. Guay, D. Dochain, and M. Perrier. Adaptive extremum seeking control of nonlinear dynamic systems with parametric uncertainties. *Automatica*, 39:1283–1293, 2003.

[15] M. Guay, D. Dochain, and M. Perrier. Adaptive extremum seeking control of continuous stirred tank bioreactors with unknown growth kinetics. *Automatica*, 40:881–888, 2004.

[16] U. Helmke, S. Riardo, and J. B. Yoshizawa. Newtons algorithm in Euclidean jordan algbras, with applications to robotics. *Communications in Information and Systems*, 3(2):283297, 2002.

[17] J. Jost. *Reimannian Geometry and Geometrical Analysis.* Springer, 2004.

[18] H. K. Khalil. *Nonlinear Systems.* Prentice Hall, 2002.

[19] S. Z. Khong, D. Nešić, C. Manzie, and Y. Tan. Multidimensional global extremum seeking via the direct optimisation algorithm. *Automatica*, 49(7):1970–1978, 2013.

[20] W. P. A. Klingenberg. *Riemannian Geometry.* de Gruyter Studies in Mathematics, 1995.

[21] A. W. Knapp. *Lie groups Beyond an Introduction.* Birkhauser, 1996.

[22] M. Krstić and H. W. Wang. Stability of extremum seeking feedback for general nonlinear dynamic systems. *Automatica*, 36:595–601, 2000.

[23] J. M. Lee. *Riemannian Manifolds, An Introduction to Curvature.* Springer, 1997.

[24] J. M. Lee. *Introduction to Topological Manifolds.* Springer, 2000.

[25] J. M. Lee. *Introduction to Smooth Manifolds.* Springer, 2002.

[26] D. G. Luenberger. The gradient projection method along geodesics. *Management Science*, 18(11):620–631, 1972.

[27] J. H. Manton. Optimisation algorithms exploting unitary constraints. *IEEE Transactions on Signal Processing*, 50(3):635650, 2002.

[28] S. Martinez, J. Cortes, and F. Bullo. Analysis and design of oscillatory control systems. *IEEE Trans Automatic Control*, 48(7):1164–1177, 2003.

[29] D. Nešić, A. Mohammadi, and C. Manzie. A framework for extremum seeking control of systems with parameter uncertainties. *IEEE Trans. Aut. Cont.*, 58(2):435–448, 2013.

[30] D. Nešić, A. Mohammadi, and C. Manzie. A systematic approach to extremum seeking based on parameter estimation. In *IEEE Conf. Decision and Control*, pages 3902–3907, Atlanta, 2010.

[31] D. Nešić, Y. Tan, W. Moase, and C. Manzie. A unifying approach to extremum seeking: adaptive schemes based on estimation of derivatives. In *IEEE Conf. Decision and Control*, pages 4625–4630, Atlanta, 2010.

[32] X. Pennec. Bi-invariant means on Lie groups with Cartan-Schouten connections. *Lecture Notes in Computer Science, Geometric Science of Information*, 8085:59–67, 2013.

[33] P. Petersen. *Riemannian Geometry.* Springer, 1998.

[34] W. Ring and B. Wirth. Optimization methods on Riemannian manifolds and their application to shape space. *SIAM journal on Control and Optimization*, 22(2):596–627, 2012.

[35] W. Rudin. *Real and Complex Analysis.* New York: McGraw-Hill, 1974.

[36] J. A. Sanders and F. Verhulst. *Averaging Methods in Nonlinear Dynamical Systems.* Springer, 1985.

[37] B. O. Shubert. A sequential method seeking the global maximum of a function. *SIAM Journal on Numerical Analysis*, 17(9):379388, 1972.

[38] S. T. Smith. Optimization techniques on Riemannian manifolds. *Fields Institute Communications*, 3(3):113–135, 1994.

[39] R. G. Strongin and Y. D. Sergeyev. *Global Optimization with Non-Convex Constraints.* Springer Science+Business Media Dordrecht, 2000.

[40] Y. Tan, D. Nešić, and I. M. Y. Mareels. On non-local stability properties of extremum seeking control. *Automatica*, 42(6):889–903, 2006.

[41] F. Taringoo, P. M. Dower, D. Nešić, and Y. Tan. *A Local Characterization of Lyapunov Functions and Robust Stability of Perturbed Systems on Riemannian Manifolds, http://arxiv.org/abs/1311.0078.* Submitted to *Automatica*, 2013.

[42] F. Taringoo, P. M. Dower, D. Nešić, and Y. Tan. *Optimization Methods on Riemannian Manifolds via Extremum Seeking Algorithms.* https://arxiv.org/abs/1412.2841, 2014.

[43] A. R. Teel and D. Popović. Solving smooth and non-smooth multivariable extremum seeking problems by the methods of nonlinear programming. In *American Control Conference*, pages 2394–2399, 2001.

[44] C. Udriste. *Convex Functions and Optimization Methods on Riemannian Manifolds.* Kluwer Academic Publishers, 1994.

[45] V. Varadarajan. *Lie groups, Lie algebras, and their representations.* Springer, 1984.

[46] Y. Yang. Globally convergent optimization algorithms on Riemannian manifolds: Uniform framework for unconstrained and constrained optimization. *Journal of Optimization Theory and Applications*, 132(2):245–265, 2007.