IEEE Control Systems Letters paper presented at
2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Step-size rules for Lie bracket-based extremum seeking with asymptotic convergence guarantees*

Victoria Grushkovskaya[1,3] and Christian Ebenbauer[2]

***Abstract*— In this paper, we derive a class of step-size rules (time-varying gains) for gradient-based extremum seeking algorithms that guarantee classical asymptotic convergence rather than practical convergence. The obtained step-size rule conditions are similar to the classical step-size rules known in stochastic approximation theory.**

I. Introduction

Extremum seeking is a powerful control technique aimed at finding and stabilizing an a priori unknown optimal state of a system using very limited knowledge about the system's model and objective function. While the first investigations of extremum seeking problems date back to the early 1920s, it can be considered a branch of modern control theory due to the intensive studies and results it has garnered since the beginning of the 21st century. Currently, there exist many significant results on theory and applications of extremum seeking control, see, e.g., [1]–[18]. We refer to [19], [20] for a comprehensive review. To steer a system to an unknown optimal state, many extremum seeking control algorithms employ time-periodic control inputs with rapidly varying frequencies and high amplitudes. Typically, this results in highly oscillating behavior of the solutions of the extremum seeking systems, which tend toward a neighborhood of the optimal state. This behavior is known as practical convergence, where the desired proximity to the optimal state is achieved by selecting suitable control design parameters. In this context, an important problem is to find control design parameters that achieve classical asymptotic convergence and thus ensure that the system's trajectory tends to the optimal state. One way to solve this problem is to design vector fields which vanish at the optimal state, as shown in, for example, [10], [14], [21]. These solutions, however, are only applicable to extremum problems with an a priori known optimal value.

In this paper, we develop control design parameters, in particular, step-size rules, for extremum seeking algorithms that are similar to the step-size rules known from stochastic gradient descent and stochastic approximation algorithms and that guarantee classical asymptotic convergence properties. In more detail, consider the stochastic gradient flow

$$dx = \tilde{\gamma}(t)(-\nabla J(x)dt + Bdw). \qquad (1)$$

For the sake of simplicity, let $J{:}\mathbb{R}{\to}\mathbb{R}$ be convex with a unique minimum x^*, $\tilde{\gamma}(t) \geq 0$ be some time-varying gain function, which can be thought of as a continuous-time step-size rule, $B \in \mathbb{R}\backslash\{0\}$, and dw be a standard Wiener process. This stochastic differential equation can be considered as a continuous-time version of the discrete-time stochastic gradient descent algorithm

$$x_{k+1}{=}x_k{+}\tilde{\gamma}_k(-h\nabla J(x_k){+}\sqrt{h}Bn_k),$$

where $h{>}0$ corresponds to the time discretization (Euler-Maruyama discretization), $\tilde{\gamma}_k{\geq}0$ to the step-size rule, and n_k is a normally distributed random variable with zero mean and unit variance. To achieve convergence to x^* in expectation *and* with zero variance ($\lim_{t\to\infty} E[x(t)]{=}x^*$, $\lim_{t\to\infty} E[(x(t){-}x^*)^2]{=}0$), i.e., convergence in the means square sense, it is well-known from stochastic approximation theory that $\tilde{\gamma}$ must converge to zero to ensure zero variance. However, $\tilde{\gamma}$ should not converge too fast to zero in order to maintain convergence in expectation. For example, for $J(x){=}x^2$, it is a rather simple calculation (using Ito calculus) to observe that the following classical conditions on $\tilde{\gamma}$ achieve the desired asymptotic convergence properties for (1) for the first and second moment (see also, e.g., [22]):

$$(\tilde{\mathrm{P}}1)\ \lim_{t\to\infty}\int_0^t \tilde{\gamma}(\tau)d\tau = \infty, \quad (\tilde{\mathrm{P}}2)\ \lim_{t\to\infty}\int_0^t \tilde{\gamma}(\tau)^2 d\tau < \infty.$$

A main motivation and goal of this paper is to develop similar conditions for extremum seeking algorithms. Gradient-based extremum seeking algorithms and stochastic gradient descent algorithms are both based on gradient approximation methods. Extremum seeking typically utilizes deterministic approximation methods while stochastic gradient descent uses stochastic approximation methods. Both approximations come with errors, e.g., deterministic error terms from remainders in Taylor-like expansions or, correspondingly, variances from sampling methods appear in the convergence analysis. Due to this and other conceptual similarities between stochastic gradient descent and extremum seeking algorithms, it is natural to ask whether conditions like $(\tilde{P}1)$ and $(\tilde{P}2)$ can be obtained for gradient-based extremum seeking in order to achieve asymptotic convergence to x^* instead of the commonly in the literature encountered practical convergence results. To address this question, we examine extremum seeking algorithms designed within the Lie bracket approximation framework as in, e.g., [6], [10], [14], [23]. However, unlike these results, we introduce a suitable class of time-varying gains (step-size rules) γ. Note that the Lie bracket approximation framework requires

*This work was not supported by any organization
[1]Department of Mathematics, University of Klagenfurt, 9020 Klagenfurt am Wörthersee, Austria viktoriia.grushkovska@aau.at
[2]Chair of Intelligent Control Systems, RWTH Aachen University, 52074 Aachen, Germany christian.ebenbauer@ic.rwth-aachen.de
[3]Institute of Applied Mathematics & Mechanics, NAS of Ukraine

uniform asymptotic stability of the optimal state for the associated Lie bracket system, which is generally not the case for systems with a time-varying gain. Therefore, the above-mentioned results are not applicable. Time-varying gains have been recently utilized, for example, in [24], [25]. However, the gains therein are increasing in time, while in our contribution we utilize decreasing and asymptotically vanishing gain functions. In addition, the analysis carried out in this paper allows us to derive rather general conditions on γ and describe the asymptotic behavior of the resulting system.

The primary contribution of this paper lies in establishing sufficient conditions on γ that guarantee asymptotic convergence to x^*. As discussed in Section II-B, while these conditions share similarities with ($\tilde{\mathrm{P}}1$) and ($\tilde{\mathrm{P}}2$), they also present certain distinctions. Furthermore, we estimate the speed of convergence of the solutions of the derived extremum seeking system. Additionally, we investigate extremum seeking systems with time-varying frequencies and establish convergence conditions for such systems. We illustrate the obtained results through a numerical example.

Notations: $R^+ = [0, \infty)$; e_j – unit vector with non-zero j-th entry; $B_\delta(x^*)$ – δ-neighborhood of an $x^* \in \mathbb{R}^n$ with $\delta > 0$;
$C(D)$ (resp., $C^p(D)$, $p \in \mathbb{N}$) – the space of continuous (resp., p times continuously differentiable) functions on $D \subset R^n$;
$\nabla f(x)$ – the gradient of f evaluated at x;
for $f, g : \mathbb{R}^n \to \mathbb{R}^n$, $L_g f(x) = \lim\limits_{s\to 0} \frac{f(x+sg(x))-f(x)}{s}$;
$[f, g](x)=L_f g(x)-L_g f(x)$ – the Lie bracket of f, g.

II. Main results

A. Problem statement and extremum seeking algorithm

Consider the optimization problem

$$x^* = \arg \min_{x\in\mathbb{R}^n} J(x). \tag{2}$$

In extremum seeking problems, the mathematical expression of $J : \mathrm{R}^n \to \mathrm{R}$ is not known but J can be evaluated at any $x \in \mathbb{R}^n$. This paper discusses the case of so-called static maps (integrator plant dynamics), i.e., our continuous-time algorithm is of the form $\dot{x} = u$ such that the to-be-designed time-varying control law $u = u(t, J(x(t))$ achieves $\lim_{t\to\infty} x(t) = x^*$ for a suitable set of initial data. This choice of problem setup is justified by the need to clarify concepts and to keep technical steps rather simple. Extensions to more general cases, such as dynamic maps or constrained problems, are reasonable.

To solve the above extremum seeking problem, consider

$$\dot{x}=\gamma(t)\sum_{j=1}^{n}\big(F_{1j}\left(J(x)\right)u^{\varepsilon}_{1j}(t) + F_{2j}\left(J(x)\right)u^{\varepsilon}_{2j}(t)\big)\,e_j, \tag{3}$$

where $x \in D$ is the state, $D \subset \mathbb{R}^n$ is a domain, $x^* \in D$. For the sake of simplicity, we take the time varying inputs as

$$\begin{aligned} u^{\varepsilon}_{1j}(t)&=2\sqrt{\frac{\pi j}{\varepsilon}}\cos\left(\frac{2\pi j t}{\varepsilon}\right),\\ u^{\varepsilon}_{2j}(t)&=2\sqrt{\frac{\pi j}{\varepsilon}}\sin\left(\frac{2\pi j t}{\varepsilon}\right), \end{aligned} \tag{4}$$

however, other choices are possible, see, e.g., [6]. The cost-dependent functions $F_{1j}, F_{2j} \in C^1(D)$ are such that $[F_{1j}, F_{2j}](z) = -1$ for all $z \in \mathbb{R}$. The function γ can be thought of as a time-varying gain and corresponds to a step-size rule when discretizing the equations for example with Euler's method. System (3) with a time-invariant gain function (constant step-size) $\gamma(t) = \gamma_0 \in \mathbb{R}$ is well-known in extremum-seeking systems. Without additional assumptions on F_1 and F_2, its solutions exhibit practical uniform asymptotic convergence to the minimizer x^* of objective function J, as shown in [10], [23] and many other works on the Lie bracket approximations of extremum-seeking systems. For example, in the paper [10], it is proven that it is possible to achieve uniform asymptotic stability of x^* by the proper choice of F_1 and F_2 vanishing at x^*. However, the use of such functions requires the knowledge of the minimal value of $J(x)$, which can be restrictive for general applications.

Remark 1: In [10], [14] we have proposed the following formula for generating F_{1j}, F_{2j}:

$$F_{1j}(z) = f(z)\sin\phi(z), F_{2j}(z) = f(z)\cos\phi(z),$$

where $f \in C(D)$ and $\phi \in C^1(D)$ are such that $r^2(z)\frac{d\phi(z)}{dz} = 1$. Some possible choices of the pairs F_{1j}, F_{2j} are, for example, $F_{1j}(z) = z$, $F_{2j}(z) = 1$ [6], $F_{1j}(z) = \sin z$, $F_{2j}(z) = \cos z$ [23], $F_{1j}(z) = (\tanh\frac{z}{2})^{1/2}\sin(2\ln(e^z - 1) - z)$, $F_{2j}(z) = (\tanh\frac{z}{2})^{1/2}\cos(2\ln(e^z - 1) - z)$ [14]. We impose the following assumptions.

Assumption 1 (Properties of J): The function $J \in C^2(D)$ has a unique extremum point x^* in D and, for all $x \in D$,

- $(\nabla J(x), x - x^*) \geq \mu_1\|x - x^*\|^2$, with some $\mu_1 > 0$;
- $\|\nabla J(x)\| \leq \mu_2\|x - x^*\|$ for all $x \in D$, with some $\mu_2 > 0$.

Assumption 2 (Regularity assumption): The functions $F_{1j}, F_{2j}(J(\cdot)) \in C^2(D \setminus \{x^*\}; \mathbb{R})$, $D \subseteq \mathbb{R}^n$; the functions $L_{F_{pj}}F_{si}(J(\cdot))$, $L_{F_{qk}}L_{F_{pj}}F_{si}(J(\cdot)) \in C(D; \mathbb{R})$, for all $p, s, q \in \{1, 2\}$, $i, j \in \{1, 2, \ldots, n\}$.

B. Convergence properties

Using for example the approach of [26], one can prove semi-global practical asymptotic stability of the point x^* for system (3) provided that $\gamma(t)$ and $\gamma'(t)$ are bounded. This means that the trajectories of the system (3) converge to a neighborhood of the point x^* (where the size of the neighborhood depends on ε) but in general not to the point x^*. In the main results of this paper, we show that under a suitable choice of γ, the trajectories of (3) indeed converge asymptotically to the point x^*.

Theorem 1: Consider system (3) *and suppose the Assumptions 1 and 2 are satisfied. Let $\gamma \in C^1(\mathbb{R}^+; \mathbb{R}^+)$ be a monotonically decreasing function which satisfy*

(P1) $\lim_{t\to\infty}\int_0^t \gamma^2(\tau)d\tau = \infty$;

(P2) $\lim_{t\to\infty}\int_0^t \max\{\gamma^3(\tau), |\gamma'(\tau)|\}d\tau < \infty$.

Then for any $\delta \in (0, \mathrm{dist}(x^*, \partial D))$ *there exists an* $\hat{\varepsilon} > 0$ *such that the solutions* $x(t)$ *of system* (3) *with any* $\varepsilon \in (0, \hat{\varepsilon})$, $t_0 \geq 0$, *and* $x(t_0) \in B_\delta(x^*)$ *satisfy the property*

$$\lim_{t\to\infty} \|x(t) - x^*\| = 0.$$

In summary, (P2) in Theorem 1 guarantees that the step-size and its derivative converge to zero. Further, (P1) guarantees that the convergence of the step-size to zero is not too fast. Comparing (P1) in Theorem 1 with ($\tilde{\mathrm{P}}$1), we observe that $\gamma(t)$ plays the role of $\sqrt{\tilde{\gamma}(t)}$. This may initially appear surprising but is indeed natural when taking into account that the averaged (Lie bracket) system of (3) has the time-varying gain functions $\gamma^2(t)$ (see Remark 5 below). Comparing (P2) in Theorem 1 with ($\tilde{\mathrm{P}}$2), we see a difference between the common step-size rule condition in stochastic gradient descent and the one obtained in this paper. In particular, also the absolute value of the derivative of γ must approach zero. However, both conditions are qualitatively similar in the sense that (P2) and ($\tilde{\mathrm{P}}$2) ensure that the deterministic error terms (Chen-Fliess remainder terms) and the stochastic error terms (variances) in the convergence analysis to not destroy to the asymptotic convergence to x^*.

The proof of Theorem 1 and some further results about specific step-size rules as well as convergence speed properties are presented in the remainder of this section.

Remark 2: Let us underline that the convergence property established in Theorem 1 is non-uniform with respect to t_0. To be precise, $\lim_{t\to\infty}\|x(t)-x^*\| = 0$ in Theorem 1 means that for any $\rho > 0$, $\delta \in (0, \mathrm{dist}(x^*, \partial D))$, there exists an $\hat{\varepsilon} > 0$ and a $T_\rho = T_\rho(t_0, \varepsilon, \delta)$ with $\varepsilon \in (0, \hat{\varepsilon})$, such that $\|x(t) - x^*\| \leq \rho$ for all $t \geq T$. Nevertheless, $\hat{\varepsilon}$ can be chosen independently of t_0, as it follows from the proof of Theorem 1.

Remark 3: Theorem 1 requires $\gamma(t)$ to be decreasing for all $t \geq 0$. In fact, it is enough to assume that $\gamma(t)$ decreases for all $t \geq T$ with some $T \geq 0$. Furthermore, the requirement of a decreasing $\gamma(t)$ can be omitted if (P1) is replaced as: for any $t \geq 0$, $\varepsilon > 0$, $\sum_{j=0}^{\infty}\gamma^2(t_0+\varepsilon j) = \infty$. Moreover, the C^1-requirement for $\gamma(t)$ can be relaxed, e.g., by the well-definiteness of its Dini-derivatives or Lipschitz continuity.

Remark 4: It is worth noting that, while the controls of type (3) achieve asymptotic convergence for static maps, the (asymptotic) convergence speed may be slow. Further, it may lack of robustness against disturbances and the ability to track a time-varying optimum. These issues are common for controls with vanishing gains, see, e.g., in [24]. However, extremum seeking algorithms with vanishing gains can be considered as deterministic alternatives to stochastic approximation algorithms and thus maybe of interest in stochastic approximation problems. The analysis of (3) in the presence of disturbances or in case of a time-varying objective function represents interesting directions of future studies.

Corollary 1: Let $\gamma(t) = \lambda(\alpha+t)^{-\kappa}, \quad \lambda \neq 0, \alpha > 0, \kappa \in (1/3, 1/2]$. *Then the conditions of Theorem 1 are satisfied.*

Some other choices of γ are discussed in Section III.

Corollary 2: Consider the system

$$\dot{x} = \sum_{j=1}^{n}\left(F_{1j}\left(J(x)\right)u_{1j}^{\varepsilon}(\omega(\tau)) + F_{2j}\left(J(x)\right)u_{2j}^{\varepsilon}(\omega(\tau))\right)e_j, \quad (5)$$

where $\omega(\tau)$ *is invertible in* $\mathbb{R}^+$ *and* $\gamma(t) = \frac{1}{\omega'(\omega^{-1}(t))}$ *is in* $C^1(\mathbb{R}^+;\mathbb{R})$, *monotonically decreasing, and satisfies* $(P1)$–$(P2)$. *Then for any* $\delta \in (0, \mathrm{dist}(x^*, \delta D))$ *there is an* $\hat{\varepsilon} > 0$ *such that the solutions* $x(\tau)$ *of system* (5) *with any* $\varepsilon \in (0, \hat{\varepsilon})$, $\tau_0 \geq 0$, $x(\tau_0) \in B_\delta(x^*)$ *satisfy* $\lim_{t\to\infty}\|x(\tau) - x^*\| = 0$.

This result follows from the fact that (5) on the time scale $t = \omega(\tau)$ has exactly the form (3) with $\gamma_j(t) = \frac{1}{\omega'(\omega^{-1}(t))}$.

Corollary 3 (Convergence speed estimate): Under the conditions of Theorem 1,

$$\|x(t) - x^*\| \leq \nu\|x^0 - x^*\|e^{-\mu\int_{t_0}^{t}\gamma^2(s)ds} + \sqrt{\varepsilon}\zeta(t), \quad (6)$$

with some $\nu, \mu > 0$ which can be made arbitrary close to 1 and μ_1, respectively, by choosing a small enough ε, and a non-negative bounded function $\zeta(t)$ tending to 0 as $t \to \infty$.

The proof is given in Section II-D.

Remark 5: The estimate (6) can be also seen from the properties of solutions of the corresponding Lie bracket system. Although the approach of [6] requires uniform asymptotic stability properties for the Lie bracket system, we may formally derive it as

$$\bar{x} = -\gamma^2(t)\nabla J(\bar{x}),\; \bar{x}(t_0) = x^0.$$

Under Assumption 1, the time derivative of the function $V = \|x - x^*\|^2$ can be estimated as $\dot{V} \leq -2\gamma^2(t)\mu_1 V$. Integrating the obtained comparison inequality, we get $V(t) \leq V(t_0)e^{-2\mu_1\int_{t_0}^{t}\gamma^2(s)ds}$, or

$$\|x(t) - x^*\| \leq \|x^0 - x^*\|e^{-\mu_1\int_{t_0}^{t}\gamma^2(s)ds} \text{ for all } t \geq t_0.$$

This estimate shows the decay behavior of solutions and illustrates the importance of (P1). The term $e^{-\mu_1\int_{t_0}^{t}\gamma^2(s)ds}$ in (6) provides a measure for the decay behavior of the oscillatory solutions of (5) towards x^*. This term does not define an asymptotic decay rate due to the presence of $\sqrt{\varepsilon}\zeta(t)$. However, the constant ε in $\sqrt{\varepsilon}\zeta(t)$ can be made arbitrarily small, as can be seen from the proof of Theorem 1. Furthermore it also follows from the proof that the convergence speed of $\zeta(t)$ to zero is governed by the properties of $\gamma(t)$ and $\gamma'(t)$.

C. Proof of Theorem 1

To simplify the presentation, we assume $x \in D \subset \mathbb{R}$, i.e.

$$\dot{x} = \gamma(t)\left(F_1\left(J(x)\right)u_1^{\varepsilon}(t) + F_2\left(J(x)\right)u_2^{\varepsilon}(t)\right),$$

with $u_1^{\varepsilon}(t) = 2\sqrt{\frac{\pi}{\varepsilon}}\cos\left(\frac{2\pi t}{\varepsilon}\right)$, $u_2^{\varepsilon}(t) = 2\sqrt{\frac{\pi}{\varepsilon}}\sin\left(\frac{2\pi t}{\varepsilon}\right)$. The proof for $D \subset \mathbb{R}^n$ goes along the same lines.

Notations: For an $\varepsilon > 0$, $j \in \mathbb{N} \cup \{0\}$, denote $t_j = t_0 + j\varepsilon$, $x(t_j) = x^j$. From $(P2)$, there is an $M_\gamma > 0$ such that

$$\varphi(t) := \max\{\gamma^3(t), |\gamma'(t)|\} \leq M_\gamma \text{ for all } t \geq 0. \quad (7)$$

Let $D_0 \subset D$ be a closed domain, $x^* \in D_0$. Define
$M_F = \sup_{x\in D_0, s=1,2} \|F_s(J(x))\|$,
$M_{2F} = \sup_{x\in D_0, i,j=1,2} \|L_{F_j} F_i(J(x))\|$,
$M_{3F} = \sup_{x\in D_0, i,j,\ell=1,2} \|L_{F_\ell} L_{F_j} F_i(J(x))\|$.
The proof of the theorem proceeds in several steps. In steps a)-d), we assume that for any $\delta>0$ with $B_\delta(x^*) \subset D_0$, there is an $\varepsilon_0>0$ such that for any $t_0 \geq 0$, the solutions $x(t)$ of (3) with $\varepsilon \in (0, \varepsilon_0)$ and $x(t_0) = x^0 \in B_\delta(x^*)$ are well-defined in D for all $t \geq t_0$. Independently of ε_0, we define $\varepsilon_1 > 0$ such that $\|x(t) - x^*\| \to \infty$ as $t \to \infty$, provided that $\varepsilon \in (0, \varepsilon_1)$. The value ε_0 will be specified in the final step e), when some auxiliary constants will be introduced. Subsequently, the assertion of the theorem will be proved with $\hat{\varepsilon} = \min\{\varepsilon_0, \varepsilon_1\}$.

a) Representation of solutions: Expanding the solutions of system (3) into the Chen–Fliess-like series (similarly to [26]), we represent the solution of system (3) with initial data $x(t_j) = x^j \in D_0$ at time t_{j+1} as

$$x^{j+1} = x^j - \varepsilon\gamma^2(t_j)\nabla J(x^j) + R(t_j, \varepsilon),$$

with $R(t_j,\varepsilon) = -\int_{t_j}^{t_{j+1}} \int_{t_j}^{t} \gamma'(s) \sum_{i=1}^{2} F_i(J(x(s))) u_i^\varepsilon(t) ds\, dt$

$$+ \int_{t_j}^{t_{j+1}} \int_{t_j}^{t} \int_{t_j}^{s} \gamma(p) \left(\sum_{i,j=1}^{2} \left(\gamma^2(p) L_{F_\ell} L_{F_j} F_i(J(x(p))) u_\ell^\varepsilon(p) \right.\right.$$
$$\left.\left. -2\gamma'(p) L_{F_j} F_i(J(x(p)))\right)\right) u_i^\varepsilon(t) u_j^\varepsilon(s) dp\, ds\, dt.$$

Such a representation follows from the zero-mean and periodicity properties of inputs (4) and the relation between F_1, F_2. For more details we refer to, e.g., [6], [10]. Using Assumption 1 and estimate (7), we get

$$\|x^j - x^* - \varepsilon\gamma^2(t_j)\nabla J(x^j)\|^2 \leq \|x^j - x^*\|^2 \Big(1 - \varepsilon\gamma^2(t_j)(2\mu_1 - \varepsilon M_\gamma^2 \mu_2)\Big).$$

For any $\mu \in (0, \mu_1)$, let $\varepsilon_1 = \min\left\{\frac{2(\mu_1-\mu)}{M_\gamma^2\mu_2}, \frac{1}{2M_\gamma^2\mu_1}\right\}$. Then $1 - 2\varepsilon\mu\gamma^2(t_j) > 0$ for all $t_j \geq 0$, $\varepsilon \in (0, \varepsilon_1)$, and

$$\|x^j - x^* - \varepsilon\gamma^2(t_j)\nabla J(x^j)\| \leq \|x^j - x^*\| \Big(1 - \varepsilon\mu\gamma^2(t_j)\Big). \quad (8)$$

b) Estimates of $\|R(t_j, \varepsilon)\|$: From (7),

$$\|R(t_j,\varepsilon)\| \leq 2M_F \sqrt{\frac{2\pi}{\varepsilon}} \int_{t_j}^{t_{j+1}} \int_{t_j}^{t} \varphi(s) ds\, dt + \frac{16\pi}{\varepsilon^{3/2}}$$
$$\times \left(\sqrt{\varepsilon} M_{2F} M_\gamma^{1/3} + M_{3F}\sqrt{2\pi}\right) \int_{t_j}^{t_{j+1}} \int_{t_j}^{t} \int_{t_j}^{s} \varphi(p) dp\, ds\, dt.$$

Integration by parts and assumption $\varepsilon \in (0, \varepsilon_1)$ yields

$$\|R(t_j,\varepsilon)\| \leq \sqrt{\varepsilon} C_R \int_{t_j}^{t_{j+1}} \varphi(t) dt \quad (\leq \varepsilon^{3/2} C_R M_\gamma). \quad (9)$$

where $C_R = \left(2M_F\sqrt{2\pi} + 16\pi(\sqrt{\varepsilon_1} M_{2F} M_\gamma^{1/3} + \sqrt{2\pi} M_{3F})\right)$. Estimates (8)–(9) imply that, for all $j \in \mathbb{N} \cup \{0\}$, $\varepsilon \in (0, \varepsilon_1)$,

$$\|x^{j+1} - x^*\| \leq \|x^j - x^*\| \big(1 - \varepsilon\mu\gamma^2(t_j)\big) + \sqrt{\varepsilon} C_R \int_{t_j}^{t_{j+1}} \varphi(t) dt. \quad (10)$$

c) Estimate of $\|x^{j+1} - x^*\|$: Denote

$$\lambda_0 = 1 - \varepsilon\mu\gamma^2(t_0),\ \lambda_j = \prod_{\ell=0}^{j} \left(1 - \varepsilon\mu\gamma^2(t_j)\right),$$
$$r_0 = C_R \int_{t_0}^{t_1} \varphi(t) dt,$$
$$r_j = r_{j-1}\left(1 - \varepsilon\mu\gamma^2(t_j)\right) + C_R \int_{t_j}^{t_{j+1}} \varphi(t) dt \text{ for } j \geq 1. \quad (11)$$

Then

$$\|x^1 - x^*\| \leq \|x^0 - x^*\| \left(1 - \varepsilon\mu\gamma^2(t_0)\right) + \sqrt{\varepsilon} C_R \int_{t_0}^{t_1} \varphi(t) dt$$
$$= \|x^0 - x^*\|\lambda_0 + \sqrt{\varepsilon} r_0,$$
$$\|x^2 - x^*\| \leq \|x^1 - x^*\| \left(1 - \varepsilon\mu\gamma^2(t_1)\right) + \sqrt{\varepsilon} C_R \int_{t_1}^{t_2} \varphi(t) dt$$
$$\leq \|x^0 - x^*\| \left(1 - \varepsilon\mu\gamma^2(t_0)\right)\left(1 - \varepsilon\mu\gamma^2(t_1)\right)$$
$$+ \sqrt{\varepsilon} r_0 \left(1 - \varepsilon\mu\gamma^2(t_1)\right) + \sqrt{\varepsilon} C_R \int_{t_1}^{t_2} \varphi(t) dt$$
$$= \|x^0 - x^*\|\lambda_1 + \sqrt{\varepsilon} r_1, \text{ etc.}$$

Thus, we may conclude that, for any $j \in \mathbb{N}$,

$$\|x^{j+1} - x^*\| \leq \|x^0 - x^*\|\lambda_j + \sqrt{\varepsilon} r_j. \quad (12)$$

Let us prove that $\lim_{j\to\infty} \|x^{j+1} - x^*\| = 0$. With this purpose, consider the sequences $\{\lambda_j\}_{j\in\mathbb{N}}$ and $\{r_j\}_{j\in\mathbb{N}}$. The values λ_j in (11) can be estimated as

$$0 \leq \lambda_j = \prod_{\ell=0}^{j} \left(1 - \varepsilon\mu\gamma^2(t_\ell)\right) \leq e^{-\varepsilon\mu \sum_{\ell=0}^{j} \gamma^2(t_\ell)}.$$

Recall that by (P1), $\lim_{j\to\infty} \sum_{\ell=0}^{j} \gamma^2(t_\ell) = \infty$. Since $\lambda_j \geq 0$, this implies $\lim_{j\to\infty} \lambda_j = 0$. Consider now the sequence $\{r_j\}_{j\in\mathbb{N}}$. Observe that the elements r_j in (11) can be represented as

$$r_j = C_R \sum_{\ell=0}^{j} \int_{t_\ell}^{t_{\ell+1}} \varphi(t) dt \prod_{m=\ell+1}^{j} \left(1 - \varepsilon\mu\gamma^2(t_m)\right). \quad (13)$$

Let us prove that $\lim_{j\to\infty} r_j = 0$, i.e. for any $\rho>0$ there is an $N_\rho \in \mathbb{N}$ such that $r_j < \rho$ for all $j > N_\rho$. By (P2) and Cauchy's convergence test, for any $\rho>0$ there is an $N_{\rho,1} \in \mathbb{N}$ such that

$$\int_{t_{\ell_1}}^{t_{\ell_2}} \varphi(t) dt < \frac{\rho}{2C_R} \text{ for all } \ell_2 > \ell_1 > N_{\rho,1}. \quad (14)$$

As $\lim_{j\to\infty} e^{-\varepsilon\mu \sum_{m=\ell+1}^{j} \gamma^2(t_m)} = 0$ for any $\ell > 0$, then for any $\rho > 0$ and the corresponding $N_{\rho,1}$ there is an $N_{\rho,2} \in \mathbb{N}$ such that

$$e^{-\varepsilon\mu\sum\limits_{\ell=N_{\rho,1}+1}^{j}\gamma^2(t_\ell)} < \frac{\rho}{2(N_{\rho,1}+1)M_\gamma C_R} \text{ for all } j > N_{\rho,2}. \quad (15)$$

Let us fix $N_\rho = \max\{N_{\rho,1}, N_{\rho,2}\}$. Then for any $j > N_\rho$,

$$\begin{aligned}
&\sum_{\ell=0}^{j}\int_{t_\ell}^{t_{\ell+1}}\varphi(t)dt\prod_{m=\ell+1}^{j}\left(1-\varepsilon\mu\gamma^2(t_m)\right)\\
&=\sum_{\ell=0}^{N_{\rho,1}}\int_{t_\ell}^{t_{\ell+1}}\varphi(t)dt\prod_{m=\ell+1}^{j}\left(1-\varepsilon\mu\gamma^2(t_m)\right)\\
&\quad+\sum_{\ell=N_{\rho,1}+1}^{j}\int_{t_\ell}^{t_{\ell+1}}\varphi(t)dt\prod_{m=\ell+1}^{j}\left(1-\varepsilon\mu\gamma^2(t_m)\right)\\
&\le\varepsilon(N_{\rho,1}+1)M_\gamma e^{-\varepsilon\mu\sum\limits_{\ell=N_{\rho,1}+1}^{j}\gamma^2(t_\ell)}+\int_{t_{N_{\rho,1}+1}}^{t_{j+1}}\varphi(t)dt.
\end{aligned}$$

Inserting (14)–(15) into the above estimate, we conclude that $r_j < \rho$ for all $j \ge N_\rho + 1$, which proves that $\lim\limits_{j\to\infty} r_j = 0$.

All in all, we conclude that $\lim\limits_{j\to\infty}\|x^{j+1}-x^*\| = 0$.

d) Convergence of the solutions as $t\to\infty$: For any $t \ge 0$,

$$\|x(t)-x^*\| \le \left\|x\left(t_0+\hat{j}\varepsilon\right)-x^*\right\|+\left\|x(t)-x\left(t_0+\hat{j}\varepsilon\right)\right\|,$$

where $\hat{j} = \left[\frac{t-t_0}{\varepsilon}\right]$. Since $t - t_0 - \left[\frac{t-t_0}{\varepsilon}\right]\varepsilon \in [0,\varepsilon)$,

$$\begin{aligned}
\left\|x(t)-x\left(t_0+\hat{j}\varepsilon\right)\right\| &\le 2\sqrt{\frac{2\pi}{\varepsilon}}M_F\int_{t_0+\hat{j}\varepsilon}^{t}\gamma(\tau)d\tau\\
&\le 2\sqrt{2\pi\varepsilon}M_F\sup_{t_0+\hat{j}\varepsilon\le s\le t}\gamma(s),
\end{aligned}$$

what follows from the integral representation of solutions.

Assumptions of the theorem imply $\gamma(t) \to 0$ as $t \to \infty$. Hence, as $t \to \infty$,

$$\|x(t)-x^*\| \le \|x^{\hat{j}}-x^*\| + 2\sqrt{2\pi\varepsilon}M_F\sup_{t_0+\hat{j}\varepsilon\le s\le t}\gamma(s)\to 0.$$

e) Well-definiteness of the solutions: It remains to find a small enough ε_0 such that the solutions of system (3) with $\varepsilon \in (0, \min\{\varepsilon_0, \varepsilon_1\})$ and $x(t_0) \in B_\delta(x^*)$ are well-defined in D_0 for all $t \ge t_0$, i.e. that $x(t) \in D_0$ for all $t \ge t_0$. Let δ_0 be such that $B_\delta(x^*) \subset B_{\delta_0}(x^*) \subset D_0$, and $d_0 = \mathrm{dist}(\partial B_{\delta_0}(x^*), D_0)$. Using the integral representation of solutions with $x(t_j) = x^j$, $j \in \mathbb{N}\cup\{0\}$, we obtain

$$\|x(t)-x^j\| \le 2\sqrt{2\pi}M_\gamma M_F\frac{t-t_j}{\sqrt{\varepsilon}} \le c_x\sqrt{\varepsilon}$$

with $c_x = 2\sqrt{2\pi}M_\gamma M_F$. Then, for any $x^j \in B_{\delta_0}(x^*)$,

$$\|x(t)-x^*\| \le \|x^j-x^*\| + \|x(t)-x^j\| \le \delta_0 + c_x\sqrt{\varepsilon}.$$

Let $\varepsilon_{01} = (d_0/c_x)^2$. Then for any $\varepsilon \in (0,\varepsilon_{01})$, $t \in [t_j, t_{j+1}]$, $x(t) \in D_0$ provided that $x^j \in B_{\delta_0}(x^*)$. It remains to define an $\varepsilon_{02} > 0$ such that $x^j \in B_{\delta_0}(x^*)$ fro all $j \in \mathbb{N}$. Recall the formulas (12) and (13), which imply

$$\|x^{j+1}-x^*\| \le \|x^0-x^*\| + \sqrt{\varepsilon}C_R\int_{t_0}^{t_{j+1}}\varphi(t), \quad (16)$$

whenever $x(t) \in D_0$ for $t \in [t_0, t_j]$. Similarly to (14), for any fixed $d > 0$ there exists an $N_d > 0$ such that $\int_{t_{\ell_1}}^{t_{\ell_2}}\varphi(t)dt < d$ for all $\ell_2 > \ell_1 > N_d$. Note that N_d does not depend on ε and t_0. Given any $\varepsilon > 0$, let $j_d = [N_d/\varepsilon]$, where $[\dots]$ stand for the integer part. Then, for any $x^0 \in B_\delta(x^*)$, $t_0 \ge 0$, $j = 1,\dots,j_d$, estimate (16) yields

$$\|x^j - x^*\| \le \delta + \sqrt{\varepsilon}C_R M_\gamma j_d\varepsilon = \delta + \sqrt{\varepsilon}C_R M_\gamma N_d.$$

Furthermore, for any $j = j_d+1, j_d+2, \dots$,

$$\begin{aligned}
\|x^j-x^*\| &\le \|x^0-x^*\| + \sqrt{\varepsilon}C_R\left(\int_{t_0}^{t_{j_d}}\varphi(t)+\int_{t_{j_d+1}}^{t_j}\varphi(t)\right)\\
&\le \delta + \sqrt{\varepsilon}C_R(M_\gamma N_d + d).
\end{aligned}$$

Taking $\varepsilon_{12} = \left(\frac{\delta_0-\delta}{C_R(M_\gamma N_d+d)}\right)^2$, we conclude that, for any $\varepsilon \in (0,\varepsilon_{02})$, $x^j \in B_{\delta_0}(x^*)$. Thus, the solutions of system (3) with $\varepsilon \in (0, \varepsilon_0 = \min\{\varepsilon_{01},\varepsilon_{02}\})$ and $x(t_0) \in B_\delta(x^*)$ are well-defined in D_0 for all $t \ge t_0$. Let us underline that ε_2 does not depend on j_d, allowing us to establish the well-definiteness property step-by-step for intervals $[t_0, t_1]$, $[t_1, t_2]$, and so forth, similarly to our approach in [10]. Thus, we complete the proof by defining $\hat{\varepsilon} = \min\{\varepsilon_0, \varepsilon_1\}$.

D. Proof of Corollary 3

As it follows from the proof of Theorem 1 (cf. Section II-C, parts d) and e)), the solutions of system (3) satisfies the following estimate, for all $t \ge t_0$:

$$\|x(t)-x^*\| \le \|x^0-x^*\|e^{-\varepsilon\mu\sum\limits_{\ell=0}^{\hat{j}-1}\gamma^2(t_0+\varepsilon\ell)} + \zeta(t)\sqrt{\varepsilon},$$

with the bounded non-negative function $\zeta(t)$ can be explicitly defined from the proof of Theorem 1), and $\hat{j} = \left[\frac{t-t_0}{\varepsilon}\right]$.Then we may estimate $\sum\limits_{\ell=0}^{\hat{j}-1}\gamma^2(t_0+\ell\varepsilon)$ as

$$\begin{aligned}
\sum_{\ell=0}^{\hat{j}-1}\gamma^2(t_0+\varepsilon\ell) &\ge \int_0^{\hat{j}}\gamma^2(t_0+\varepsilon s)ds = \frac{1}{\varepsilon}\int_{t_0}^{t_0+\varepsilon\hat{j}}\gamma^2(s)ds\\
&= \frac{1}{\varepsilon}\int_{t_0}^{t}\gamma^2(s)ds - \frac{1}{\varepsilon}\int_{t_0+\varepsilon\hat{j}}^{t}\gamma^2(s)ds\\
&\ge \frac{1}{\varepsilon}\int_{t_0}^{t}\gamma^2(s)ds - \gamma^2(t_0).
\end{aligned}$$

Thus,

$$\|x(t)-x^*\| \le \|x^0-x^*\|e^{-\mu\int_{t_0}^{t}\gamma^2(s)ds+\varepsilon\mu\gamma^2(t_0)} + \zeta(t)\sqrt{\varepsilon},$$

implying the assertion of the corollary with $\nu = e^{\varepsilon\mu\gamma^2(t_0)}$.

III. Simulations

Consider system (3) with $F_1(z) = \sin(x)$, $F_2(z) = \cos(z)$:

$$\dot{x} = \gamma(t)\left(\sin\left(J(x)\right)u_1^\varepsilon(t) + \cos\left(J(x)\right)u_2^\varepsilon(t)\right). \quad (17)$$

For numerical simulations, we set $J(x) = \frac{x^2}{2}$, $\varepsilon = 0.1$. Figure 1 shows the trajectory of the system with $x(0) = 1$ and $\gamma(t)$ from Corollary 1, $\gamma(t) = \frac{\lambda}{(\alpha+t)^\kappa}$, with $\alpha = 0.1$,

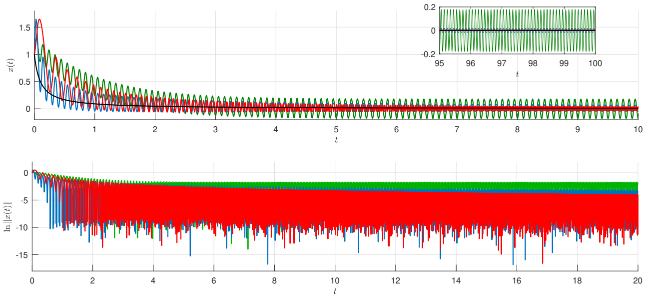

Fig. 1. Time plots of the trajectories of systems (17) and (18) (top) and the corresponding functions $\ln\|x(t)\|$ (bottom). Curves in blue, green, and red correspond to (17) with $\gamma(t)=\frac{1}{\sqrt{0.1+t}}$, $\gamma(t)\equiv 1$, and to (18), respectively. The plot of the function $\|x(0)\|e^{-\int_0^t\gamma^2(s)ds}=\frac{1}{1+10t}$ is depicted in black (top).

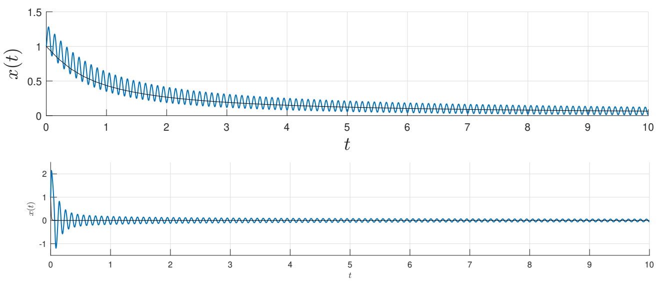

Fig. 2. Time plots of the trajectories of system (17) with $\gamma(t)=\sqrt{\sin\frac{1}{0.5+t}}$ (blue, top) and $\gamma(t)=\frac{1}{\sqrt{0.1+t}}+10e^{-10t}$ (blue, bottom), $x(0)=1$, and the plots of the corresponding functions $\|x(0)\|e^{-\int_0^t\gamma^2(s)ds}$ (black).

$\lambda=1$, $\kappa=\frac{1}{2}$. To illustrate Corollary 3, we add the graph of $\psi(t)=\|x^0-x^*\|e^{-\mu\int_{t_0}^t\gamma^2(s)ds}$. As mentioned in Remark 5, the trajectory of (17) oscillates around the curve $\psi(t)$ with vanishing amplitude of oscillations. In this case, $t_0=0$, $x^0=1$, $\mu=1$. Figure 1 also shows the solution of system (3) with $\gamma(t)\equiv 1$ to compare the behavior of (3) with an extremum seeking algorithm with time-invariant gain. Furthermore, we consider system (5) with $F_1(z)=\sin(x)$, $F_2(z)=\cos(z)$, $\omega(\tau)=$ which corresponds to $\gamma(t)=\frac{\lambda}{(\alpha+t)^\kappa}$.

$$\dot{x}=\sin\left(J(x)\right)u_1^\varepsilon(\omega(t))+\cos\left(J(x)\right)u_2^\varepsilon(\omega(t)). \tag{18}$$

We put the same parameters for the numerical simulation: $\varepsilon=0.1$, $\alpha=0.1$, $\lambda=1$, $\kappa=\frac{1}{2}$, $x(0)=1$. The time-plot of the corresponding solution of system (18) is presented on Figure 1 in red. To better illustrate the behavior of all three systems, we have also included the corresponding logarithmic plots on Fig. 1, right. Of further interest is the investigation and comparison of other possible $\gamma(t)$. For example, Fig. 2 (left) shows the trajectories of system (17) with $\gamma(t)=\sqrt{\sin\frac{1}{0.5+t}}$. The other parameters remain the same as before. Although the trajectory of the system exhibits a slower decay rate, this example illustrates a variety of functions γ satisfying the conditions of Theorem 1. Fig. 2 (right) presents the results of numerical simulation for the case $\gamma(t)=\frac{1}{\sqrt{0.1+t}}+10e^{-10t}$.

REFERENCES

[1] M. Krstić and K. B. Ariyur, *Real-Time optimization by Extremum Seeking Control*. Wiley-Interscience, 2003.

[2] Y. Tan, D. Nešić, and I. Mareels, "On non-local stability properties of extremum seeking control," *Automatica*, vol. 42, no. 6, pp. 889–903, 2006.

[3] D. Nešić, "Extremum seeking control: Convergence analysis," in *Proc. 2009 European Control Conf.*, pp. 1702–1715, 2009.

[4] L. Fu and Ü. Özgüner, "Extremum seeking with sliding mode gradient estimation and asymptotic regulation for a class of nonlinear systems," *Automatica*, vol. 47, no. 12, pp. 2595–2603, 2011.

[5] S.-J. Liu and M. Krstić, *Stochastic averaging and stochastic extremum seeking*. Springer Science & Business Media, 2012.

[6] H.-B. Dürr, M. S. Stanković, C. Ebenbauer, and K. Johansson, "Lie bracket approximation of extremum seeking systems," *Automatica*, vol. 49, pp. 1538–1552, 2013.

[7] M. Guay and D. Dochain, "A time-varying extremum-seeking control approach," *Automatica*, vol. 51, pp. 356–363, 2015.

[8] A. Scheinker and M. Krstić, *Model-free stabilization by extremum seeking*. Springer, 2017.

[9] H.-B. Dürr, M. Krstić, A. Scheinker, and C. Ebenbauer, "Extremum seeking for dynamic maps using Lie brackets and singular perturbations," *Automatica*, vol. 83, pp. 91–99, 2017.

[10] 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.

[11] V. Grushkovskaya and A. Zuyev, "Extremum seeking approach for nonholonomic systems with multiple time scale dynamics," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 5392–5398, 2020.

[12] M. Abdelgalil and H. Taha, "Lie bracket approximation-based extremum seeking with vanishing input oscillations," *Automatica*, vol. 133, p. 109735, 2021.

[13] D. Bhattacharjee and K. Subbarao, "Extremum seeking control with attenuated steady-state oscillations," *Automatica*, vol. 125, p. 109432, 2021.

[14] V. Grushkovskaya and C. Ebenbauer, "Extremum seeking control of nonlinear dynamic systems using Lie bracket approximations," *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 7, pp. 1233–1255, 2021.

[15] J. I. Poveda, M. Benosman, A. R. Teel, and R. G. Sanfelice, "Robust coordinated hybrid source seeking with obstacle avoidance in multivehicle autonomous systems," *IEEE Transactions on Automatic Control*, vol. 67, no. 2, pp. 706–721, 2021.

[16] O. Romero and M. Benosman, "Time-varying continuous-time optimisation with pre-defined finite-time stability," *International Journal of Control*, vol. 94, no. 12, pp. 3237–3254, 2021.

[17] C. T. Yilmaz, M. Diagne, and M. Krstic, "Exponential extremum seeking with unbiased convergence," in *2023 62nd IEEE Conference on Decision and Control (CDC)*, pp. 6749–6754, IEEE, 2023.

[18] H. Liu, Y. Tan, and D. Oetomo, "A novel extremum seeking control to enhance convergence and robustness in the presence of nonlinear dynamic sensors," *Journal of Systems Science and Complexity*, vol. 37, no. 1, pp. 3–21, 2024.

[19] Y. Tan, W. Moase, C. Manzie, D. Nešić, and I. M. Y. Mareels, "Extremum seeking from 1922 to 2010," in *Proc. 29th Chinese Control Conf.*, pp. 14–26, 2010.

[20] A. Scheinker, "100 years of extremum seeking: A survey," *Automatica*, vol. 161, p. 111481, 2024.

[21] R. Suttner and S. Dashkovskiy, "Exponential stability for extremum seeking control systems," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 15464–15470, 2017.

[22] S. Meyn, *Control Systems and Reinforcement Learning*. Cambridge University Press, 2022.

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

[24] C. T. Yilmaz, M. Diagne, and M. Krstic, "Exponential and prescribed-time extremum seeking with unbiased convergence," *arXiv preprint arXiv:2401.00300*, 2023.

[25] C. T. Yilmaz, M. Diagne, and M. Krstic, "Perfect tracking of time-varying optimum by extremum seeking," *arXiv:2402.14178v1*, 2024.

[26] V. Grushkovskaya and A. Zuyev, "Motion planning and stabilization of nonholonomic systems using gradient flow approximations," *Nonlinear Dynamics*, vol. 111, no. 23, pp. 21647–21671, 2023.