2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Interconnection Conditions for the Stability of Nonlinear Sampled-Data Extremum Seeking Schemes

Karla Kvaternik and Lacra Pavel

Abstract— **The application of numerical optimization methods to the problem of extremum seeking control (ESC) has the potential to greatly diversify the types and capabilities of ESC schemes. The first uniform treatment of such sampled-data ESC schemes was given in [1]. We approach the problem from the point of view of interconnected systems' theory, deriving a different, more structurally concrete set of conditions that guarantee the closed-loop stability of such schemes. Our main assumptions concern the interconnection terms arising from the dynamic coupling between a numerical optimization algorithm and a continuous-time nonlinear plant. We demonstrate how these assumptions are satisfied for a special case involving an approximate gradient descent. Our primary motivation in deriving these new conditions is their natural suitability for the development and analysis of decentralized ESC schemes.**

I. INTRODUCTION

With the appearance of its first rigorous stability analysis in [2], extremum seeking control (ESC) has received renewed interest from the research community within recent years. The interest in ESC continues to be incited by the proliferation of its practical applications [3], [4], [5], [6].

ESC is a method of regulating a plant's output to a value that corresponds to an optimum of its steady-state *reference-to-output* (RO) map. The analytic structure of this RO map is assumed to be unknown, and therefore the reference input that produces the optimum output value cannot be computed offline.

Most of the literature on ESC is focused on the analysis of continuous-time schemes. The authors in [7] extend the work in [2] to demonstrate semiglobal practical stability of the original scheme and several variations thereof. In [8], a different flavor of ESC is introduced, where the adjustable parameters are unknown but the structure of the RO map is partially known. Another approach to ESC is given in [9], where finite-time regulation is combined with numerical optimization to minimize an unknown state-to-output map. An ESC scheme for discrete-time plants with a special structure is analyzed in [10].

In this paper, we focus on sampled-data ESC of the sort proposed in [1]. Our interest is partly motivated by the prospect of interchangeably choosing an optimizer from amongst a variety of well-developed numerical optimization methods in the nonlinear optimization literature. In [1], the authors provide a set of conditions guaranteeing the closed-loop stability of a very general sampled-data ESC scheme, and their work appears to be the first attempt at providing a unifying treatment of the subject.

As in [1], we aim to derive a set of sufficient (small-gain type) conditions for the closed-loop stability of generic sampled-data ESC schemes. However, our approach differs twofold. First, we view the problem from the perspective of large-scale, interconnected systems theory (c.f. [11], for example). Second, we take advantage of Lyapunov stability arguments that are more specialized to the analysis of optimization algorithms (c.f. [12], for example). This approach allows us to derive a set of conditions that are more closely related to the structural features of the subsystems involved, and are more constructive in that sense. Moreover, our analysis allows us to explicitly identify how the relevant problem parameters affect the tradeoffs in the performance of such schemes.

Instead of treating the optimizer as a difference inclusion with general Lyapunov properties, we give it a structure in which the search vector is explicit. Although less general, this form is common to a large number of optimization methods. In the literature on Lyapunov analysis of such optimization methods, one finds joint conditions on this search vector and candidate Lyapunov functions. These conditions appear to be a natural starting point for the analysis of algorithms with perturbations on their ideal search vectors.

We propose that the plant-optimizer interconnection can often be modeled as an additive perturbation of each subsystem's isolated (i.e., open-loop) dynamics. We show how a scheme involving a gradient descent algorithm that employs a forward-difference estimate of the gradient can be modeled in this way. In treating the interconnection as an additive perturbation, we generalize the result in [1] by removing all assumptions concerning the measurements made by the optimizer in forming its search direction.

We anticipate that the set of conditions we derive here will be valuable in the development and analysis of decentralized ESC schemes.

This paper is organized as follows. In Section II we describe our problem setting and state our assumptions on the open-loop stability properties of the plant and optimizer, and the structure of their interconnection. Our main result, Theorem 3.1, is proved in Section III. In Section IV, we study a special case involving an approximate gradient descent algorithm, and show how our assumptions apply.

A. Notation

All vector norms $\|\cdot\|$ are Euclidean. $(v^k)_{k=0}^{\infty}$ denotes sequence of numbers v^k, $k \in \mathbb{N}$. Temporal sequences are

This work was supported by NSERC (through the Vanier fund and the NSERC/CRD), and Bell Labs/ Alcatel-Lucent.

K. Kvaternik and L. Pavel are with the Edward S. Rogers Department of Electrical and Computer Engineering, University of Toronto {kvaternik,pavel}@utoronto.ca

indexed by subscript - i.e., t_k denotes the kth time instant and $x(t_k^-) = \lim_{t \uparrow t_k} x(t)$. The unit vector along the ith coordinate is denoted by e_i. $\mathbb{R}_+$ is the set of positive real numbers. For a differentiable function $f : \mathbb{R}^n \to \mathbb{R}^m$, defined by $x \mapsto f(x)$, the (i, j)th component of $\nabla f : \mathbb{R}^n \to \mathbb{R}^{n \times m}$ is defined as $[\nabla f(x)]_{i,j} = \frac{\partial f_j(x)}{\partial x_i}$, so that if $f : \mathbb{R}^n \to \mathbb{R}$, the i'th component of f's gradient is denoted by $\nabla f^T e_i$.

II. PROBLEM SETTING

We consider the class of nonlinear systems described by

$$\Sigma : \begin{cases} \dot{x} = f(x, v) \\ y = h(x), \end{cases} \tag{1}$$

where $x \in \mathbb{R}^n$, $v \in \mathbb{R}^p$ and $y \in \mathbb{R}$. To ensure the existence of Σ's solutions, we assume that f is locally Lipschitz. We also make the standard assumption that there exists a continuous *equilibrium map* $l : \mathbb{R}^p \to \mathbb{R}^n$ such that $f(x, v) = 0$ if, and only if $x = l(v)$.

In ESC literature it is typically assumed that for any fixed v, the equilibrium $l(v)$ has some kind of stability property. We make our assumption in terms of the error variable

$$z(t) \triangleq x(t) - l(v(t)), \tag{2}$$

whose dynamics for a fixed $v(t) \equiv v$ are $\dot{z} = f(z + l(v), v)$.

A2.1: There exists a radially unbounded C^1 function $V_\Sigma : \mathbb{R}^n \to \mathbb{R}_+$ such that:

(a) $V_\Sigma(z) > 0$, $\forall z \in \mathbb{R}^n \backslash \{0\}$, and $V_\Sigma(0) = 0$,
(b) there exists a real number $\gamma > 0$ such that $\dot{V}_\Sigma(z(t)) = \nabla V_\Sigma(z)^T f(z + l(v), v) \leq -\gamma V_\Sigma(z(t))$, $\forall z \in \mathbb{R}^n$, $\forall v \in \mathbb{R}^p$. $\diamondsuit$

Next, we define the *reference-to-output* (RO) map $J : \mathbb{R}^p \to \mathbb{R}$ by the composition $v \mapsto h(l(v))$. For the ESC problem to make sense, we need to assume that J has an extremum, which, without loss of generality, we take to be a minimum.

A2.2: There exists a $v^* \in \mathbb{R}^p$ such that $\forall v \in \mathbb{R}^p$, $J(v) \geq J(v^*)$ $\diamondsuit$

The goal in ESC is to find v^* by adjusting Σ's reference input v and monitoring its output $h(x(t))$. In this paper we give conditions under which this task can be accomplished through the sample-and-hold interaction between Σ and a numerical optimization algorithm $\mathscr{O}$ with the following structure:

$$\mathscr{O} : \left\{ v^{k+1} = v^k + s(v^k), \quad \forall k \in \mathbb{N}, \right. \tag{3}$$

where $s(v^k)$ is a search vector at the kth iterate. The generic form (3) is common to many standard numerical optimization methods.

In this paper, we make no assumptions on how the search vector $s(v^k)$ is produced. In this sense we generalize the work in [1], wherein the authors place conditions on the "dither" functions used to probe the RO map, and on how the optimizer's measurements of the RO map enter its update equation. On the other hand, the authors in [1] study a more general class of optimizers – those whose update law is described by difference inclusions.

In the following subsections, we state a collection of assumptions on the open-loop behaviour of $\mathscr{O}$, and the nature of the interconnection between Σ and $\mathscr{O}$.

A. The Open-Loop Operation of $\mathscr{O}$

Given a static cost function $J(v)$, we assume that $\mathscr{O}$ produces a sequence $(v^k)_{k=1}^\infty$ that asymptotically converges to v^* for any initial condition v^0. Although there are many ways to make this assumption precise, our choice is motivated by Lyapunov arguments from the optimization literature [12].

A2.3: There exists a radially unbounded C^1 function $V_\mathscr{O} : \mathbb{R}^p \to \mathbb{R}_+$ with the following properties:

(a) $V_\mathscr{O}(v) > 0$, $\forall v \in \mathbb{R}^p \backslash \{v^*\}$, and $V_\mathscr{O}(v^*) = 0$,
(b) $\nabla V_\mathscr{O}(v)^T s(v) < 0$, for all $v \in \mathbb{R}^p \backslash \{v^*\}$, and $\nabla V_\mathscr{O}(v^*)^T s(v^*) = 0$,
(c) there exists a real number $\kappa_s > 0$ such that $\|s(v)\|^2 \leq -\kappa_s \nabla V_\mathscr{O}(v)^T s(v)$, $\forall v \in \mathbb{R}^p$,
(d) the gradient of $V_\mathscr{O}$ is Lipschitz – i.e., there exists a real number $L_{\nabla V_\mathscr{O}} > 0$, such that for all $v_1, v_2 \in \mathbb{R}^p$, $\|\nabla V_\mathscr{O}(v_1) - \nabla V_\mathscr{O}(v_2)\| \leq L_{\nabla V_\mathscr{O}} \|v_1 - v_2\|$. $\diamondsuit$

The first two properties in A2.3 are analogous to those demanded of V_Σ in A2.1, while properties similar to the last two appear in Lyapunov-based stability proofs of various optimization algorithms [12].

Remark 2.1: Property (c) in A2.3 is a joint condition on $\nabla V_\mathscr{O}(v)$ and the search vector $s(v)$. For a given function $V_\mathscr{O}(v)$, it is possible that the range of κ_s for which this property is satisfied is tunable through the design of $s(v)$. For example, (as we will see in Section IV), such is the case when $s(v) = -\alpha \nabla J(v)$, and α is a tunable step size. $\diamondsuit$

In Lemma 2.2, we show why A2.3 guarantees that $\mathscr{O}$ generates sequences that asymptotically converge to v^*. However, first we need the technical Lemma 2.1.

Lemma 2.1: Given any differentiable function $g : \mathbb{R}^m \to \mathbb{R}$ and any $a, b \in \mathbb{R}^m$, we can write

$$\begin{aligned} g(a+b) = g(a) + \nabla g(a)^T b \\ + \int_0^1 [\nabla g(a + \tau b) - \nabla g(a)]^T b \, d\tau. \end{aligned} \tag{4}$$

Proof: Let $u(\tau) = a + \tau b$. Then by the chain rule,

$$\tfrac{d}{d\tau} g(a + \tau b) = \nabla_u g(u)^T \nabla_\tau u(\tau) = \nabla g(a + \tau b)^T b.$$

By Leibniz's rule, $\int_0^1 \frac{d}{d\tau} g(a + \tau b) \, d\tau = g(a+b) - g(a)$, and therefore

$$g(a+b) = g(a) + \int_0^1 \nabla g(a + \tau b)^T b \, d\tau. \tag{5}$$

The result then follows by adding and subtracting $\nabla g(a)^T b$ to the right-hand side of (5). $\square$

Lemma 2.2: Define $\Delta V_\mathscr{O}^k \triangleq V_\mathscr{O}(v^k + s(v^k)) - V_\mathscr{O}(v^k)$, and assume that A2.2 and A2.3 hold. Then, $\Delta V_\mathscr{O}^k < 0$, provided that $\kappa_s < \frac{2}{L_{V_\mathscr{O}}}$.

Proof: By Lemma 2.1, we may expand $V_\mathscr{O}$ about v^k as

$$\begin{aligned} V_\mathscr{O}(v^k + s(v^k)) \leq V_\mathscr{O}(v^k) + \nabla V_\mathscr{O}(v^k)^T s(v^k) \\ + \int_0^1 \left[\nabla V_\mathscr{O}(v^k + \tau s(v^k)) - \nabla V_\mathscr{O}(v^k) \right]^T s(v^k) \, d\tau. \end{aligned}$$

Since the gradient of $V_{\mathscr{O}}$ is Lipschitz and $v^k + s(v^k) = v^{k+1}$, we have that

$$\Delta V_{\mathscr{O}}^k \leq \nabla V_{\mathscr{O}}(v^k)^T s(v^k) + \frac{L_{V_{\mathscr{O}}}}{2}\|s(v^k)\|^2. \quad (6)$$

From property (c) in A2.3, we obtain

$$\Delta V_{\mathscr{O}}^k \leq (1 - \frac{L_{V_{\mathscr{O}}}}{2}\kappa_s)\nabla V_{\mathscr{O}}(v^k)^T s(v^k), \quad (7)$$

and the conclusion then follows from A2.3 (b), and the fact that $\kappa_s \in (0, \frac{2}{L_{V_{\mathscr{O}}}})$. □

Lemma 2.2 indicates that $V_{\mathscr{O}}$ decreases along successive iterates generated by (3), as long as $v^k \neq v^*$. From the Lyapunov theory of difference equations, we conclude that v^* is asymptotically stable for $\mathscr{O}$ (c.f. Theorem 5.9.2 in [13], or Corollary 4.8.2 in [14], for example).

Remark 2.2: Some optimization algorithms do not necessarily produce a descent sequence on the cost function J. Therefore, J is not always a natural choice for a Lyapunov function. ♢

B. The Closed-Loop Operation of $\mathscr{O}$

The optimizer $\mathscr{O}$ in (3) is designed to minimize the RO map $J(v)$. In order to formulate its search vector at each iteration, $\mathscr{O}$ must make measurements of $J(v)$ whose analytic structure is not known. In closed-loop with Σ, these measurements are corrupted by the transient error $z(t)$ defined in (2); $\mathscr{O}$ measures $h(x(t))$ instead of the true RO map $J(v) = h(l(v))$. It is therefore reasonable to view the effect of Σ on $\mathscr{O}$ as a perturbation of $\mathscr{O}$'s ideal evolution (3). When feedback interconnected with Σ through a sample-and-hold, the optimizer evolves according to

$$\hat{\mathscr{O}} : \begin{cases} v^{k+1} = v^k + \hat{s}(v^k, z^k), & \forall k \in \mathbb{N} \\ v(t) = v^k, & \forall t \in [t_{k-1}, t_k), \end{cases} \quad (8)$$

where $\hat{s}(v^k, z^k)$ is the perturbed search vector and z^k is defined as the transient error (2) just before the kth time instant:

$$z^k \triangleq x(t_k^-) - l(v^k). \quad (9)$$

For simplicity, we take the sampling interval to remain constant across iterations – i.e. $t_{k+1} - t_k = T > 0$, $\forall k \in \mathbb{N}$.

We make two assumptions on the structure of $\hat{s}(v^k, z^k)$; assumption A2.4 states that the perturbation on $s(v^k)$ affects it additively, and A2.5 requires that the nonvanishing component of this perturbation is "small" relative to V_Σ. We state:

A2.4: There exists a continuous $p : \mathbb{R}^n \to \mathbb{R}^p$ such that for all $v \in \mathbb{R}^p$ and $z \in \mathbb{R}^n$, $\hat{s}(v,z) = s(v) + p(z)$. ♢

Aside from requiring $p(z)$ to enter additively, we make no further assumptions on its structure; $p(z)$ is free to embody other errors that are independent of the transient error, and therefore may not vanish when z is set to 0.

Next, we express the perturbation $p(v)$ in terms of its vanishing and nonvanishing components

$$p(z^k) = \Big(p(z^k) - p(0)\Big) + p(0) \triangleq p_v(z^k) + p_o, \quad (10)$$

where $p_o \triangleq p(0)$ and $p_v(0) = 0$, and state:

A2.5: There exists a real number $\kappa_\Sigma > 0$ such that for all $z \in \mathbb{R}^n$, $\kappa_\Sigma V_\Sigma(z) \geq \|p_v(z)\|^2$. ♢

In Section IV, we show how these assumptions are satisfied for the case in which $\mathscr{O}$ is the gradient descent algorithm that employs an Euler approximation of $\nabla J(v)$ at each iteration, and Σ is a linear, stable system.

C. The Closed-Loop Operation of Σ

In order to express the perturbation on Σ arising from its interconnection to $\hat{\mathscr{O}}$, we make the following assumption, which is common in ESC literature [1], [2].[1]

A2.6: The equilibrium map $l(v)$ is Lipschitz on $\mathbb{R}^p$ with constant L_l. ♢

The effect of $\hat{\mathscr{O}}$ on Σ can be derived from the properties of V_Σ given in Assumption A2.1. We summarize this effect in the next Lemma.

Lemma 2.3: Let T be the constant sampling period $t_{k+1} - t_k$ and define $\Delta V_\Sigma^k \triangleq V_\Sigma(z^{k+1}) - V_\Sigma(z^k)$, with z^k as in (9). Under Assumptions A2.1 and A2.6,

$$\begin{aligned} \Delta V_\Sigma^k \leq & -(1 - e^{-\gamma T})V_\Sigma(z^k) + e^{-\gamma T}\|\hat{s}(v^k, z^k)\|^2 \\ & + \tfrac{1}{4}e^{-\gamma T}(L_{V_\Sigma} L_l)^2, \end{aligned} \quad (11)$$

for all $((z^k)^T, (v^k)^T)^T \in \Omega_z \times \Omega_{\mathscr{O}}$, where $\Omega_z \subset \mathbb{R}^n$ and $\Omega_{\mathscr{O}} \subset \mathbb{R}^p$ are arbitrarily large, compact sets containing the origin and v^*, respectively.

Proof: From A2.1 we have that $\forall z(0) \in \mathbb{R}^n$, $V_\Sigma(z(t)) \leq e^{-\gamma t}V_\Sigma(z(0))$. Consider the interconnection of Σ to $\hat{\mathscr{O}}$, (c.f. (1) and (8)). According to our sample-and-hold convention in (8), $z(t_k^-) \neq z(t_k) = z(t_k^+)$ since the transition from v^k to v^{k+1} occurs at t_k. Using (2) and (9), we obtain

$$\begin{aligned} V_\Sigma(z(t_{k+1}^-)) \equiv V_\Sigma(z^{k+1}) &\leq e^{-\gamma(t_{k+1} - t_k)}V_\Sigma(z(t_k)) \\ &= e^{-\gamma T}V_\Sigma(x(t_k) - l(v^{k+1})) \\ &= e^{-\gamma T}V_\Sigma(x(t_k^-) - l(v^k) + l(v^k) - l(v^{k+1})) \\ &= e^{-\gamma T}V_\Sigma(z^k + [l(v^k) - l(v^{k+1})]). \end{aligned} \quad (12)$$

By A2.1, V_Σ is C^1, which means that it is locally Lipschitz on any compact set. Let L_{V_Σ} be its Lipschitz constant on the compact set

$$\begin{aligned} S = \{z \in \mathbb{R}^n : z = \zeta + l(v) - l(v + \hat{s}(v, \zeta)), \\ \zeta \in \Omega_z,\ v \in \Omega_{\mathscr{O}}\}. \end{aligned} \quad (13)$$

Next, add and subtract $e^{-\gamma T}V_\Sigma(z^k)$ to the right-hand side of (12) and use A2.6 to obtain

$$\begin{aligned} V_\Sigma(z^{k+1}) \leq & e^{-\gamma T}V_\Sigma(z^k + [l(v^k) - l(v^{k+1})]) \\ & - e^{-\gamma T}V_\Sigma(z^k) + e^{-\gamma T}V_\Sigma(z^k) \\ \leq & e^{-\gamma T}V_\Sigma(z^k) + e^{-\gamma T}L_{V_\Sigma}L_l\|v^{k+1} - v^k\|. \end{aligned}$$

Since $\|v^{k+1} - v^k\| = \|\hat{s}(v^k, z^k)\|$ by (8), we write

$$V_\Sigma(z^{k+1}) \leq e^{-\gamma T}V_\Sigma(z^k) + e^{-\gamma T}L_{V_\Sigma}L_l\|\hat{s}(v^k, z^k)\|. \quad (14)$$

[1] In both [1] and [2], the equilibrium map $l(v)$ is required only to be locally Lipschitz. Although it is possible to carry out our analysis in terms of the weaker local Lipschitz property, we choose to work with the global property in order to simplify our presentation.

From Young's inequality, we have that for any $\varepsilon_1 > 0$, $\|\hat{s}(v^k, z^k)\| \le \varepsilon_1 \|\hat{s}(v^k, z^k)\|^2 + \frac{1}{4\varepsilon_1}$. Taking $\varepsilon_1 = \frac{1}{L_{V_\Sigma} L_l}$ we write

$$V_\Sigma(z^{k+1}) \le e^{-\gamma T} V_\Sigma(z^k) + e^{-\gamma T} \|\hat{s}(v^k, z^k)\|^2 + \tfrac{1}{4} e^{-\gamma T} (L_{V_\Sigma} L_l)^2,$$

and the required form (11) then follows by subtracting $V_\Sigma(z^k)$ from both sides. □

The conclusion in Lemma 2.3 is only valid when $((z^k)^T, (v^k)^T)^T \in \Omega_z \times \Omega_{\mathscr{O}}$, where Ω_z and $\Omega_{\mathscr{O}}$ are some compact sets that are chosen to be arbitrarily large. Once fixed, these sets determine the size of L_{V_Σ}, which ultimately affects the size of the neighbourhood of $(0^T, v^{*T})$ to which the closed-loop trajectories converge. In the next section, we complete our closed-loop analysis, and provide conditions on the initialization of $\Sigma - \mathscr{O}$, which ensure that the resulting trajectories remain inside $\Omega_z \times \Omega_{\mathscr{O}}$.

III. MAIN RESULT

In order to prove our main theorem, we make use of the following assumption on the relationship between the ideal search vector $s(v^k)$ and $V_{\mathscr{O}}$:

A3.1: There exists a real number $\kappa_{V_{\mathscr{O}}} > 0$ such that for all $v \in \mathbb{R}^p$, $\nabla V_{\mathscr{O}}^T \big(s(v) + \kappa_{V_{\mathscr{O}}} \nabla V_{\mathscr{O}}(v)\big) \le 0$. ♢

Remark 3.1: Assumption A3.1 expresses that the search vector $\frac{s(v)}{\|s(v)\|}$ should be close to the direction $\frac{-\nabla V_{\mathscr{O}}(v)}{\|\nabla V_{\mathscr{O}}(v)\|}$. We note that A3.1 resembles property (c) imposed on $V_{\mathscr{O}}$ in assumption A2.3. In fact, by geometric reasoning we can show that if the angle between the vectors $\frac{s(v)}{\|s(v)\|}$ and $\frac{-\nabla V_{\mathscr{O}}(v)}{\|\nabla V_{\mathscr{O}}(v)\|}$ is bounded by some number $\theta \in (0, \pi/2)$ for all $v \in \mathbb{R}^p$, then given A2.3, A3.1 is automatically satisfied for all $\kappa_{V_{\mathscr{O}}} \le \kappa_s \cos^2(\theta)$. ♢

The following theorems give sufficient conditions under which the $\Sigma - \hat{\mathscr{O}}$ feedback interconnection exhibits semiglobal asymptotic practical stability with respect to $(0, v^*) \in \mathbb{R}^n \times \mathbb{R}^p$.

Theorem 3.1: Consider the composite Lyapunov function $V(z^k, v^k) = V_\Sigma(z^k) + V_{\mathscr{O}}(v^k)$ and assume that A2.1 to A2.6, and A3.1 hold, with

$$\kappa_s < \frac{1}{L_{\nabla V_{\mathscr{O}}} + 2}. \tag{15}$$

Then, there exists a neighbourhood Ω_o of $((0)^T, (v^*)^T)$ that can be made arbitrarily large, and the associated positive, real numbers κ_Σ^* and T^*, such that whenever $\kappa_\Sigma < \kappa_\Sigma^*$, $T > T^*$ and $((z^0)^T, (v^0)^T)^T \in \Omega_o$, $V(z^k, v^k)$ decreases along the trajectories of the closed-loop system (1)–(8) according to

$$\Delta V^k \le -C_\Sigma V_\Sigma(z^k) - C_{\mathscr{O}} \|\nabla V_{\mathscr{O}}(v^k)\|^2 + C, \tag{16}$$

where $\Delta V^k = V(z^{k+1}, v^{k+1}) - V(z^k, v^k)$, the real numbers C_Σ, $C_{\mathscr{O}}$ and C are positive, and given by

$$C = \Big(2L_{\nabla V_{\mathscr{O}}} + 4e^{-\gamma T} + \tfrac{1}{\kappa_{\nabla V_{\mathscr{O}}}(1-\kappa_s(L_{\nabla V_{\mathscr{O}}}+2))}\Big)\|p_o\|^2 + \tfrac{1}{4} e^{-\gamma T} (L_{V_\Sigma} L_l)^2 \tag{17}$$

$$C_{\mathscr{O}} = \kappa_{\nabla V_{\mathscr{O}}}\big(\tfrac{1}{2} - \kappa_s(\tfrac{1}{2} L_{\nabla V_{\mathscr{O}}} + 2e^{-\gamma T} - 1)\big) \tag{18}$$

$$C_\Sigma = 1 - e^{-\gamma T} - \kappa_\Sigma \Big(2L_{\nabla V_{\mathscr{O}}} + 4e^{-\gamma T} + \tfrac{1}{\kappa_{\nabla V_{\mathscr{O}}}(1-\kappa_s(L_{\nabla V_{\mathscr{O}}}+2))}\Big). \tag{19}$$

Proof: Let $\Delta V_{\hat{\mathscr{O}}} = V_{\mathscr{O}}(v^k + \hat{s}(v^k, z^k)) - V_{\mathscr{O}}(v^k)$, and $\Delta V_\Sigma = V_\Sigma(z^{k+1}) - V_\Sigma(z^k)$. By arguments similar to those used in Lemma 2.2 to obtain (6), we may express $\Delta V_{\hat{\mathscr{O}}}$ as

$$\Delta V_{\hat{\mathscr{O}}}^k \le \nabla V_{\mathscr{O}}(v^k)^T \hat{s}(v^k, z^k) + \tfrac{L_{\nabla V_{\mathscr{O}}}}{2} \|\hat{s}(v^k, z^k)\|^2.$$

For notational simplicity, we drop all arguments from now on. Suppose that at some k, $((z^k)^T, (v^k)^T)^T \in \Omega_z \times \Omega_{\mathscr{O}}$. Then we may apply Lemma 2.3 to write

$$\begin{aligned}\Delta V^k &= \Delta V_\Sigma^k + \Delta V_{\hat{\mathscr{O}}}^k \\ &\le -(1 - e^{-\gamma T}) V_\Sigma + \nabla V_{\mathscr{O}}^T \hat{s} + \big(\tfrac{L_{\nabla V_{\mathscr{O}}}}{2} + e^{-\gamma T}\big)\|\hat{s}\|^2 \\ &\quad + \tfrac{1}{4} e^{-\gamma T} (L_{V_\Sigma} L_l)^2.\end{aligned}$$

From A2.4 and (10),

$$\begin{aligned}\|\hat{s}(v^k, z^k)\|^2 &= \|s(v^k) + p_v(z^k) + p_o\|^2 \\ &\le 2\|s(v^k)\|^2 + 4\|p_v(z^k)\|^2 + 4\|p_o\|^2,\end{aligned}$$

where the inequality is obtained by twice applying $\|a+b\|^2 \le 2\|a\|^2 + 2\|b\|^2$. We therefore have

$$\begin{aligned}\Delta V^k &\le -(1 - e^{-\gamma T}) V_\Sigma + \nabla V_{\mathscr{O}}^T s + \nabla V_{\mathscr{O}}^T p_v + \nabla V_{\mathscr{O}}^T p_o - \tfrac{\|p_o\|^2}{4\kappa_{p_o}} \\ &\quad + \big(L_{\nabla V_{\mathscr{O}}} + 2e^{-\gamma T}\big)\|s\|^2 + \big(2L_{\nabla V_{\mathscr{O}}} + 4e^{-\gamma T}\big)\|p_v\|^2 \\ &\quad + \Big[\big(2L_{\nabla V_{\mathscr{O}}} + 4e^{-\gamma T}\big)\|p_o\|^2 + \tfrac{1}{4} e^{-\gamma T}(L_{V_\Sigma} L_l)^2 + \tfrac{\|p_o\|^2}{4\kappa_{p_o}}\Big].\end{aligned}$$

where we have also added and subtracted the term $\frac{\|p_o\|^2}{4\kappa_{p_o}}$, in which κ_{p_o} is some real number to be specified later. Let $C = \big(2L_{\nabla V_{\mathscr{O}}} + 4e^{-\gamma T} + \frac{1}{4\kappa_{p_o}}\big)\|p_o\|^2 + \frac{1}{4} e^{-\gamma T}(L_{V_\Sigma} L_l)^2$. Applying A2.3 (c) to express $\|s\|^2$ in terms of $\nabla V_{\mathscr{O}}^T s$, we write

$$\begin{aligned}\Delta V^k &\le -(1 - e^{-\gamma T}) V_\Sigma + (1 - \kappa_s(L_{\nabla V_{\mathscr{O}}} + 2e^{-\gamma T}))\nabla V_{\mathscr{O}}^T s + C \\ &\quad + \nabla V_{\mathscr{O}}^T p_v + \nabla V_{\mathscr{O}}^T p_o + \big(2L_{\nabla V_{\mathscr{O}}} + 4e^{-\gamma T}\big)\|p_v\|^2 - \tfrac{\|p_o\|^2}{4\kappa_{p_o}}.\end{aligned}$$

Next, we focus on the second term in the above inequality. Noting first that by (15), the quantity $(1 - \kappa_s(L_{\nabla V_{\mathscr{O}}} + 2e^{-\gamma T})) > 0$ for all $T \ge 0$, we apply A3.1 to express $\nabla V_{\mathscr{O}}^T s$ in terms of $\|\nabla V_{\mathscr{O}}\|^2$:

$$\begin{aligned}\Delta V^k &\le -(1 - e^{-\gamma T}) V_\Sigma + \nabla V_{\mathscr{O}}^T p_v + \nabla V_{\mathscr{O}}^T p_o \\ &\quad - \kappa_{\nabla V_{\mathscr{O}}}\big(1 - \kappa_s(L_{\nabla V_{\mathscr{O}}} + 2e^{-\gamma T})\big)\|\nabla V_{\mathscr{O}}\|^2 \\ &\quad + \big(2L_{\nabla V_{\mathscr{O}}} + 4e^{-\gamma T}\big)\|p_v\|^2 - \tfrac{1}{4\kappa_{p_o}}\|p_o\|^2 + C.\end{aligned}$$

Using A2.5, we absorb the $\big(2L_{\nabla V_{\mathscr{O}}} + 4e^{-\gamma T}\big)\|p_v\|^2$ term into V_Σ, and introduce the negative term $\frac{-1}{4\kappa_{p_v}}\|p_v\|^2$ (with κ_{p_v} to be specified later), which allows us to dominate the cross term $\nabla V_{\mathscr{O}}^T p_v$. For the same reason, we extract a $-(\kappa_{p_v} + \kappa_{p_o})\|\nabla V_{\mathscr{O}}\|^2$ term from $-\kappa_{\nabla V_{\mathscr{O}}}\big(1 - \kappa_s(L_{\nabla V_{\mathscr{O}}} + 2\kappa_1)\big)\|\nabla V_{\mathscr{O}}\|^2$, and write

$$\begin{aligned}\Delta V^k &\le -\big(1 - e^{-\gamma T} - \kappa_\Sigma(2L_{\nabla V_{\mathscr{O}}} + 4e^{-\gamma T} + \tfrac{1}{4\kappa_{p_v}})\big)V_\Sigma \\ &\quad - \big(\kappa_{\nabla V_{\mathscr{O}}}\big(1 - \kappa_s(L_{\nabla V_{\mathscr{O}}} + 2e^{-\gamma T})\big) - \kappa_{p_v} - \kappa_{p_o}\big)\|\nabla V_{\mathscr{O}}\|^2 \\ &\quad - \kappa_{p_v}\|\nabla V_{\mathscr{O}}\|^2 - \kappa_{p_o}\|\nabla V_{\mathscr{O}}\|^2 + \nabla V_{\mathscr{O}}^T p_v + \nabla V_{\mathscr{O}}^T p_o \\ &\quad - \tfrac{1}{4\kappa_{p_v}}\|p_v\|^2 - \tfrac{1}{4\kappa_{p_o}}\|p_o\|^2 + C.\end{aligned}$$

Then, we define

$$C_\Sigma = 1 - e^{-\gamma T} - \kappa_\Sigma (2L_{\nabla V_\mathscr{O}} + 4e^{-\gamma T} + \tfrac{1}{4\kappa_{p_v}}) \qquad \text{and} \quad (20)$$

$$C_\mathscr{O} = \kappa_{\nabla V_\mathscr{O}} \big(1 - \kappa_s (L_{\nabla V_\mathscr{O}} + 2e^{-\gamma T})\big) - \kappa_{p_v} - \kappa_{p_o}, \qquad (21)$$

and apply completion of squares to obtain

$$\begin{aligned}\Delta V^k &\leq -C_\Sigma V_\Sigma - C_\mathscr{O} \|\nabla V_\mathscr{O}\|^2 + C \\ &\quad - \kappa_{p_v} \left\| \nabla V_\mathscr{O} - \tfrac{1}{2\kappa_{p_v}} p_v \right\|^2 - \kappa_{p_o} \left\| \nabla V_\mathscr{O} - \tfrac{1}{2\kappa_{p_o}} p_o \right\|^2 \\ &\leq -C_\Sigma V_\Sigma - C_\mathscr{O} \|\nabla V_\mathscr{O}\|^2 + C,\end{aligned}$$

which is the required form (16).

Next, we show that C_Σ and $C_\mathscr{O}$ are positive for a sufficiently small κ_Σ, and a sufficiently large T. To ensure that $C_\mathscr{O} > 0$, we pick any $\kappa_{p_v} > 0$ and $\kappa_{p_o} > 0$ so small that $\kappa_{p_v} + \kappa_{p_o} < \kappa_{\nabla V_\mathscr{O}} (1 - \kappa_s (L_{\nabla V_\mathscr{O}} + 2e^{-\gamma T})|_{T=0})$. One possible choice, which leads to the expressions (17), (18) and (19), is to let

$$\kappa_{p_v} = \kappa_{p_o} = \tfrac{1}{4} \kappa_{\nabla V_\mathscr{O}} (1 - \kappa_s (L_{\nabla V_\mathscr{O}} + 2)).$$

We rearrange (20) as

$$C_\Sigma = 1 - \big(1 + 4\kappa_\Sigma\big) e^{-\gamma T} - \kappa_\Sigma \big(2L_{\nabla V_\mathscr{O}} + \tfrac{1}{4\kappa_{p_v}}\big), \qquad (22)$$

pick any $\varepsilon_2 \in (0,1)$ and define

$$\kappa_\Sigma^* \triangleq \frac{1 - \varepsilon_2}{2L_{\nabla V_\mathscr{O}} + \frac{1}{4\kappa_{p_v}}}. \qquad (23)$$

If $\kappa_\Sigma < \kappa_\Sigma^*$, then

$$1 - \kappa_\Sigma \big(2L_{\nabla V_\mathscr{O}} + \tfrac{1}{4\kappa_{p_v}}\big) > \varepsilon_2 > 0,$$

and we see that choosing any $T > T^*$, where

$$T^* = \frac{1}{\gamma} \ln \left(\frac{1 + 4\kappa_\Sigma}{\varepsilon_2} \right), \qquad (24)$$

renders $C_\Sigma > 0$.

Finally, given that V_Σ is only locally Lipschitz, we need to specify a set of initial conditions for which our use of Lemma 2.3 remains valid for all subsequent $k > 0$. Let us fix an arbitrarily large, compact set $\Omega_z \times \Omega_\mathscr{O} \ni ((0)^T, (v^*)^T)$ and set L_{V_Σ} to be the Lipschitz constant associated with V_Σ on the set S discussed in the proof of Lemma 2.3 (c.f. (13)). Let $\Omega_o \subset \mathbb{R}^n \times \mathbb{R}^p$ be the largest sublevel set of $V(z,v)$ contained inside $\Omega_z \times \Omega_\mathscr{O}$. By standard arguments, this set is nonempty and positively invariant with respect to $\Sigma - \mathscr{O}$. Since V_Σ is assumed to be continuous and radially unbounded, Ω_o can be made arbitrarily large through our choice of $\Omega_z \times \Omega_\mathscr{O}$. □

Remark 3.2: Inequality (15) imposes a growth condition on $\|s\|$ relative to $\|V_\mathscr{O}\|$, as is sometimes done in open-loop stability analyses of optimization algorithms [12]. In the feedback interconnection with Σ, this condition becomes more stringent. The inequality $\kappa_\Sigma < \kappa_\Sigma^*$ translates into a condition on the "strength" of Σ's stability at $l(v^k)$ relative to the strength of the destabilizing effect of the interconnection terms $p_v(z^k)$ and p_o. Finally, the lower-bound (24) on the sampling period T is analogous to the time-scale separation requirement in continuous-time ESC schemes. The authors in [1] refer to T as the *waiting time*. ♢

Remark 3.3: It is worth noting that C in (17) is comprised of two components: one that can be made arbitrarily small by increasing T, and the other which is multiplied by $\|p_o\|$. In Section IV we show that when the optimizer $\mathscr{O}$ is the gradient descent algorithm employing the forward-Euler approximation of the gradient of J, the non-vanishing component of the perturbation (i.e. p_o) is parameterized by the Euler step size μ, and can be made arbitrarily small by choosing μ small. ♢

Although C_Σ and $C_\mathscr{O}$ are related to the convergence rate of the scheme, C dictates the size of the ultimate bound on the sequence $\big(\|((z^k)^T, (v^k - v^*)^T)^T\|\big)_{k=0}^\infty$. We state this more precisely in Definition 3.1 and Theorem 3.2, where we demonstrate the semi-global, practical, asymptotic stability of $(0^T, (v^*)^T)$ for $\Sigma - \mathscr{O}$.

Definition 3.1: The point $(0^T, (v^*)^T)$ is said to be *semiglobally practically asymptotically stable* for the closed-loop system $\Sigma - \mathscr{O}$ if:

1) There exist two compact subsets of $\mathbb{R}^n \times \mathbb{R}^p$, P and Ω_o, with $P \subset \Omega_o$, both containing $(0^T, (v^*)^T)$, and both being positively invariant with respect to $\Sigma - \mathscr{O}$. Furthermore, each trajectory of $\Sigma - \mathscr{O}$ initiated inside $\Omega_o \backslash P$ must enter P in finitely many iterations.
2) $\Sigma - \mathscr{O}$ is parameterized by a set of tunable variables that can be adjusted to render Ω_o arbitrarily large, and P arbitrarily small. ♢

The proof of our next theorem may be compared to that of Theorem 5.14.2 in [13].

Theorem 3.2: Assume that the conditions of Theorem 3.1 are satisfied with $\kappa_\Sigma < \kappa_\Sigma^*$. and $T > T^*$. Furthermore, assume that the nonvanishing perturbation p_o is parameterized by a tunable variable μ as $p_o = \mu \overline{p}_o$. Then, the closed-loop system $\Sigma - \mathscr{O}$ is semiglobally practically asymptotically stable at $(0^T, (v^*)^T)^T$.

Proof: Let w denote the point $(z^T, v^T)^T \in \mathbb{R}^n \times \mathbb{R}^p$ and $w^* = (0^T, v^{*T})$. We define the set

$$Z \triangleq \{w \in \mathbb{R}^n \times \mathbb{R}^p : C_\Sigma V_\Sigma(z) + C_\mathscr{O} \|\nabla V_\mathscr{O}(v)\|^2 \leq C\}, \quad (25)$$

on which $V(w^k)$ is no longer guaranteed to decrease with successive iterations of $\Sigma - \mathscr{O}$. By A2.1 and A2.3, the function $F : w \mapsto C_\Sigma V_\Sigma(z) + C_\mathscr{O} \|\nabla V_\mathscr{O}(v)\|^2$ is continuous and positive definite[2]. Consequently, Z is compact for a sufficiently small $C > 0$. Moreover, since

$$\begin{aligned} C = \mu^2 \Big(2L_{\nabla V_\mathscr{O}} + 4e^{-\gamma T} + \tfrac{1}{\kappa_{\nabla V_\mathscr{O}} (1 - \kappa_s (L_{\nabla V_\mathscr{O}} + 2))} \Big) \|\overline{p}_o\|^2 \\ + \tfrac{1}{4} e^{-\gamma T} (L_{V_\Sigma} L_l)^2, \end{aligned} \qquad (26)$$

can be made arbitrarily small for a sufficiently small μ and a sufficiently large T, we see that Z itself can be made arbitrarily small by the continuity of F.

We may now construct the required set P discussed in Definition 3.1. By the compactness of Z and the continuity of V, there exists a number $\beta = \max\{V(w) : w \in Z\}$. Then,

[2]To see that $\|\nabla V_\mathscr{O}(v)\|$ is positive definite, suppose that there exists some $\tilde{v} \neq v^*$ for which $\|\nabla V_\mathscr{O}(\tilde{v})\| = 0$. In that case, $\nabla V_\mathscr{O}(\tilde{v})^T s(\tilde{v}) = 0$ regardless of s, thus violating A2.3 (b).

$\Omega_\beta = \{w \in \mathbb{R}^n \times \mathbb{R}^p : V(w) \le \beta\}$ is the smallest sublevel set of V strictly containing Z. We claim that the set

$$\Omega_{\beta+C} \triangleq \{w \in \mathbb{R}^n \times \mathbb{R}^p : V(w) \le \beta + C\} \tag{27}$$

is positively invariant with respect to $\Sigma - \hat{\mathscr{O}}$. To see this, suppose that at some $k_o \in \mathbb{N}$, w^{k_o} is inside $\Omega_{\beta+C}$. There are two possibilities: either $w^{k_o} \in \Omega_{\beta+C} \backslash Z$ or $w^{k_o} \in Z$. In the first case, $\Delta V^{k_o} = V(w^{k_o+1}) - V(w^{k_o}) < 0$, which means that $V(w^{k_o+1}) < V(w^{k_o}) \le \beta + C$, and therefore w^{k_o+1} is also inside $\Omega_{\beta+C}$. In the second case, $\Delta V^{k_o} \ge 0$ (by the definition of Z), but it is greater than zero by at most C – i.e. $V(w^{k_o+1}) - V(w^{k_o}) \le C$. Since $w^{k_o} \in Z \subset \Omega_\beta$, $V(w^{k_o}) \le \beta$. Therefore $V(w^{k_o+1}) \le V(w^{k_o}) + C \le \beta + C$, which means again that w^{k_o+1} is inside $\Omega_{\beta+C}$. Induction then allows us to conclude the positive invariance of $\Omega_{\beta+C}$. Since Ω_β is the smallest sublevel set of V containing Z, it is clear that $\Omega_{\beta+C}$ is the smallest positively invariant set containing Z.

Next, let ε be a positive, (arbitrarily small) real number and consider the larger sublevel set

$$P \triangleq \{w \in \mathbb{R}^n \times \mathbb{R}^p : V(w) \le \beta + C + \varepsilon\}, \tag{28}$$

which is compact since both V_Σ and $V_{\mathscr{O}}$ are positive definite and radially unbounded. We note that by the construction of P from Z, and the fact that Z can be made arbitrarily small via T and μ, P can likewise be made arbitrarily small.

Choose any Ω_o according to Theorem 3.1, large enough to strictly contain P. Such a choice is always possible by the radial unboundedness of V_Σ and $V_{\mathscr{O}}$. In the following, we show that all trajectories initiated inside $\Omega_o \backslash P$ enter P in finitely many iterations. Since Ω_o is compact, there exists a number

$$a = \min\{-\Delta V(w) : w \in \Omega_o \backslash P\}. \tag{29}$$

Suppose that $\Sigma - \hat{\mathscr{O}}$ is initialized at $w^0 \in \Omega_o \backslash P$. Then $V(w^{k+1}) < V(w^k) - a$ and therefore $V(w^k) < V(w^0) - ka$, which implies that $w^k \in P$ for all $k > K(a(\varepsilon), w^0)$, where

$$K = \left\lceil \frac{V(w^0) - \beta - C - \varepsilon}{a} \right\rceil. \tag{30}$$

Since w^0 is arbitrary, it stands that all trajectories initiated inside $\Omega_o \backslash P$ enter P in finitely many iterations, and Definition 3.1 is satisfied. □

IV. SPECIAL CASE

To illustrate the application of our assumptions, in this section we analyze an ESC scheme involving the plant

$$\begin{aligned} \dot{x} &= Ax + g(v) \\ y &= h(x), \qquad y \in \mathbb{R} \end{aligned} \tag{31}$$

and the optimizer

$$v^{k+1} = v^k - \alpha \nabla J(v^k), \tag{32}$$

which is the gradient descent algorithm with a fixed step size α. If A is strictly Hurwitz, then A2.1 is satisfied with $V_\Sigma(z) = \frac{1}{2} z^T z$, which is locally Lipschitz on any arbitrarily large compact set $\Omega_z \subset \mathbb{R}^n$. The reference-to-output map to be minimized is $J(v) = h(-A^{-1} g(v))$. If J is strictly convex and bounded from below, then A2.2 is satisfied. If g is Lipschitz then A2.6 is satisfied. Since for a sufficiently small step size α, (32) produces a descent sequence on J, it is appropriate to use $V_{\mathscr{O}}(v) = J(v) - J(v^*)$. A2.3 is then satisfied provided that J is C^1, $\alpha > 0$ and $\nabla J(v)$ is Lipschitz with a constant $L_{\nabla V_{\mathscr{O}}}$. The second property in A2.3 is satisfied for all $\kappa_s \ge \alpha$, while A3.1 is satisfied for all $\kappa_{V_{\mathscr{O}}} \le \alpha$ – i.e., this example jointly satisfies A2.3 and A3.1 with $\kappa_{V_{\mathscr{O}}} = \kappa_s = \alpha$.

The ESC problem makes sense if the analytic structure of either h or g (or both) is unknown. In that case $\nabla J(v)$ must be estimated through the measurements of $J(v)$. Let us consider the simplest derivative-free approximation of the gradient descent algorithm – an optimizer $\hat{\mathscr{O}}$ that uses a fixed-distance, forward-difference (FD) approximation of ∇J. At the kth iteration, $\hat{\mathscr{O}}$ must make $p+1$ measurements of $h(x(t))$ – once to collect the approximate value of $J(v^k)$, then p more times to measure the variation of J along each coordinate axis in $\mathbb{R}^p$. For notational simplicity, we assume that the time between intra-iteration measurements remains constant, and equal to the inter-iteration interval T – i.e., for the kth iteration, the measurements occur at time instants $(t_{k,0}, t_{k,1}, \ldots, t_{k,p})$, where $t_{k,i} = (t_k + iT)^-$, $0 \le i \le p$. Whereas the ith component of the ideal search vector is $s_i(v^k) = -\alpha \nabla J(v^k)^T e_i$, the ith component of the actual search vector is

$$\hat{s}_i^k = \tfrac{-\alpha}{\mu}\left(h(l(v^k + \mu e_i) + z^{k,i}) - h(l(v^k) + z^k)\right), \tag{33}$$

where $\hat{s}_i^k = \hat{s}_i(v^k, z^{k,i}, z^k)$ and $z^{k,i} \triangleq x(t_{k,i}^-) - l(v^k + \mu e_i)$.

We now show that (33) can be modeled as the ideal $s_i(v^k)$, additively perturbed by p_i^k, which reflects errors due to both the FD approximation of $\nabla J(v^k)^T e_i$ and the transient dynamics of (31).

First we examine the error due to the FD approximation alone. If we expand $J(v^k + \mu e_i)$ using (4) from Lemma 2.1 as

$$\begin{aligned} J(v^k + \mu e_i) &= J(v^k) + \nabla J(v^k)^T (\mu e_i) \\ &\quad + \int_0^1 \left[\nabla J(v^k + \tau_3 \mu e_i) - \nabla J(v^k)\right]^T (\mu e_i)\, d\tau_1, \end{aligned}$$

then we see that the FD approximation of $\nabla J(v^k)^T e_i$ is

$$\tfrac{1}{\mu}\left(J(v^k + \mu e_i) - J(v^k)\right) = \nabla J(v^k)^T e_i + I_{i,1}, \tag{34}$$

where

$$I_{i,1} \triangleq \int_0^1 \left[\nabla J(v^k + \tau_1 \mu e_i) - \nabla J(v^k)\right]^T e_i\, d\tau_1. \tag{35}$$

Next, we focus on the transient errors in (33). Using (5) from Lemma 2.1 and assuming that h is continuously differentiable, we expand $h(l(v^k + \mu e_i) + z_i^k)$ about $l(v^k + \mu e_i)$, and $h(l(v^k) + z^k)$ about $l(v^k)$, to obtain

$$h(l(v^k + \mu e_i) + z^{k,i}) = h(l(v^k + \mu e_i)) + I_{i,2} \tag{36}$$

$$h(l(v^k) + z^k) = h(l(v^k)) + I_{i,3}, \tag{37}$$

where

$$I_{i,2} \triangleq \int_0^1 \nabla h(l(v^k + \mu e_i) + \tau_2 z^{k,i})^T z^{k,i}\, d\tau_2, \tag{38}$$

and

$$I_{i,3} \triangleq \int_0^1 \nabla h(l(v^k) + \tau_3 z^k)^T z^k \, d\tau_3. \tag{39}$$

First, we recognize that $h(l(v^k + \mu e_i)) \equiv J(v^k + \mu e_i)$ and that $h(l(v^k)) \equiv J(v^k)$. Then, subtracting (37) from (36), dividing by μ, and recalling (34), we obtain

$$\begin{aligned} \hat{s}_i^k &= -\alpha \left[\nabla J(v^k)^T e_i + I_{i,1} + \tfrac{1}{\mu} I_{i,2} - \tfrac{1}{\mu} I_{i,3} \right] \\ &= s_i(v^k) + p_i^k, \end{aligned} \tag{40}$$

where $s_i(v^k) = -\alpha \nabla J(v^k)^T e_i$ and

$$p_i^k \equiv p_i(v^k, z^{k,i}, z^k, \mu) \triangleq -\alpha \left(I_{i,1} + \tfrac{1}{\mu} I_{i,2} - \tfrac{1}{\mu} I_{i,3} \right). \tag{41}$$

Hence, we have shown that A2.4 is satisfied for this example. Similar techniques can be employed to model transient and other errors on the ideal search direction for other algorithms as well.

Finally, we must show that A2.5 holds for our choice of V_Σ and the perturbation $p^k \triangleq [p_1^k, \ldots, p_p^k]^T$. For simplicity, we suppose that the gradient of $h(x)$ is bounded (as would be the case if $h(x) = Hx$, $H \in \mathbb{R}^{1\times n}$) – i.e., there exists a number $b_{\nabla h} > 0$, such that for all $x \in \mathbb{R}^n$, $\|\nabla_x h(x)\| \leq b_{\nabla h}$. Then, from our definitions of $I_{i,1}$, $I_{i,2}$ and $I_{i,3}$ in (35), (38) and (39), we obtain

$$\begin{aligned} |p_i^k| &\leq \alpha |I_{i,1}| + \tfrac{\alpha}{\mu} |I_{i,2}| + \tfrac{\alpha}{\mu} |I_{i,2}| \\ &\leq \tfrac{\alpha L_{\nabla V_\mathcal{O}} \mu}{2} + \tfrac{\alpha}{\mu} b_{\nabla h} \|z^{k,i}\| + \tfrac{\alpha}{\mu} b_{\nabla h} \|z^k\|. \end{aligned} \tag{42}$$

We would like to express $\|z^{k,i}\|$ in terms of $\|z^k\|$. Let $\gamma = \max\{|\Re(\lambda)| : \det(\lambda I - A) = 0\}$. Then, by solving (31), we can show that for all $0 \leq i \leq p$,

$$z^{k,i} = e^{-\gamma i T} z^k + \sum_{j=1}^{i} e^{-\gamma j T} \left[l(v^k + \mu e_{i+1-j}) - l(v^k + \mu e_{i+2-j}) \right]$$

which, by the fact that l is Lipschitz with constant L_l, implies that

$$\|z^{k,i}\| \leq e^{-\gamma i T} \|z^k\| + L_l \sqrt{2} \mu \sum_{j=1}^{i} e^{-\gamma j T}.$$

Therefore, (42) becomes

$$|p_i^k| \leq K_1 + K_2 \|z^k\|, \tag{43}$$

where

$$K_1 = \tfrac{\alpha L_{\nabla V_\mathcal{O}} \mu}{2} + \alpha b_{\nabla h} L_l \sqrt{2} \sum_{j=1}^{i} e^{-\gamma j T}, \tag{44}$$

$$K_2 = \tfrac{\alpha b_{\nabla h}}{\mu} \left(e^{-\gamma i T} + 1 \right). \tag{45}$$

With further algebra we can show (from (43) and Young's inequality) that $\|p^k\|^2 = |p_1^k|^2 + \cdots + |p_p^k|^2$ can be expressed as in (10) – in terms of a vanishing component $\|p_v(z^k)\|^2$ and a nonvanishing component $\|p_o\|^2$. From (43), we notice that $\|p_v(z^k)\|^2$ will have the form $K_3 \|z^k\|^2$, for some $K_3 > 0$. An important feature of the expressions (44) and (45), is that at least one tunable parameter appears in each term - namely one of μ, α and T. This feature implies that K_3 can be made so small that A2.5 can be satisfied. Moreover, it is clear that p_o is likewise parameterized by these tunable parameters.

The analysis of this special case suggests that the interconnection assumptions that we propose can be satisfied under a reasonable set of conditions on the plant and optimizer.

V. CONCLUSIONS

We considered the problem of sampled-data extremum seeking control (ESC) from the point of view of interconnected systems theory. In contrast to [1], we examined a more structured class of optimizers, which led to the derivation of a different set of sufficient conditions for the stability of such schemes. Our motivation in deriving these new conditions is their natural applicability to the development and analysis of distributed extremum seeking schemes. A distinction between our work and [1] is that we have proposed modeling the dynamic coupling between the plant and optimizer as an additive perturbation on the optimizer's search vector. This approach enabled us to generalize the result in [1], by removing the assumptions on the measurements made by the optimizer.

REFERENCES

[1] A.R. Teel and D. Popović. Solving smooth and nonsmooth multivariable extremum seeking problems by the methods of nonlinear programming. In *Proceedings of the American Control Conference, VA, USA*, 2001.

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

[3] D. Popović, M. Janković, S. Manger, and A.R. Teel. Extremum seeking methods for optimization of variable cam timing engine operation. *IEEE Transactions on Control Systems Technology*, 14:398–407, 2006.

[4] P.M. Dower, P.M. Farrell, and D. Nešić. Extremum seeking control of cascaded raman optical amplifiers. *IEEE Transactions on Control Systems Technology*, 16:396–407, 2008.

[5] K.B. Ariyur and M. Krstić. *Real-time optimization by extremum-seeking control*. Wiley, 2003.

[6] C. Centioli, F. Iannone, G. Mezza, M. Panella, L. Pangione, S. Podda, A. Tuccillo, V. Vitale, and L. Zaccarian. Mazimization of the lower hybrid power coupline in the frascati tokamak upgrade via extremum seeking. *Control Engineering Practice*, 16:1468–1478, 2008.

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

[8] V. Adetola and M. Guay. Parameter convergence in adaptive extremum-seeking control. *Automatica*, 43:105–110, 2007.

[9] C. Zhang and R Ordó nez. Numerical optimization-based extremum seeking control with application to abs design. *IEEE Transactions on Automatic Control*, 52:454–467, 2007.

[10] J-Y. Choi, M. Krstić, K.B. Ariyur, and J.S. Lee. Extremum seeking control for discrete-time systems. *IEEE Transactions on Automatic Control*, 47:318–323, 2002.

[11] D.D. Šiljak. *Large-scale dynamic systems*. Elsevier North-Holland, 1978.

[12] B.T. Polyak. *Introduction to optimization*. Optimization Software, Inc., 1987.

[13] R.P. Agarwal. *Difference equations and inequalities - theory, methods and applications*. Marcel Dekker, 2000.

[14] V. Lakshmikantham and D. Trigiante. *Theory of difference equations*. Marcel Dekker, 2002.