Extremum Seeking Control: Convergence Analysis*

Dragan Nešić

***Abstract*— This paper summarizes our recent work on dynamical properties for a class of extremum seeking (ES) controllers that have attracted a great deal of research attention in the past decade. Their local stability properties were already investigated, see [2]. We first show that semi-global practical convergence is possible if the controller parameters are carefully tuned and the objective function has a unique (global) extremum. An interesting tradeoff between the convergence rate and the size of the domain of attraction of the scheme is uncovered: the larger the domain of attraction, the slower the convergence of the algorithm. The amplitude, frequency and shape of the dither signal are important design parameters in the extremum seeking controller. In particular, we show that changing the amplitude of the dither adaptively can be used to deal with global extremum seeking in presence of local extrema. Moreover, we show that the convergence of the algorithm is proportional to the power of the dither signal. Consequently, the square-wave dither yields the fastest convergence among all dithers of the same frequency and amplitude. We consider extremum seeking of a class of bioprocesses to demonstrate our results and motivate some open research questions for multi-valued objective functions.**

I. Introduction

In many engineering applications the system needs to operate close to an extremum of a given *objective (cost) function* during its steady-state operation. Moreover, the objective function is often not available analytically to the designer but instead one can measure the value of the objective function by probing the system.

Extremum seeking is an optimal control approach that deals with situations when the plant model and/or the cost to optimize are not available to the designer but it is assumed that measurements of plant input and output signals are available. Using these available signals, the goal is to design an extremum seeking controller that dynamically searches for the optimizing inputs. This situation arises in a range of classical, as well as certain emerging, engineering applications. Indeed, this method was successfully applied to biochemical reactors [9], [4], ABS control in automotive brakes [8], variable cam timing engine operation [14], electromechanical valves [13], axial compressors [21], mobile robots [11], mobile sensor networks [5], [12], optical fibre amplifiers [7] and so on [2]. A good survey of the literature on this topic prior to 1980 can be found in [16] and a more recent overview can be found in [2]. Åström and Wittenmark rated extremum seeking as one of the most promising adaptive control methods [1, Section 13.3].

There are two main approaches to extremum seeking: (i) adaptive control extremum seeking; (ii) nonlinear programming based extremum seeking. Adaptive control methods provide a range of adaptive controllers that solve the extremum seeking problem for a large class of systems [2]. The controller makes use of a certain excitation (dither) signal which provides the desired sub-optimal behaviour if the controller parameters are tuned appropriately. On the other hand, nonlinear programming based extremum seeking methods combine the classical nonlinear programming methods for numerical optimization with an approximate online generation of the gradient of the objective function by applying constant probing inputs successively [20].

The main goal of this paper is to report on our recent results on stability properties of a class of adaptive extremum seeking controllers. The first local stability analysis of this class of controllers was reported in 2000 by Krstić and Wang [10]. This seminal paper used techniques of averaging and singular perturbations to show that if the adaptive extremum seeking controller is tuned appropriately, then sub-optimal extremum seeking is achieved if the system is initialized close to the extremum.

We introduced a simplified adaptive scheme in [17] where it was shown under slightly stronger conditions that non-local (even semi-global) extremum seeking is achieved if the controller is tuned appropriately. Moreover, by using the singular perturbations techniques and averaging, we demonstrated that this simplified scheme operates on average in its slow time scale as the steepest descent optimization scheme. We reported a detailed analysis of this simplified scheme in [17]. In [19] we analysed the flexibility in choosing the shape of the excitation dither signal to ensure faster convergence. It was shown for static maps that a square wave dither yields fastest convergence over all dither signals with the same amplitude and frequency. We reported conditions that ensure global extremum seeking in the presence of local extrema in [18]. Adaptive schemes with multi-valued objective functions that arise, for instance, in bioprocesses, were investigated in [4]. Multi-valued functions pose some open research questions that we briefly mention in the last section. In the sequel, we present an overview of our recent results in [4], [17], [18], [19].

Mathematical preliminaries: We denote the set of real numbers as $\mathbb{R}$. Given a sufficiently smooth function $h : \mathbb{R}^p \to \mathbb{R}$, we denote its i^{th} derivative with respect to j^{th} variable as $D^i_j h(x_1, \ldots, x_p)$. When $i = 1$ and $j = 1$ we write simply $Dh(x_1, \ldots, x_p) := D^1_1(x_1, \ldots, x_p)$. The continuous

*This work was supported by the Australian Research Council under the Discovery Grants and Australian Professorial Fellow schemes. The author would like to thank Y. Tan, I. M. Y. Mareels and G. Bastin for the fruitful collaboration that has lead to this work.

D. Nešić is with Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, 3052, Victoria, Australia. E-mail: d.nesic@ee.unimelb.edu.au

function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is of class $\mathcal{KL}$ if it is nondecreasing in its first argument and converging to zero in its second argument. Given a measurable function x, we define its $\mathcal{L}_\infty$ norm $\| \cdot \| = \operatorname{ess\,sup}_{t \geq 0} |x(t)|$.

We will show in the next section that the closed loop systems with an adaptive extremum seeking controller can be written as a parameterized family of systems:

$$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}, \boldsymbol{\varepsilon}) \ , \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$, $t \in \mathbb{R}_{\geq 0}$ and $\boldsymbol{\varepsilon} \in \mathbb{R}^\ell_{>0}$ are respectively the state of the system, the time variable and the parameter vector. The stability of the system (1) can depend in an intricate way on the parameter $\boldsymbol{\varepsilon}$ and we will need the following definition (see [17] for motivating examples):

Definition 1 *The system (1) with parameter $\boldsymbol{\varepsilon}$ is said to be semi-globally practically asymptotically (SPA) stable, uniformly in $(\varepsilon_1, \ldots, \varepsilon_j)$, $j \in \{1, \ldots, \ell\}$, if there exists $\beta \in \mathcal{KL}$ such that the following holds. For each pair of strictly positive real numbers (Δ, ν), there exist real numbers $\varepsilon_k^* = \varepsilon_k^*(\Delta, \nu) > 0, k = 1, 2, \ldots, j$ and for each fixed $\varepsilon_k \in (0, \varepsilon_k^*), k = 1, 2, \ldots, j$ there exist $\varepsilon_i = \varepsilon_i(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{i-1}, \Delta, \nu)$, with $i = j+1, j+2, \ldots, \ell$, such that the solutions of (1) with the so constructed parameters $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_\ell)$ satisfy:*

$$|\mathbf{x}(t)| \leq \beta(|\mathbf{x}_0|, (\varepsilon_1 \cdot \varepsilon_2 \cdot \cdots \cdot \varepsilon_\ell)(t - t_0)) + \nu, \tag{2}$$

for all $t \geq t_0 \geq 0$, $\mathbf{x}(t_0) = \mathbf{x}_0$ with $|\mathbf{x}_0| \leq \Delta$. If we have that $j = \ell$, then we say that the system is SPA stable, uniformly in $\boldsymbol{\varepsilon}$.

Note that in Definition 1 we can construct a small "box" around the origin for the parameters $\varepsilon_k, k = 1, 2, \ldots, j$ so that the stability property holds uniformly for all parameters in this box, whereas at the same time we can not do so for the parameters $\varepsilon_k, k = j+1, \ldots, l$. Sometimes we abuse terminology and refer to $(\varepsilon_1 \cdots \varepsilon_\ell)$ in the estimate (2) as the "convergence speed" (although the real convergence speed depends also on the function β).

II. An adaptive control scheme

Consider the system:

$$\dot{\mathbf{x}} = f(\mathbf{x}, u), \qquad y = h(\mathbf{x}) \ , \tag{3}$$

where $\mathbf{f} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ and $h : \mathbb{R}^n \to \mathbb{R}$ are sufficiently smooth. $\mathbf{x}$ is the measured state, u is the input and y is the output. We suppose that there exists a unique $\mathbf{x}^*$ such that $y^* = h(\mathbf{x}^*)$ is the extremum of the map $h(\cdot)$. Due to uncertainty, we assume that neither $\mathbf{x}^*$ nor $h(\cdot)$ is precisely known to the control designer. The main objective in extremum seeking control is to force the solutions of the closed loop system to eventually converge to $\mathbf{x}^*$ and to do so without any precise knowledge about $\mathbf{x}^*$ or $h(\cdot)$.

Consider a family of control laws of the following form:

$$u = \alpha(\mathbf{x}, \theta), \tag{4}$$

where $\theta \in \mathbb{R}$ is a scalar parameter. The closed-loop system (3) with (4) is then

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \alpha(\mathbf{x}, \theta)). \tag{5}$$

The requirement that θ is scalar and that (3), (4) is SISO is to simplify presentation. Multidimensional parameter situations can be tackled, see [2].

We proposed in [17] a first order extremum seeking scheme (see Figure1) that yields the following closed loop dynamics:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \alpha(\mathbf{x}, \hat{\theta} + a \sin(\omega t))) \tag{6}$$

$$\dot{\hat{\theta}} = kh(\mathbf{x}) b \sin(\omega t), \tag{7}$$

where (k, a, b, ω) are tuning parameters. Compared with the extremum seeking scheme in [10], the proposed extremum seeking scheme in Figure 1 is simpler, containing only an integrator (without low-pass and high-pass filters that are used in [10]).

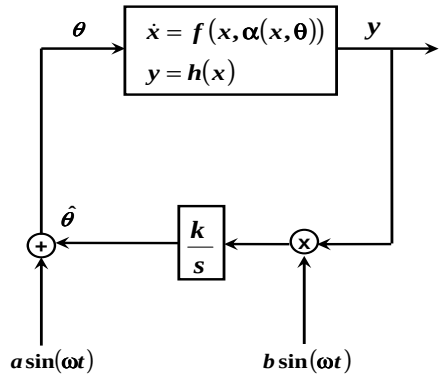

Fig. 1. A first order extremum seeking controller

III. Global extremum seeking in absence of local extrema

In this section we summarize results from [17] that provide guarantees for SPA stability under the following assumptions:

Assumption 1 *There exists a function $\mathbf{l} : \mathbb{R} \to \mathbb{R}^n$ such that $f(\mathbf{x}, \alpha(\mathbf{x}, \theta)) = 0$ if and only if $\mathbf{x} = \mathbf{l}(\theta)$.*

Assumption 2 *For each $\theta \in \mathbb{R}$, the equilibrium $\mathbf{x} = \mathbf{l}(\theta)$ of (5) is globally asymptotically stable, uniformly in θ.*

Assumption 3 *Denoting $Q(\cdot) = h \circ \mathbf{l}(\cdot)$, there exists a unique θ^* maximizing $Q(\cdot)$ and, the following holds*[1]*:*

$$DQ(\theta^*) = 0 \qquad D^2Q(\theta^*) < 0 \tag{8}$$
$$DQ(\theta^* + \zeta)\zeta < 0 \qquad \forall \zeta \neq 0 \ . \tag{9}$$

Note that (9) in Assumption 3 guarantees that there do not exist any local extrema. We will consider the case with local extrema in the next section.

Introduce the change of the coordinates, $\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{x}^*$, $\tilde{\theta} = \hat{\theta} - \theta^*$ and the system takes the form:

$$\begin{aligned} \dot{\tilde{\mathbf{x}}} &= \mathbf{f}(\tilde{\mathbf{x}} + \mathbf{x}^*, \alpha(\tilde{\mathbf{x}} + \mathbf{x}^*, \tilde{\theta} + \theta^* + a\sin(\omega t))) \\ \dot{\tilde{\theta}} &= kh(\tilde{\mathbf{x}} + \mathbf{x}^*)b\sin(\omega t). \end{aligned} \tag{10}$$

Note that the point $(\mathbf{x}^*, \theta^*)$ is in general *not an equilibrium point* of the system (6), (7). We introduce $k \stackrel{\triangle}{=} \omega\delta K$, $\sigma \stackrel{\triangle}{=} \omega t$, where ω and δ are small parameters and $K > 0$ is fixed. The system equations expanded in time σ are:

$$\begin{aligned} \omega\frac{d\tilde{\mathbf{x}}}{d\sigma} &= \mathbf{f}(\tilde{\mathbf{x}} + \mathbf{x}^*, \alpha(\tilde{\mathbf{x}} + \mathbf{x}^*, \tilde{\theta} + \theta^* + a\sin(\sigma))) \\ \frac{d\tilde{\theta}}{d\sigma} &= \delta Kh(\tilde{\mathbf{x}} + \mathbf{x}^*)b\sin(\sigma). \end{aligned} \tag{11}$$

The system (11) has the form (1) where the parameter vector is defined as $\boldsymbol{\varepsilon} := [a \ b \ \delta \ \omega]^T$. For simplicity of presentation in the sequel we let $b = a$ and

$$\boldsymbol{\varepsilon} := [a^2 \ \delta \ \omega]^T \ . \tag{12}$$

The system (11) has a two-time-scale structure and our first main result is proved by applying the singular perturbations and averaging methods (see [17]).

Theorem 1 *Suppose that Assumptions 1, 2 and 3 hold. Then, the system (10) is SPA stable, uniformly in (a^2, δ).*

□

Theorem 1 provides the parameter tuning guidelines since it shows that to achieve a certain domain of attraction one first needs to reduce a and δ sufficiently and then for fixed values of these parameters reduce ω sufficiently. Hence, one can achieve any given domain of attraction but the convergence speed will be reduced simultaneously (cf. Definition 1). This tradeoff was first observed in [17]. In the convergence speed analysis of the extremum seeking scheme, the "worst case" convergence speed is considered. That is, the convergence speed of the overall system depends on the convergence speed of the slowest sub-system. The first order extremum seeking controller (10), according to Theorem 1, yields the following stability bound:

$$\begin{aligned} |\mathbf{z}(t)| &\leq \beta(|\mathbf{z}(t_0)|, (a^2\delta\omega)(t - t_0)) + \nu, \\ &= \beta\left(|\mathbf{z}(t_0)|, (a^2k)\frac{(t - t_0)}{K}\right) + \nu, \end{aligned} \tag{13}$$

for all $t \geq t_0 \geq 0$ and $|\mathbf{z}(t_0)| \leq \Delta$, where $\mathbf{z} \stackrel{\triangle}{=} (\tilde{\mathbf{x}}^T, \tilde{\theta}^T)^T$ and k, K were defined before. Since $K > 0$ is fixed, the parameter $a^2 \cdot k$ affects the convergence speed. The smaller $a^2 \cdot k$, the slower the convergence and the larger the domain of attraction.

Note that since $h(\cdot)$ is continuous, then for any $\nu > 0$, there exists $\nu_1 > 0$ such that

$$|\tilde{\mathbf{x}}| \leq \nu_1 \implies |h(\tilde{\mathbf{x}} + \mathbf{x}^*) - y^*| \leq \nu \ . \tag{14}$$

Theorem 1 can be interpreted as follows. For any (Δ, ν) we can adjust $\boldsymbol{\varepsilon}$ so that for all $|\mathbf{z}| \leq \Delta$ we have that $\limsup_{t\to\infty} |y(t) - y^*| \leq \nu$. In other words, the output of the system can be regulated arbitrarily close to the extremum value y^* from an arbitrarily large set of initial conditions by adjusting the parameters $\boldsymbol{\varepsilon}$ in the controller. In particular, the parameters $\boldsymbol{\varepsilon}$ are chosen so that Definition 1 holds with (Δ, ν_1) and ν_1 is defined in (14).

Theorem 1 is a stronger result than [10, Theorem 1] since we prove SPA stability, as opposed to local stability in [10]. However, our results are stated under stronger assumptions (Assumptions 1-3) than those in [10]. Assumptions 1-3 appear to be natural when non-local stability is investigated. Moreover, we note that it is not crucial in Assumptions 1 – 3 that all conditions hold globally. For instance, instead of requiring (9) in Assumption 3, we can assume:

$$DQ(\theta^* + \zeta)\zeta < 0 \qquad \forall \zeta \in \mathcal{D}, \ \zeta \neq 0 \ , \tag{15}$$

where $\mathcal{D}$ is a bounded neighborhood of θ^*. We note that these conditions are not very restrictive, whereas their global version is (Assumptions 2 and 3). Indeed, if the maximum is isolated and all functions are sufficiently smooth, we can conclude that the condition (8) implies that there exists a set $\mathcal{D}$ satisfying (15). Similarly, we could assume only local stability in Assumption 2. If all of our assumptions were regional (as opposed to global) we could still state SPA stability with respect to the given bounded region.

The proof of Theorem 1 in [17] provides an interesting insight into the way the extremum seeking controller operates. The parameter ω is used to separate time scales between the plant (boundary layer) and the extremum seeking controller (reduced system), where the plant states are fast and they quickly die out (Assumption 2). Using the singular perturbation method, we obtain that the reduced system in the variable "θ_r" in time "$\sigma = \omega t$" is time varying and it has the form:

$$\frac{d\theta_r}{d\sigma} = Ka\delta Q(\theta^* + \theta_r + a\sin(\sigma))\sin(\sigma) \ , \tag{16}$$

for which we introduce an "averaged" system:

$$\frac{d\theta_{av}}{d\sigma} = \frac{K}{2}a^2\delta \cdot DQ(\theta^* + \theta_{av}) \ . \tag{17}$$

Hence, the averaged system (17) can be regarded as the "gradient system" whose globally asymptotically stable equilibrium θ^* corresponds to the global maximum of the unknown map Q (Assumption 3).

As already indicated, the first order extremum seeking scheme works on average as a "gradient search" method. Both the excitation signal and the integrator are necessary

[1] Without loss of generality we assume that the extremum is a maximum.

to achieve this. The excitation signal $a\sin(\omega t)$ is added to system (3) to get probing while the multiplication (modulation) of output and the excitation signal extracts the gradient of the unknown mapping $Q(\cdot)$. The role of the integrator is to get on average the steepest decent along the gradient of $Q(\cdot)$. Hence, the first order scheme is the simplest controller structure that achieves extremum seeking.

Note that we did not prove SPA stability, uniform in the whole vector $\boldsymbol{\varepsilon}$ in Theorem 1 for system (10) with parameter $\boldsymbol{\varepsilon}$. To prove such a result we need a stronger assumption:

Assumption 4 *We have that (8) holds and there exists $\alpha_Q \in \mathcal{K}_\infty$ such that $DQ(\theta^* + \zeta)\zeta \leq -\alpha_Q(|\zeta|)$ for all $\zeta \in \mathbb{R}$.*

Next we will use Assumptions 4 and 5 that will be stated below after some auxiliary results are presented to prove semi-global stability results uniform in the parameter $\boldsymbol{\varepsilon}$.

We introduce "boundary layer" using $\bar{\mathbf{x}} := \tilde{\mathbf{x}} - \mathbf{l}(\theta^* + \tilde{\theta} + a\sin(\sigma)) =: \tilde{\mathbf{x}} - \tilde{\mathbf{l}}(\tilde{\theta} + a\sin(\sigma))$ and rewrite (11) in the time scale "t" as follows:

$$\dot{\tilde{\mathbf{x}}} = \tilde{\mathbf{f}}(\bar{\mathbf{x}} + \tilde{\mathbf{l}}(\tilde{\theta} + a\sin(\sigma)), \tilde{\theta} + a\sin(\sigma)) + \omega a\widetilde{\Delta}_1 \quad (18)$$

$$\dot{\tilde{\theta}} = a\delta K\tilde{h} \circ \tilde{\mathbf{l}}(\tilde{\theta} + a\sin(\sigma))\sin(\sigma) + a\delta\widetilde{\Delta}_2, \quad (19)$$

where $\tilde{\mathbf{f}}(\mathbf{x}, \theta) := \mathbf{f}(\mathbf{x}+\mathbf{x}^*, \alpha(\mathbf{x}+\mathbf{x}^*, \theta^*+\theta))$, $\tilde{h}(\mathbf{x}) := h(\mathbf{x}+\mathbf{x}^*)$ and $\widetilde{\Delta}_1, \widetilde{\Delta}_2$ are appropriate functions that depend on the state variables, parameters and time (see [17]). We consider the system (18), (19) as a feedback connection of two systems. It was shown in [17] that under our assumptions, the two systems are input to state stable (ISS) in an appropriate sense (see [15]):

Proposition 1 *Suppose that Assumption 4 holds. Then, there exist $\beta_1 \in \mathcal{KL}$ and for any $\Delta_1 > \nu_1 > 0$ and $\omega^* > 0$ there exist $\gamma_1^\theta, \gamma_2^\theta \in \mathcal{K}_\infty$, $a^* > 0$ and $\delta^* > 0$ such that for all $a \in (0, a^*)$, $\delta \in (0, \delta^*)$, $\omega \in (0, \omega^*)$ and $\max\{|\tilde{\theta}_0|, ||\bar{\mathbf{x}}||\} \leq \Delta_1$, with $\tilde{\theta}_0 := \tilde{\theta}(t_0)$ we have that the solutions of the subsystem (19) satisfy:*

$$|\tilde{\theta}(t)| \leq \max\left\{\beta_1(|\tilde{\theta}_0|, (a^2\delta\omega)(t-t_0)), \gamma_{\boldsymbol{\varepsilon}}^1(||\bar{\mathbf{x}}||), \nu_1\right\}, \quad (20)$$

for all $t \geq t_0 \geq 0$, where $\gamma_{\boldsymbol{\varepsilon}}^1(s) := \gamma_1^\theta\left(\frac{1}{a}\gamma_2^\theta(s)\right)$. □

Proposition 2 *Suppose that Assumptions 1 and 2 hold. Then, there exist $\beta_2 \in \mathcal{KL}$ and for any positive Δ_2, ν_2, a^* and δ^* there exist $\gamma_1^z, \gamma_2^z \in \mathcal{K}_\infty$ and $\omega^* > 0$, such that for all $a \in (0, a^*)$, $\delta \in (0, \delta^*)$ $\omega \in (0, \omega^*)$ and $\max\{|\bar{\mathbf{x}}_0|, ||\tilde{\theta}||\} \leq \Delta$, with $\bar{\mathbf{x}}_0 := \bar{\mathbf{x}}(t_0)$, we have that the solutions of the subsystem (18) satisfy:*

$$|\bar{\mathbf{x}}(t)| \leq \max\left\{\beta_2(|\bar{\mathbf{x}}_0|, t-t_0), \gamma_{\boldsymbol{\varepsilon}}^2(||\tilde{\theta}||), \nu_2\right\}, \quad (21)$$

for all $t \geq t_0 \geq 0$, where $\gamma_{\boldsymbol{\varepsilon}}^2(s) := \gamma_1^z(\omega a\gamma_2^z(s))$. □

Note that the ISS gains $\gamma_{\boldsymbol{\varepsilon}}^1$ and $\gamma_{\boldsymbol{\varepsilon}}^2$ in Propositions 1 and 2 depend on a and ω. Moreover, the gain $\gamma_{\boldsymbol{\varepsilon}}^1$ increases to infinity as a is reduced to zero. Typically, this behavior leads to lack of stability in the interconnection. However, in this case, it is possible to counteract this increase of $\gamma_{\boldsymbol{\varepsilon}}^1$ through the decrease of $\gamma_{\boldsymbol{\varepsilon}}^2$ as $\gamma_{\boldsymbol{\varepsilon}}^2$ decreases to zero as a decreases. Moreover, it is sometimes possible to achieve this in a manner that will guarantee SPA stability, uniform in $\boldsymbol{\varepsilon}$, see [17]. A condition that is needed for this to hold is summarized in the next assumption.

Assumption 5 *Let the gains $\gamma_{\boldsymbol{\varepsilon}}^1, \gamma_{\boldsymbol{\varepsilon}}^2$ come from Propositions 1 and 2. Assume that there exists $\gamma \in \mathcal{K}_\infty$ such that for any $0 < s_1 < s_2$ there exist ω^* and a^* such that for all $\omega \in (0, \omega^*)$, $a \in (0, a^*)$ and $s \in [s_1, s_2]$ we have that the following small gain conditions hold:*

$$\gamma_{\boldsymbol{\varepsilon}}^1 \circ \gamma_{\boldsymbol{\varepsilon}}^2(s) \leq \gamma(s) < s, \qquad \gamma_{\boldsymbol{\varepsilon}}^2 \circ \gamma_{\boldsymbol{\varepsilon}}^1(s) \leq \gamma(s) < s. \quad (22)$$

The following result was proved in [17]:

Theorem 2 *Suppose that Assumptions 1, 2, 4 and 5 hold. Then, the closed-loop system (10) is SPA stable, uniformly in $\boldsymbol{\varepsilon} = (a^2, \delta, \omega)$.* □

Note that the conditions (22) do not imply each other, as can be easily seen from the case when $\gamma_2^\theta \circ \gamma_1^z(s) = s^q$, $q > 1$ and $\gamma_2^z \circ \gamma_1^\theta(s) = s^p$, $p > 1$ in which case the conditions (22) become respectively $\gamma_1^\theta\left(\omega a^{q-1}\left(\gamma_2^z(s)\right)^q\right) < s$ and $\gamma_1^z\left(\omega a^{1-p}\left(\gamma_2^\theta(s)\right)^p\right) < s$. It is obvious, that in the first case we can choose ω and a independent of each other so that the first condition in (22) holds, whereas it is impossible to do so for the second condition in (22). This also illustrates that conditions of Assumption 5 may not hold for some gains. If all the gains $\gamma_1^\theta, \gamma_2^\theta, \gamma_1^z, \gamma_2^z$ are linear then Assumption 5 holds. In particular, the small gain conditions (22) become independent of a and can be achieved by reducing ω only.

It is an open question whether there is a genuine gap between Theorems 1 and 2, that is whether there exists an example satisfying all conditions of Theorem 1 but not conditions of Theorem 2 that *is not* SPA stable uniformly in $\boldsymbol{\varepsilon}$.

Example 1 *Consider the system:*

$$\dot{x} \;=\; -x + u^2 + 4u; \qquad y = -(x+4)^2\,. \quad (23)$$

Let the control input be $u = \theta$. Then, using our notation we have $\theta^* = -2$, $x^* = -4$ and $y^* = 0$. We choose the initial value $x(0) = 2$ that is far away from the optimal value $x^* = -4$. Let $\hat{\theta}(0) = 0$. We present simulations for various values of a, δ, ω to illustrate that the speed of convergence, the domain of attraction and the accuracy of the algorithm indeed depend on the parameters as suggested by Theorems 1 and 2.

We apply the first order scheme in Figure 1. By choosing $a = 0.3$, $\delta = 0.5(K = 4)$ and $\omega = 0.5$, the performance of the scheme is shown in Figure 2, where $|\mathbf{z}| = |(\tilde{x}, \tilde{\theta})|$ and $\tilde{x}$ and $\tilde{\theta}$ are the same as in Equation (10). It can be seen that, the state $\mathbf{z}$ converges to the neighborhood of the origin. The output also converges to the vicinity of the extremum value. We increase a such that $a = 0.6$ while keeping $\delta = 0.5$ and $\omega = 0.5$. Increasing a will get a fast convergent speed, while the domain of the attraction would be smaller. It can be seen

clearly from Figure 2 that, though both $y(t)$ and $|\mathbf{z}|$ converge very fast, they converge to a much larger *neighborhood of their optimal values.*

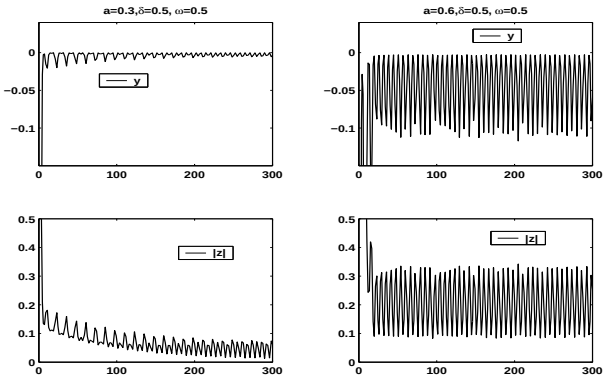

Fig. 2. The performance of the simplest extremum seeking scheme

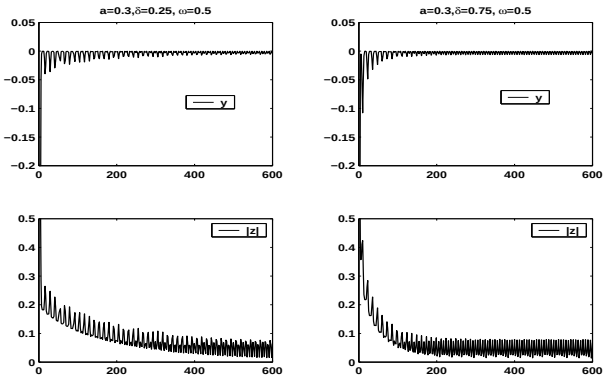

Fig. 3. The performance of the simplest extremum seeking scheme

Next, we fix $a = 0.3$ and $\omega = 0.5$. First, we let $\delta = 0.25$ and observe that the state $\mathbf{z}$ converges to the neighborhood of the origin, see Figure 3. The output also converges to the vicinity of its extremum value. Then, we increase δ to be 0.75 and observe that the convergence speed has increased, see Figure 3. However, when we further increase δ to be larger than 1.40 unstable performance is observed.

IV. Global extremum seeking in presence of local extrema

In this section, we summarize the results from [18] where global extremum seeking was investigated for objective functions that have local extrema. In particular, we use the following assumption:

Assumption 6 *There exists a unique global maximum $\theta^* \in \mathbb{R}$ of $Q(\cdot)$ such that*

$$Q(\theta^*) > Q(\theta), \quad \forall \theta \in \mathbb{R},\ \theta \neq \theta^*. \tag{24}$$

Assumption 6 implies that besides the global maximum θ^* there may also exist local maxima $\tilde{\theta}^*$. It is weaker than Assumption 3 that does not allow for local maxima.

The extremum seeking feedback scheme in Figure 4 was proposed in [18] to deal with this problem. One of the main differences between this scheme and the one in the previous section is that the amplitude of the excitation signal in Figure 4 is time varying, whereas in Figure 1 the amplitude is fixed. The model of the system in Figure 4 can be written as

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}, \alpha(\mathbf{x}, \hat{\theta} + a \cdot \sin(\omega \cdot t))) \\ \dot{\hat{\theta}} &= \omega \cdot \delta \cdot h(\mathbf{x}) \cdot \sin(\omega \cdot t) \\ \dot{a} &= -\omega \cdot \delta \cdot \varepsilon \cdot g(a), \qquad a(0) = a_0\ , \end{aligned} \tag{25}$$

where $g(\cdot)$ is a locally Lipschitz function that is zero at zero and positive otherwise. The simplest choice is $g(a) = a$. Here ε, ω, δ and a_0 are to be chosen by the designer.

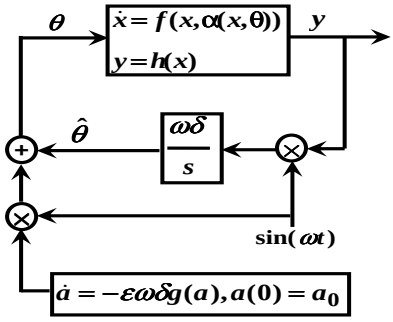

Fig. 4. A global extremum seeking feedback scheme.

Denoting $\sigma = \omega \cdot t$, the equations (25) in time "σ" are:

$$\begin{aligned} \omega \cdot \tfrac{d\mathbf{x}}{d\sigma} &= \mathbf{f}(\mathbf{x}, \alpha(\mathbf{x}, \hat{\theta} + a \sin(\sigma))) \\ \tfrac{d\hat{\theta}}{d\sigma} &= \delta \cdot h(\mathbf{x}) \cdot \sin(\sigma) \\ \tfrac{da}{d\sigma} &= -\varepsilon \cdot \delta \cdot g(a), \qquad a(0) = a_0. \end{aligned} \tag{26}$$

The system (26) is in standard singular perturbation form, where the singular perturbation parameter is ω. To obtain the fast and slow systems, we set $\omega = 0$ and "freeze" $\mathbf{x}$ at its "equilibrium", $\tilde{\mathbf{x}} = \mathbf{l}(\hat{\theta} + a \cdot \sin(\sigma))$ to obtain the "reduced" system in the variables (θ, a) in the time scale $\sigma = \omega \cdot t$:

$$\tfrac{d\theta}{d\sigma} = \delta \cdot Q(\theta + a \cdot \sin(\sigma)) \cdot \sin(\sigma) = \delta \cdot \mu(\sigma, \theta, a) \tag{27}$$

$$\tfrac{da}{d\sigma} = -\varepsilon \cdot \delta \cdot g(a), \qquad a(0) = a_0, \tag{28}$$

where $Q(\cdot) = h \circ \mathbf{l}(\cdot)$ is the output equilibrium map (cf. Assumption 6). We can write the "averaged" system of (27) by using:

$$\mu_{av}(\theta, a) := \frac{1}{2\pi} \int_0^{2\pi} \mu(t, \theta, a) dt, \tag{29}$$

where $\mu(\cdot, \cdot, \cdot)$ comes from (27). Indeed, using the above definition, we can analyze the closed loop system ((27)-(28)) via the auxiliary "averaged" system:

$$\tfrac{d\theta}{d\sigma} = \delta \cdot \mu_{av}(\theta, a) \tag{30}$$

$$\tfrac{da}{d\sigma} = -\delta \cdot \varepsilon \cdot g(a), \qquad a(0) = a_0 > 0\ . \tag{31}$$

By introducing the new time $\tau := \varepsilon \cdot \delta \cdot \sigma$, we can rewrite the above equations as follows:

$$\begin{aligned} \varepsilon \cdot \tfrac{d\theta}{d\tau} &= \mu_{av}(\theta, a) \\ \tfrac{da}{d\tau} &= -g(a), \qquad a(0) = a_0 > 0 \ , \end{aligned} \tag{32}$$

that are in standard singular perturbation form. However, there are several differences with the classical singular perturbation literature. In our case the equation:

$$0 = \mu_{av}(\theta, a) \tag{33}$$

may not have k *isolated* real roots $\theta = \ell_i(a)$. Indeed, some of the real roots may only be defined for $a \in [0, \bar{a}]$ and such that for some i and j we have $\ell_i(a) \neq \ell_j(a), a \in [0, \bar{a})$ and $\ell_i(\bar{a}) = \ell_j(\bar{a})$. Moreover, it shown in [18] that there exists a continuous function $p(\theta, a)$ such that:

$$\mu_{av}(\theta, a) = a \cdot p(\theta, a) \tag{34}$$

and this means that we will be unable to prove stability of the "boundary layer" system *uniform in* a that is a standard assumption in the singular perturbation literature. Note also that we are interested in convergence properties of this system initialized from a set of initial conditions satisfying $a(0) = a_0$ which is a weaker property from the standard stability properties considered in the singular perturbation literature.

Another assumption that characterizes solutions of the equation (33) is needed in the main result.

Assumption 7 *There exists an isolated real root* $\theta = \ell(a) : \mathbb{R}_{\geq 0} \to \mathbb{R}$ *of the equation (33) such that:*

1) ℓ *is continuous and* $D_1 p(\ell(a), a) < 0$, $\forall a \geq 0$, *where* $p(\theta, a)$ *is defined in (34).*
2) *There exists* $a^* > 0$ *such that for all* $a \geq a^*$, $\theta = \ell(a)$ *is the unique real root of (33).*
3) $\ell(0) = \theta^*$, *where* θ^* *is the global extremum, which comes from Assumption 6.*

Before showing the stability properties of the closed loop system (25), the following proposition shows stability properties of the "reduced" system (27)-(28).

Proposition 3 *Suppose that Assumptions 6 and 7 hold. Then, for any strictly positive* (Δ, ν) *and* $a_0 > a^*$ *there exist* $\beta = \beta_{a_0,\Delta,\nu} \in \mathcal{KL}$ *and* $\varepsilon^* = \varepsilon^*(a_0, \Delta, \nu) > 0$ *and for any* $\varepsilon \in (0, \varepsilon^*)$ *there exists* $\delta^* = \delta^*(\varepsilon) > 0$ *such that for any such* a_0, ε *and* $\delta \in (0, \delta^*)$ *we have that for all* $(\theta(\sigma_0), a(\sigma_0))$ *satisfying* $a(\sigma_0) = a_0$ *and* $|\theta(\sigma_0) - \ell(a_0)| \leq \Delta$ *and all* $t \geq t_0 \geq 0$ *the solutions of the system (27), (28) satisfy:*

$$|\theta(\sigma) - \ell(a(\sigma))| \leq \beta(|\theta(\sigma_0) - \ell(a(\sigma_0))|, \delta(\sigma - \sigma_0)) + \nu. \tag{35}$$

From the semi-global practical asymptotical (SPA) stability properties of the "reduced" system ((27)- (28)), the stability properties of the overall system (25) are stated in the following theorem.

Theorem 3 *Suppose that Assumptions 1, 2, 6 and 7 hold. Then, for any strictly positive* (Δ, ν) *and* $a_0 > a^*$ *there exist* $\mathcal{KL}$ *functions* $\beta_{\mathbf{x}}$, β_θ *and* $\varepsilon^* = \varepsilon^*(a_0, \Delta, \nu) > 0$ *and for any* $\varepsilon \in (0, \varepsilon^*)$ *there exists* $\delta^* = \delta^*(\varepsilon) > 0$, *for any* $\delta \in (0, \delta^*(\varepsilon))$ *there exists* $\omega^* = \omega^*(\delta) > 0$ *and for any such* a_0, ε, $\delta \in (0, \delta^*)$ *and* $\omega \in (0, \omega^*)$, *we have that for all* $(x(t_0), \hat{\theta}(t_0), a(t_0))$ *satisfying* $a(t_0) = a_0$, $|\hat{\theta}(t_0) - \ell(a_0)| \leq \Delta$, $|\mathbf{x}(t_0) - \mathbf{l}(\hat{\theta}(t_0))| \leq \Delta$ *and all* $t \geq t_0 \geq 0$ *the solutions of the system (25) satisfy:*

$$\begin{aligned} |\mathbf{x}(t) - \mathbf{l}(\hat{\theta}(t))| &\leq \beta_{\mathbf{x}}\left(|\mathbf{x}(t_0) - \mathbf{l}(\hat{\theta}(t_0))|, (t - t_0)\right) + \nu, \\ |\hat{\theta}(t) - \ell(a(t))| &\leq \beta_\theta\left(|\hat{\theta}(t_0) - \ell(a(t_0))|, \omega\delta(t - t_0)\right) + \nu. \end{aligned}$$

□

Theorem 3 presents a tuning mechanism for the controller parameters (choice of $\omega, \delta, \varepsilon$) and its initialization (choice of a_0) that guarantees semi-global practical convergence to the global extremum despite the presence of local extrema. Simulations in [18] illustrate that such convergence is indeed achieved. We note that since the static mapping $Q(\cdot)$ is not known, it is in general not possible to check *a priori* whether Assumption 7 holds, let alone analytically compute the values of a^*, ε^*, δ^* and ω^*. However, our result suggests that if there is some evidence that Assumptions 6 and 7 may hold, then increasing sufficiently a_0 and reducing sufficiently ε, δ and ω will indeed result in global convergence. In practice, determining how large $a_0, \varepsilon, \delta, \omega$ should be, may have to be determined through experiments.

The stability result of Theorem 3 is different from the stability result in [17, Theorem 1], where the closed-loop system is proved to be SPA stable, uniformly in the tuning parameters: the amplitude of the sine wave dither a and δ. First of all, in Theorem 3, the initial value of the amplitude of the dither plays an important role in the proposed ES feedback scheme to guarantee that the output of the system (25) converges to the global extremum semi-globally practically. Such an initial value a_0 is also a tuning parameter in the proposed ES feedback scheme. Secondly, a_0 does affect the convergence speed as β_θ and $\beta_{\mathbf{x}}$ are dependent on the choice of a_0, though analyzing how a_0 affects the convergence speed is much more difficult than other parameters. Thirdly, Theorem 3 clearly indicates that the choice of ε^* depends on the choice of a_0 and the choice of δ^* depends on the choice of ε and the choice of ω^* depends on the choice of δ.

A consequence of Theorem 3 is that we can tune the extremum seeking controller to achieve $\limsup_{\sigma\to\infty} |\theta(\sigma) - \ell(a(\sigma))| \leq \nu$ from an arbitrarily large set of initial conditions and for arbitrarily small $\nu > 0$. Moreover, from (28) it is obvious that there exists $\beta_a \in \mathcal{KL}$ with $\beta_a(s, 0) = s$, such that for all $a(\sigma_0) = a_0 \in \mathbb{R}_{>0}$ we have:

$$|a(\sigma)| \leq \beta_a(a(\sigma_0), \varepsilon \cdot \delta \cdot (\sigma - \sigma_0)), \ \forall \sigma \geq \sigma_0 \geq 0 \ , \tag{36}$$

and since $\ell(\cdot)$ is continuous and $\ell(0) = \theta^*$, it follows that

$$\lim_{\sigma\to\infty} \ell(a(\sigma)) = \theta^* \Rightarrow \limsup_{\sigma\to\infty} |\theta(\sigma) - \theta^*| \leq \nu \ ,$$

which implies semi-global practical extremum seeking since θ^* is the global extremum of $Q(\cdot)$. In the time "t", (36) becomes

$$|a(t)| \leq \beta_a(a(t_0), \omega \cdot \varepsilon \cdot \delta \cdot (t - t_0)), \ \forall t \geq t_0 \geq 0. \quad (37)$$

Note that since $Q(\cdot)$ is continuous, then for any $\nu > 0$, there exists $\nu_1 > 0$ such that

$$|\hat{\theta}| \leq \nu_1 \implies |Q(\hat{\theta}(t)) - \theta^*| = |y(t) - y^*| \leq \nu \ . \quad (38)$$

Theorem 3 can be interpreted as follows. For any (a_0, Δ, ν_1), where $a_0 > a^*$, ν_1 is defined in (38), we can adjust ε, δ and ω appropriately so that for all $|\mathbf{z}(t_0)| \leq \Delta$, where $\mathbf{z} \triangleq \begin{bmatrix} \mathbf{x} - \mathbf{l}(\hat{\theta}) \\ \hat{\theta} - \ell(a) \end{bmatrix}$, we have that $\limsup_{t \to \infty} |y(t) - y^*| \leq \nu$. In other words, the output of the system can be regulated arbitrarily close to the global extremum value y^* from an arbitrarily large set of initial conditions by adjusting the parameters $(\omega, \delta, \varepsilon, a_0)$ in the controller. This is despite the possible presence of local extrema (cf. Assumption 6).

The system model (26) suggests a three-time-scale dynamics when ω, δ and ε are very small. Indeed, the solutions first converge to a small neighborhood of the set $\mathcal{X} := \{(\mathbf{x}, \hat{\theta}) : \mathbf{x} - \mathbf{l}(\hat{\theta}) = 0\}$ (fast transient) and then with the speed proportional to $\omega \cdot \delta$ to a neighborhood of the set $\mathcal{L} := \{(\hat{\theta}, a) : \hat{\theta} - \ell(a) = 0\}$ (medium transient) and then with the speed propositional to $\omega \cdot \delta \cdot \varepsilon$ to a neighborhood of the point $(\theta, a) = (\theta^*, 0)$ (slow transient). Moreover, during the slow transient, the solutions stay in a ν-neighborhood of the set $\mathcal{X}$.

Example 2 *Assumption 7 is crucial to prove the global convergence of the proposed scheme. From the result of Theorem 3, $\hat{\theta}$ converges to $\ell(0)$. If $\ell(0)$ is not the global extremum, the output of the overall system would converge to a local extremum. An illustrative example is used to show that global extremum seeking can be achieved by the proposed scheme in Figure 4 in the presence of local extrema. Consider the following dynamic system*

$$\dot{x}_1 = -x_1 + x_2, \quad \dot{x}_2 = x_2 + u, \quad y = h(\mathbf{x}), \quad (39)$$

where $h(\mathbf{x}) = -(x_1 + 3x_2)^4 + \frac{8}{15}(x_1 + 3x_2)^3 + \frac{5}{6}(x_1 + 3x_2)^2 + 10$. The control input is chosen as

$$u = -x_1 - 4x_2 + \theta \ . \quad (40)$$

Moreover, we have $Q(\theta) = -\theta^4 + \frac{8}{15}\theta^3 + \frac{5}{6}\theta^2 + 10$ that has a global maximum at $\theta^* = 1$ and a local maximum at $\tilde{\theta}^* = -0.6$ as seen in Figure 5. Hence, Assumption 6 holds. The bifurcation diagram in Figure 5 implies that Assumption 7 also holds. First, we use the extremum seeking scheme from Section III, in which the amplitude of the sinusoidal signal is fixed to be small $a = 0.1$. The initial condition is chosen to be $\hat{\theta}(0) = -1$ such that the local maximum $\hat{\theta} = -0.6$ lays between the initial condition and the global maximum. By choosing $\varepsilon = 1$, $\delta = 0.005$, $\omega = 0.1$, $x_1(0) = x_2(0) = 0$, the simulations show that the extremum seeking scheme is stuck in the local maximum $\tilde{\theta}^* = -0.6$, see Figure 6.

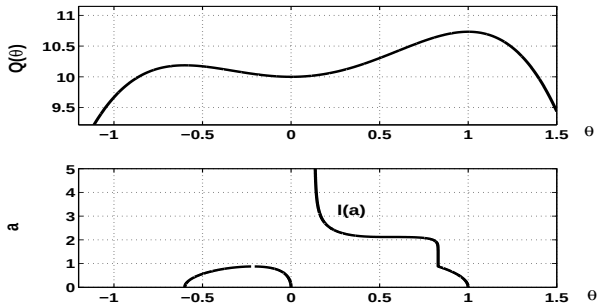

Fig. 5. A 4^{th}-order polynomial and its bifurcation diagram for which Assumption 7 holds.

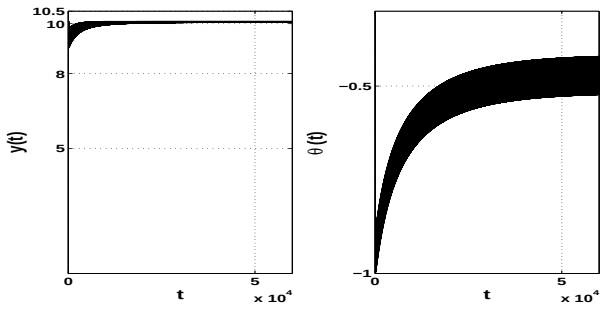

Fig. 6. The performance of the extremum seeking feedback scheme when $a = 0.1$ is fixed.

All conditions in Theorem 3 hold and hence the conclusion of Theorem 3 holds. Let $g(a) = a$, $a_0 = 3$, which turns out to be sufficiently large (see Figure 5), using the same parameters as above, $\hat{\theta}(t)$ converges to a small neighborhood of the global maximum $\theta^* = 1$ (the global extremum) as seen in Figure 7. This example shows that the proposed scheme can ensure that global extremum seeking is achieved despite the existence of local extrema.

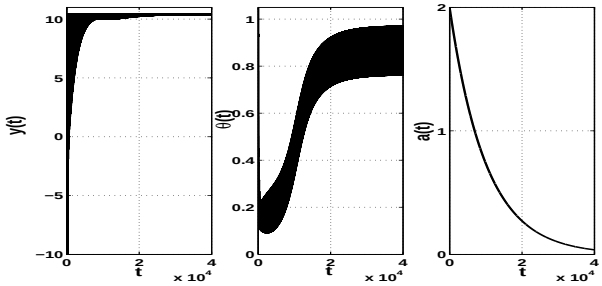

Fig. 7. The performance of the proposed ES feedback scheme.

V. Dither shape effects

In this section, we consider a static mapping $h(\cdot)$ with the first order extremum seeking controller from Section III, see Figure 8. Our goal is to consider the effects of the shape of the dither signal $d(t)$ on convergence properties of the algorithm. We present the main result in [19] that describes in detail how different dithers affect the domain of

attraction and speed of convergence, as well as the accuracy of extremum seeking control. It is shown that the square wave produces the fastest convergence among all signals with the same amplitude and frequency, if the amplitude a and the parameter δ in the controller are sufficiently small. Moreover, it is shown that in the limit as the amplitude is reduced to zero, all dithers yield almost the same domain of attraction and accuracy.

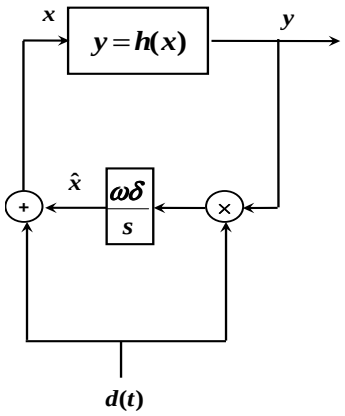

Fig. 8. A peak seeking feedback scheme with arbitrary dither.

The model of the closed loop system in Figure 8 is:

$$\dot{x} = \delta \cdot \omega \cdot h(x + d(t)) \cdot d(t), \tag{41}$$

where $h : R \to R$ is sufficiently smooth. The signal $d(\cdot)$ is referred to as "dither" and $\delta > 0$ and $\omega > 0$ are parameters that the designer can choose. We use the following assumptions:

Assumption 8 *: There exists a maximum x^* of $h(\cdot)$ such that*

$$Dh(x^*) = 0; \qquad D^2h(x^*) < 0 \ . \tag{42}$$

Assumption 9 *Dither signals $d(\cdot)$ are periodic functions of period $T > 0$ (and frequency $\omega = \frac{2\pi}{T}$) that satisfy:*

$$\int_0^T d(s)ds = 0; \ \frac{1}{T}\int_0^T d^2(s)ds > 0; \ \max_{s\in[0,T]} |d(s)| = a;$$

where $a > 0$ is the amplitude of the dither.

The parameter ω in (41) is chosen to be the same as the frequency of the dither signal.

For comparison purposes, three special kinds of dither are used repeatedly in our examples: sine wave, square wave and triangle wave. The sine wave is defined in the usual manner. The square wave and triangle wave of unit amplitude and period 2π are defined as follows:

$$sq(t) := \begin{cases} 1, \ t \in [2\pi k, \pi(2k+1)) \\ -1, \ t \in [\pi(2k+1), 2\pi(k+1)) \end{cases}$$

$$tri(t) := \begin{cases} \frac{2}{\pi}(t - 2\pi k), \ t \in [2\pi k, \pi(2k+\frac{1}{2})) \\ \frac{2}{\pi}(-t - \pi(2k-1)), \ t \in [\pi(2k+\frac{1}{2}), \pi(2k+\frac{3}{2})) \\ \frac{2}{\pi}(t - \pi(2k+2)), \ t \in [\pi(2k+\frac{3}{2}), 2\pi(k+1)) \end{cases}$$

Note that by definition, the signals $\sin(t)$, $sq(t)$ and $tri(t)$ are of unit amplitude and period 2π. We can generate similar signals of arbitrary amplitude a and frequency ω, e.g., $a \cdot sq(\omega t)$. We will often use the "power" (average of the square of the signal) of the normalized dithers (unit amplitude and period 2π):

$$P_{sq} = 1; \ P_{sin} = \frac{1}{2}; \ P_{tri} = \frac{1}{3} \tag{43}$$

In general, we use P_d to denote the power of dithers $d(\cdot)$ with amplitude equal to 1 and period 2π. The power for dither $d(\cdot)$ with amplitude $a \neq 1$ is equal to $a^2 P_d$. We emphasize that our results apply to arbitrary dithers satisfying Assumption 9.

Remark 1 *Assumption 9 is needed in our analysis that is based on averaging of (41). We note that most extremum seeking literature (see [2]) uses dither signals of the form $d(t) = a\sin(\omega t)$ which obviously satisfy our Assumption 9.*

Introducing the coordinate change $\tilde{x} = x - x^*$, we can rewrite (41) as follows:

$$\dot{\tilde{x}} \ = \ \delta\omega h(\tilde{x} + x^* + d(t)) \cdot d(t) =: \delta\omega \bar{f}(t, \tilde{x}, d). \tag{44}$$

It was shown in Section III that under a stronger version of Assumption 8 (uniqueness of the maximum) and with the sinusoidal dither $d(t) = a \cdot \sin(\omega t)$, where $\omega = 1$ (that satisfies Assumption 9) we have that for any compact set $\mathcal{D}$ and any $\nu > 0$ we can chose the amplitude of the dither $a > 0$ and $\delta > 0$ and find a class $\mathcal{KL}$ function β, which depends on δ and $d(\cdot)$, such that the solutions of the closed loop system (44) satisfy:

$$|\tilde{x}(t_0)| \in \mathcal{D} \implies |\tilde{x}(t)| \leq \beta\left(|\tilde{x}(t_0)|, t - t_0\right) + \nu, \tag{45}$$

for all $t \geq t_0 \geq 0$.

Note that $\mathcal{D}$, ν and β are *performance indicators* since they quantify different aspects of the performance of the extremum seeking algorithm. We will show later that each of these indicators is affected by our choice of dither $d(\cdot)$ and the parameter $\delta > 0$. In particular, we have that:

- **Speed of convergence** of the algorithm is captured by the function β. Obviously, we would like convergence to be as fast as possible.
- **Domain of convergence** is quantified by the set $\mathcal{D}$. In particular, we would like to make the domain of convergence (attraction) as large as possible.
- **Accuracy** of the algorithm is quantified by the number $\nu > 0$ since all trajectories starting in the set $\mathcal{D}$ eventually end up in the ball B_ν, where we have that $|x(t) - x^*| \leq \nu$. Indeed, the smaller the number ν, the closer we eventually converge to the maximum x^* (hence, the accuracy of the algorithm is better).

It turns out that a direct analysis of the system (44) to estimate $\mathcal{D}, \nu, \beta$ is hard but the system can be analyzed via

an appropriate auxiliary averaged system. We will carry out such an analysis in the next section.

Consider the following auxiliary gradient system:

$$\dot{\zeta} = Dh(\zeta + x^*) \ . \tag{46}$$

Because of Assumption 8, the system (46) has the property that x^* is an asymptotically stable equilibrium[2]. Let $\mathcal{D}$ denote the domain of attraction of x^* for the system (46) and note that since $h(\cdot)$ is assumed smooth, the set $\mathcal{D}$ is a neighborhood of x^*. In other words, a consequence of Assumption 8 is that there exists $\beta \in \mathcal{KL}$ and a set $\mathcal{D}$ such that for all $t \geq 0$ the solutions of (46) satisfy:

$$\zeta_0 \in \mathcal{D} \quad \Rightarrow \quad |\zeta(t)| \leq \beta(|\zeta_0|, t) \tag{47}$$

Using this auxiliary system, we can state our main result:

Theorem 4 *Suppose that Assumption 8 holds and consider the closed loop system (41) with an arbitrary dither $d(\cdot)$ for which Assumption 9 holds, where $a > 0$ is the dither amplitude. Let $\mathcal{D}$ and β come from (47). Then, for any strict compact subset $\hat{\mathcal{D}}$ of $\mathcal{D}$ and any $\nu > 0$, there exists $a^* > 0$ and $\delta^* > 0$ such that for any $a \in (0, a^*]$, $\delta \in (0, \delta^*]$ and any $\omega > 0$ we have that solutions of (41) satisfy:*

$$\tilde{x}_0 \in \hat{\mathcal{D}} \Rightarrow |\tilde{x}(t)| \leq \beta\left(|\tilde{x}_0|, \delta\omega a^2 P_d(t - t_0)\right) + \nu \tag{48}$$

A sketch of proof of Theorem 4 can be found in [19].

We emphasize that the auxiliary gradient system (46) plays a crucial role in terms of achievable performance of the extremum seeking controller. Indeed, $\mathcal{D}$ and β are *independent* of the choice of dither and Theorem 4 specifies how they affect the achievable domain of attraction $\hat{\mathcal{D}}$ of the closed loop (41), as well as the speed of convergence via the function β.

We now discuss Theorem 4 in more detail to explain how dither shape affects the domain of attraction, accuracy and convergence speed of the closed loop system. We note that the controller parameter (a, δ) needs to be tuned appropriately in order for Theorem 4 to hold.

Domain of attraction: It is shown that any dither satisfying Assumption 9 can yield a domain of attraction $\hat{\mathcal{D}}$ that is an arbitrary strict subset of the domain of attraction of the gradient system (46) if a and δ are sufficiently small. We emphasize that $\omega > 0$ can be arbitrary and a and δ do not depend on it.

Accuracy: The ultimate bound that is quantified by the number ν can be made arbitrarily small by any dither satisfying Assumption 9 if a and δ are sufficiently small. Hence, in the limit, all dithers perform equally well in terms of domain of attraction and accuracy.

Convergence speed: We emphasize that β in (48) is the same as β in (47) for any dither $d(\cdot)$. The main difference in speed of convergence comes from the scaling factor within the function β:

$$\delta \cdot \omega \cdot a^2 \cdot P_d \ , \tag{49}$$

where δ is a controller parameter, a and ω are respectively the amplitude and frequency of dither and P_d is the power of the normalized dither (with unit amplitude and period 2π). Note also that $\omega\delta$ is the integrator constant in Figure 8. Also, note that Theorem 4 holds for sufficiently small a and δ that are independent of ω which is an arbitrary positive number. Obviously, if the product (49) is larger than 1 then the closed loop system (41) converges faster than the auxiliary gradient system (46). Similarly, if the product (49) is smaller than 1, the system (41) converges slower than the gradient system (46).

The first observation is that for sufficiently small a and δ the bound in Theorem 4 holds *for any* ω. Hence, for fixed a, δ and P_d we have that the larger the ω, the faster the convergence. In other words, Theorem 4 shows that in our case study we can achieve arbitrarily fast convergence of the extremum seeking closed loop by making ω sufficiently large. Simulations in Example 3 verify our analysis. We also emphasize that this result is in general not possible to prove for general dynamic plants. For instance, the results in Sections III and IV that are stated for general dynamical systems provide a similar bound as in (48) under the stronger assumption that ω is sufficiently small.

Suppose now that a, δ and ω are fixed and we are only interested in how the shape of dither affects the convergence. As we change dither, its (normalized) power P_d changes and as we can see from (43) that the square wave will yield twice larger normalized power than the sine wave and three times larger power than the triangle wave. Consequently, we can expect twice faster convergence with the square wave than with the sine wave and three times faster convergence than with the triangle wave. Simulation results in Example 4 that we present in the sequel are consistent with the above analysis.

Note that Theorem 1 is a weaker version of Theorem 4 for the sine wave dither only. Indeed, Theorem 1 does not consider arbitrary dither and the domain of attraction and convergence estimates are not as sharp as in Theorem 4. For instance, the relationship of convergence rate and the domain of attraction to the auxiliary system (46) was not shown in Section III as this was impossible to do using the Lyapunov based proofs used to prove Theorem 1. On the other hand, using the trajectory based proofs adopted in this paper, we can prove tight estimates as outlined in Theorem 4. Moreover, in Theorem 1 it was not clear how the dither power P_d affects the convergence rate of the average system. Note that the function β in (48) is the same for any dither satisfying Assumption 9 and the only difference comes the parameters in (49). However, the values of a^* and δ^* are typically different for different dithers.

We note that one can state and prove a more general version of Theorem 4 that applies to general dynamical plants and in this case $h(\cdot)$ is an appropriate reference-to-output map. With extra assumptions on the plant dynamics, one can use singular perturbation theory to prove this more general result (see for instance [17] for a Lyapunov based proof in the case of sine wave dither). However, in this case we will

[2]Moreover, all local maxima of $h(\cdot)$ are asymptotically stable equilibria of (46) and all local minima of $h(\cdot)$ are unstable equilibria of (46).

need to require that ω is sufficiently small.

The following proposition is obvious and it states that the power of the normalized square wave is larger than or equal to the power of any other normalized dither satisfying Assumption 9. In other words, for fixed δ and a for which (48) holds, the square wave is guaranteed to produce the fastest convergence over all dithers with the same amplitude and frequency.

Proposition 4 *Consider arbitrary $d(\cdot)$ satisfying Assumption 9. Then, we have that the power of the normalized dither satisfies:*

$$0 < P_d \leq P_{sq} = 1 \ .$$

It has been shown in Theorem 4 that the convergence speed of the extremum seeking systems depends on the choice the dither shape P_d, amplitude a and frequency ω as well as δ. It also is shown that the domain of the attraction and accuracy of all dithers are almost the same as $\left\{ \begin{array}{l} a \to 0 \\ \delta \to 0 \end{array} \right.$. In this part, we use examples to illustrate various behaviors and simulations to confirm our theoretical findings. Our results should motivate the users of extremum seeking control to experiment with different dithers in order to achieve the desired trade-off between convergence, domain of attraction or accuracy.

The following example illustrates that increasing the frequency of dither while keeping a, δ and P_d the same yields faster convergence.

Example 3 *Consider the quadratic mapping*

$$h(x) = -(x+4)^2 \tag{50}$$

where $Dh(\tilde{x}+x^*) = -2\tilde{x}$. It is trivial to see that in this case $\mathcal{D} = R$ and $\beta(s,t) = se^{-2t}$ (see Theorem 4). The dither is chosen to be $d(t) = a\sin(\omega t)$. Hence, from (43) we have $P_{sin} = 1/2$. The initial condition is chosen as $x_0 = -2$. When we fix $a = 0.5$ and $\delta = 0.1$, the output response with different frequencies is shown in Figure 9. It is clear that the larger the ω is, the faster the convergence is.

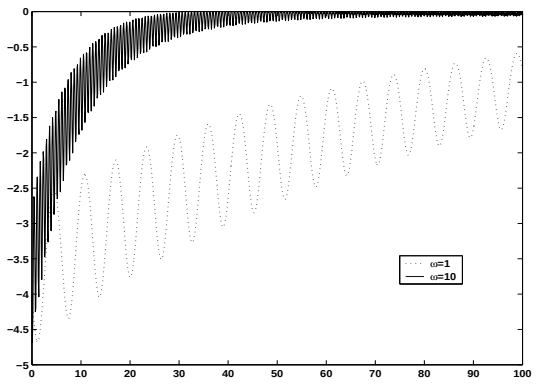

Fig. 9. The output of the ES with different frequencies

Next, we will discuss the situation when $h(\cdot)$ is a quadratic mapping, where we assume $\omega = 1$. Consider the simplest possible case of quadratic maps:

$$h(x) = -x^2 + a_1 x + a_0 \ ,$$

we have that

$$Dh(x) = -2x + a_1$$

and since $x^* = \frac{a_1}{2}$, we can write with $\tilde{x} := x - x^*$:

$$Dh(\tilde{x} + x^*) = -2\tilde{x} \ .$$

Hence, the auxiliary gradient system (46) takes the following form:

$$\dot{\zeta} = -2\zeta \ .$$

It is not hard to show that in this case for arbitrary dither $d(\cdot)$ satisfying Assumption 9 we have that the average system is of the form:

$$\dot{\tilde{x}} = -2a^2 P_d \delta \tilde{x} \ . \tag{51}$$

Hence, in this special case we have that the average system for any dither satisfying Assumption 9 is globally exponentially stable. Indeed, for square wave, sine wave and triangular wave we have from (43) that the following holds for all $x_0 \in R$, $t \geq 0$, respectively:

$$\begin{aligned} sq: \ |\tilde{x}(t)| &= \exp\left(-2a^2\delta t\right)|\tilde{x}_0| \\ sin: \ |\tilde{x}(t)| &= \exp\left(-a^2\delta t\right)|\tilde{x}_0| \\ tri: \ |\tilde{x}(t)| &= \exp\left(-\frac{2}{3}a^2\delta t\right)|\tilde{x}_0| \ . \end{aligned}$$

The square wave produces the fastest speed of convergence for the average system among all dithers with the same amplitude. The same can be concluded for the actual system using the proof of Theorem 4. This suggests that the square-wave dither should be the prime candidate to use in the ES system for fast convergence speed, although this dither is rarely considered in the literature [2]. Indeed, all references that we are aware of use a sinusoidal dither signal. The simulation results shown in Example 4 illustrates that the convergence speed of extremum seeking controller with the square wave is fastest among all dithers with the same amplitude.

Example 4 *The simulation is done for the following system where $\omega = 1$:*

$$\dot{x} = \delta h(x + d(t))d(t)$$

where $h(x) = -(x+4)^2$. In the new coordinate, $\tilde{x} = x - x^* = x + 4$, we have

$$\dot{\tilde{x}} = \delta h(\tilde{x} + x^* + d(t))d(t) = -\delta(\tilde{x} + d(t))^2 d(t).$$

The averaged system is given in (51). The simulation result is shown in Figure 10, where $a = 0.1$ and $\delta = 0.5$. Simulations show that the extremum seeking controller with the square wave dither converges fastest.

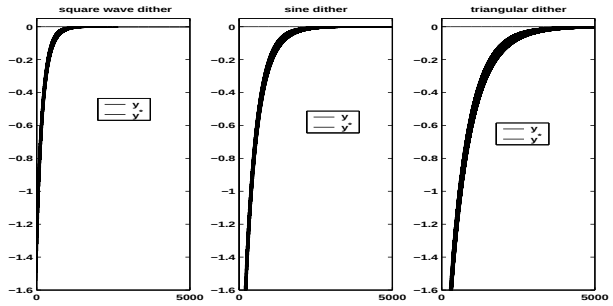

Fig. 10. The performance of the extremum seeking schemes of different excitation signal

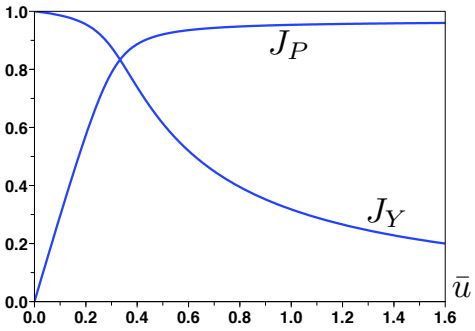

Fig. 11. Productivity and yield for system (53) with $c = 3, K_m = 0.1$.

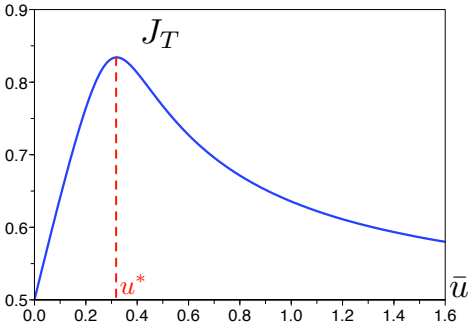

Fig. 12. Overall performance index J_T for system (53) with $c = 3$, $K_m = 0.1$ and $\lambda = 0.5$.

VI. Extremum seeking of bioprocesses

In this section we consider the steady-states of bioprocesses that need to achieve an optimal trade-off between yield and productivity maximization, see [4]. To illustrate the ideas, consider a single irreversible enzymatic reaction of the form $\mathbf{X}_1 \longrightarrow \mathbf{X}_2$ with $\mathbf{X}_1$ the substrate (or reactant) and $\mathbf{X}_2$ the product. The reaction takes place in the liquid phase in a continuous stirred tank reactor. The substrate is fed into the reactor with a constant concentration c at a volumetric flow rate u. The reaction medium is withdrawn at the same volumetric flow rate u so that the liquid volume V is kept constant. The process dynamics are described by the following standard mass-balance state space model:

$$\dot{x}_1 = -r(x_1) + (u/V)(c - x_1) \tag{52a}$$
$$\dot{x}_2 = r(x_1) - (u/V)x_2 \tag{52b}$$

where x_1 is the substrate concentration, x_2 is the product concentration and $r(x_1)$ is the reaction rate (called *kinetics*). Obviously this system makes physical sense only in the non-negative orthant $x_1 \geqslant 0$, $x_2 \geqslant 0$. Moreover the flow rate u (which is the control input) is non-negative by definition and physically upper-bounded (by the feeding pump capacity) $0 \leqslant u \leqslant u^{max}$. In [4] we investigated two different cases depending on the form of the rate function $r(x_1)$. The first one is the Michaelis-Menten kinetics which is the most basic model for enzymatic reactions $r(x_1) = \frac{v_m x_1}{K_m + x_1}$ with v_m the maximal reaction rate and K_m the half-saturation constant. To normalise the model we use $v_m V$ and v_m^{-1} as the units of u and time respectively. We also assume that the process is equipped with an on-line sensor that measures the product concentration x_2 in the outflow. So the normalised model becomes

$$\dot{x}_1 = -\frac{x_1}{K_m + x_1} + u(c - x_1) \tag{53a}$$
$$\dot{x}_2 = \frac{x_1}{K_m + x_1} - ux_2. \tag{53b}$$

It can be readily verified that, for any positive constant input flow rate $\bar{u} \in (0, u^{max}]$, there is a unique steady-state $\bar{x}_1 = \varphi_1(\bar{u})$, $\bar{x}_2 = \varphi_2(\bar{u})$ that is globally asymptotically stable in the non-negative orthant.

The industrial objective of the process is the production of the reaction product. For process optimization, two steady-state performance criteria are considered : the *productivity* J_P is the amount of the product harvested in the outflow per unit of time $J_P := \bar{u}\bar{x}_2 = \bar{u}\varphi_2(\bar{u})$; the *yield* is the amount of product made per unit of substrate fed to the reactor: $J_Y := \frac{\bar{x}_2}{c} = \frac{\varphi_2(\bar{u})}{c}$. The sensitivity of J_P and J_Y with respect to $\bar{u}$ is illustrated in Figure 11. The process must be operated at a steady-state that achieves a trade-off between yield and productivity. To this end we define an overall performance index as a convex combination of J_P and J_Y:

$$J_T(\bar{u}) \triangleq \lambda J_P + (1 - \lambda)J_Y = \varphi_2(\bar{u})\left[\lambda\bar{u} + \frac{1 - \lambda}{c}\right] \tag{54}$$

for $\lambda \in [0, 1]$. This cost function is illustrated in Figure 12 where it is readily seen that it has a unique global maximum u^*. The corresponding optimal steady-state is naturally defined as $x_1^* = \varphi_1(u^*)$, $x_2^* = \varphi_2(u^*)$.

The kinetic rate function $r(x_1)$, the dynamical model (52) and the function $J_T(\bar{u})$ are unknown to the user while the goal is to maximize the composite cost $J_T = \lambda u x_2 + (1 - \lambda)c^{-1}x_2$ but he does not know that J_T is a function of $\bar{u}$ of the form (54) shown in Figure 12. We apply the extremum seeking scheme from Section III with the following definition

of the output and input:

$$y = h(x) := \lambda u x_2 + (1-\lambda)c^{-1}x_2;\ u := \alpha(\hat{\theta} + a\sin(\omega t)) \tag{55}$$

where x_2 is measured, λ chosen by the designer and $\alpha(\cdot)$ is a sigmoid function (see [4]). In Figure 13 the operation of the extremum seeking control algorithm is illustrated for appropriately tuned parameters $a = 0.02$, $K\delta = 1$, $\omega = 0.1$. We see that there is a time scale separation between the system itself and the climbing mechanism se predicted by Theorem 1.

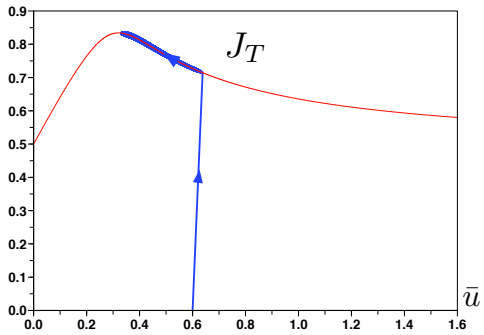

Fig. 13. Extremum seeking for system (53) with $a = 0.02$, $k = 1$, $\omega = 0.1$.

The following corollary holds for all bioprocesses that satisfy the following assumption:

Assumption 10 *The following conditions hold: (i) For each admissible value of the flow rate $\bar{u}$ the system must have a single globally asymptotically stable equilibrium; (ii) The performance cost function must be single-valued and "well-shaped" in the sense that, for the admissible range of flow rate values $0 \leqslant \bar{u} \leqslant u^{max}$, it must have a single maximum value $J_T(u^*)$ without any other local extrema.*

A direct consequence of Theorem 1 is the following:

Corollary 1 *For any initial condition $(x_1(0) \geqslant 0, x_2(0) \geqslant 0, \theta_0(0))$ and for any $\nu > 0$, there exist parameters (a, k, ω) such that, for the closed-loop system* (53), (55), (7), *$x_1(t) \geqslant 0$, $x_2(t) \geqslant 0$, $\theta_0(t)$ bounded and*

$$\limsup_{t\to\infty}\Big(|x_1(t) - x_1^*| + |x_2(t) - x_2^*| + |u(t) - u^*|\Big) \leqslant \nu.$$

Note that the above corollary can be restated in the spirit of Theorem 1 to guarantee semi-global practical convergence. This is a stronger result than the main result in [22] that deals only with local convergence.

It was shown in [4] that some bioprocesses may not satisfy Assumption 10 and, in particular, J_T may turn out to be multivalued. Such situation is quite natural in the context of bioreactors and yet it is not covered with any of the presented results. A preliminary analysis of this situation was given in [4] and we summarize below.

We consider again the simple model (52) but we now assume that, in addition to the Michaelis-Menten kinetics, the reaction rate is subject to exponential substrate inhibition. The rate function is as follows:

$$r(x_1) = \frac{v_m x_1}{K_m + x_1} e^{-bx_1^p}$$

where b and p are two positive constant parameters. The dynamical model is written:

$$\dot{x}_1 = -\frac{v_m x_1}{K_m + x_1} e^{-bx_1^p} + u(c - x_1) \tag{56a}$$

$$\dot{x}_2 = \frac{v_m x_1}{K_m + x_1} e^{-bx_1^p} - ux_2 \tag{56b}$$

Depending on the value of $\bar{u} \in (0, u^{max}]$, the system may have one, two or three steady-states $(\bar{x}_1, \bar{x}_2)$ with $\bar{x}_1$ solution of:

$$\frac{v_m \bar{x}_1}{K_m + \bar{x}_1} e^{-b\bar{x}_1^p} = \bar{u}(c - \bar{x}_1)$$

and $\bar{x}_2 = c - \bar{x}_1$.

The productvity $J_P = \bar{u}\bar{x}_2$ is represented in Figure 14 as a function of $\bar{u}$. In this example, J_P is clearly a multivalued function of $\bar{u}$. However it can be seen that it has a unique global maximum for $\bar{u} = u^*$. Moreover, the graph of Figure

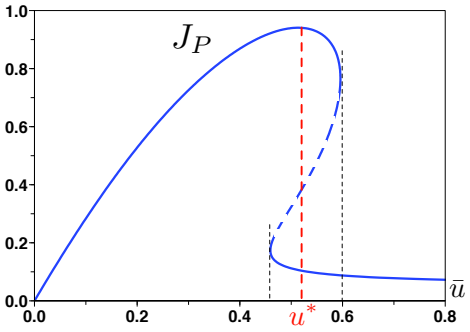

Fig. 14. Productivity J_P for system (56) with $c = 3$, $v_m = 2$, $K_m = 1$, $b = 0.08$, $p = 3.4$.

14 can also be regarded as a bifurcation diagram with respect to the parameter $\bar{u}$ where the solid branches correspond to stable equilibria and the dashed branch to unstable equilibria. Hence it can be seen that the maximum point is located on a stable branch.

Here we assume that the industrial objective is to achieve the maximization of the productivity J_P. A fully satisfactory operation of the extremum seeking control law (with $y(t) = u(t)x_2(t)$) can be observed in Figure 15 and Figure 16.

The result of Figure 15 is expected since we are in conditions quite similar to the previous case of Section IV. The result of Figure 16 is more informative since here the convergence towards the maximum of the cost function is operated in two successive stages. In a first stage, there is a fast convergence to the nearest stable state which is located *on the lower stable branch* followed by a quasi-steady-state progression along that branch. Then, when the

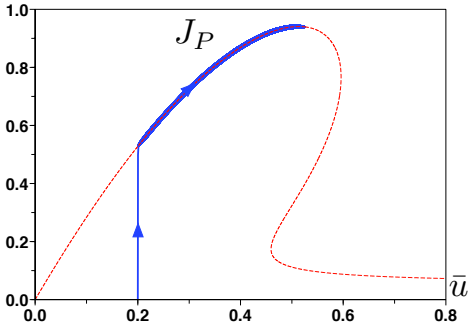

Fig. 15. Extremum seeking for system (56) with $a = 0.003$, $k = 10$, $\omega = 0.01$.

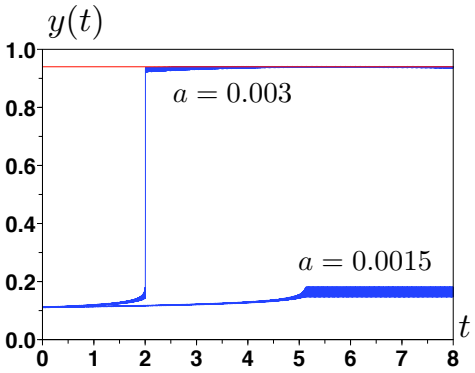

Fig. 17. Output signal $y(t)$: when a is too small, the trajectory is stuck on the lower branch.

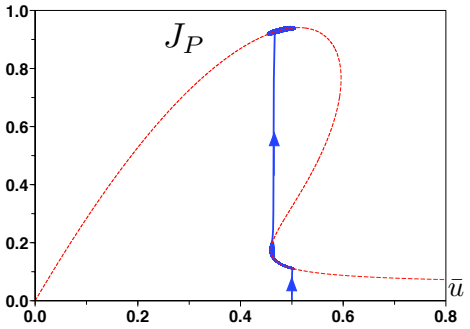

Fig. 16. Extremum seeking for system (56) with $a = 0.003$, $k = 6$, $\omega = 0.01$.

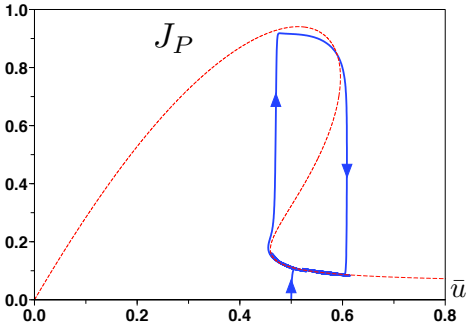

Fig. 18. Extremum seeking for system (56) with $a = 0.015$, $k = 6$, $\omega = 0.01$.

state reaches the bifurcation point, there is a fast jump up to the *good upper branch* and a final climbing up to the maximum point. It is very important to emphasize here that, in order to get the result of Figure 16, the amplitude a of the dither signal must be large enough. Otherwise, the trajectory of the closed loop system definitely remains stuck on the lower branch at the bifurcation point as shown in Figure17. On the other side, too large values of the dither amplitude are also prohibited because they produce cyclic trajectories as shown in Figure18. From all these observations, we can conclude that by tuning the amplitude of the dither signal properly, it is possible to pass through the discontinuities of the stable branches of the cost function and to converge to the global maximum. This further motivates tuning the amplitude of dither in the extremum seeking controller. While a preliminary analysis of this issue was presented in [4], a careful analysis and tuning guidelines in this case remain an open research problem.

VII. Conclusions

A summary our recent results on dynamical properties of a class of adaptive extremum seeking controllers was presented and applied to a various models of a continuously stirred reactor. Our analysis shows how various tuning parameters in the extremum controller affect the overall convergence properties of the algorithm. Such results will be useful to practitioners since they provide controller tuning guidelines that can ensure larger domains of attraction, faster convergence or better accuracy of the extremum seeking algorithms. Moreover, it was shown that global extremum seeking in presence of local extrema can be achieved using appropriate tuning of controller parameters.

References

[1] K. J. Åström and B. Wittenmark, *Adaptive control (2nd ed.)* Reading, MA: Addison-Wesley, 1995.

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

[3] A. Banaszuk, K.B. Ariyur, M. Krstić, C.A. Jacobson, "An adaptive algorithm for control of combustion instability", *Automatica*, vol. 40 (2004), 1965-1972.

[4] G. Bastin, D. Nešić, Y. Tan and I.M.Y Mareels, "On extremum seeking in bioprocesses with multivalued cost functions", Biotechnology Progress, in print, 2009.

[5] E. Biyik and M. Arcak, "Gradient climbing in formation via extremum seeking and passivity-based coordination rules", *Proc. 46th IEEE Conf. Decis. Contr.*, New Orleans, USA, 2007, pp. 3133–3138.

[6] P. F. Blackman, "Extremum-seeking regulators", In J. H. Westcott, *An exposition of adaptive control* New York: The Macmillan Company, 1962.

[7] P.M. Dower, P. Farrell and D. Nešić, “Extremum seeking control of cascaded Raman optical amplifiers”, IEEE Transactions on Control Technology, vol. 16 (2008), pp. 396–407.
[8] S. Drakunov, Ü. Özgüner, P. Dix and B. Ashra, “ABS control using optimum search via sliding modes”, *IEEE Transactions on Control Systems Technology*, vol. 3 (1995), 79–85.
[9] M. Guay, D. Dochain, M. Perrier, “Adaptive extremum seekingcontrol of continuous stirred tank bioreactors with unknown growth kinetics”, *Automatica*, vol. 40 (2004), 881-888.
[10] M. Krstić and H.-H. Wang, “Stability of extremum seeking feedback for general nonlinear dynamic systems”, *Automatica*, vol. 36 (2000), 595–601.
[11] C.G. Mayhew, R.G. Sanfelice and A.R. Teel, “Robust source-seeking hybrid controllers for autonomous vehicles”, *Proc. Amer. Contr. Conf.*, New York, July 11-13, 2007, 1185–1190.
[12] P. Ögren, E. Fiorelli and N.E. Leonard, “Cooperative control of mobile sensor networks: adaptive gradient climbing in a distributed environment”, *IEEE Trans. Automat. Contr.*, vol. 49 (2004), 1292–1302.
[13] K. S. Peterson and A. G. Stefanopoulou, “Extremum seeking control for soft landing of an electromechanical valve actuator”, *Automatica*, vol. 40 (2004), 1063-1069.
[14] D. Popović, M. Janković, S. Manger and A.R.Teel, “Extremum seeking methods for optimization of variable cam timing engine operation”, *Proc. American Control Conference*, Denver, Colorado, June 4-6, 3136–3141, 2003.
[15] E. Sontag, “Comments on integral variants of ISS”, *Systems & Control Letters*, **34** pp. 93–100, 1998.
[16] J. Sternby, “Extremum control systems: An area for adaptive control?”, *Proc. Joint Amer. Contr. Conf.*, San Francisco, CA, (1980), WA2-A.
[17] Y. Tan, D. Nešić and I.M.Y. Mareels, “On non-local stability properties of extremum seeking control”, Automatica, vol. 42 (2006), pp. 889–903.
[18] Y. Tan, D. Nešić, I.M.Y. Mareels and A. Astolfi, “On global extremum seeking in the presence of local extrema”, Automatica 45 (2009), pp. 245–251.
[19] Y. Tan, D. Nešić and I.M.Y. Mareels, “On the choice of dither in extremum seeking systems: A case study”, Automatica, vol. 44 (2008), pp. 1446–1450.
[20] A.R. Teel and D. Popović, “Solving smooth and nonsmoooth multivariable extremum seeking problems by the methods of nonlinear programming”, *Proc. Amer. Contr. Conf.*, Arlington, VA, June 25-27, (2001), 2394-2399.
[21] H.-H. Wang, S. Yeung, and M. Krstić, “Experimental Application of Extremum Seeking on an Axial-Flow Compressor”, *IEEE Transactions on Control Systems Technology*, **8**, pp. 300–309, 2000.
[22] H-H. Wang, M. Krstić and G. Bastin, “Optimizing bioreactors by extremum seeking”, *International Journal of Adaptive Control and Signal Processing*, Vol. 13, pp. 651-669, 1999.
[23] D. J. Wilde, *Optimum Seeking Methods*. Englewood NJ: Prentice-Hall Inc., 1964.