Extremum Seeking for Static Maps With Delays

Tiago Roux Oliveira, *Member, IEEE*, Miroslav Krstić, *Fellow, IEEE*, and Daisuke Tsubakino, *Member, IEEE*

***Abstract*—In this paper, we address the design and analysis of multi-variable extremum seeking for static maps subject to arbitrarily long time delays. Both Gradient and Newton-based methods are considered. Multi-input systems with different time delays in each individual input channel as well as output delays are dealt with. The phase compensation of the dither signals and the inclusion of predictor feedback with a perturbation-based (averaging-based) estimate of the Hessian allow to obtain local exponential convergence results to a small neighborhood of the optimal point, even in the presence of delays. The stability analysis is carried out using backstepping transformation and averaging in infinite dimensions, capturing the infinite-dimensional state due the time delay. In particular, a new backstepping-like transformation is introduced to design the predictor for the Gradient-based extremum seeking scheme with multiple and distinct input delays. The proposed Newton-based extremum seeking approach removes the dependence of the convergence rate on the unknown Hessian of the nonlinear map to be optimized, being user-assignable as in the literature free of delays. A source seeking example illustrates the performance of the proposed delay-compensated extremum seeking schemes.**

***Index Terms*—Averaging in infinite dimensions, backstepping transformation, delay systems, Gradient and Newton-based extremum seeking (ES), predictor feedback, source seeking.**

I. Introduction

EXTREMUM seeking (ES) is a real-time, model-independent adaptive control technique for tuning parameters to optimize an unknown nonlinear map. The most popular ES approach relies on a small periodic excitation, usually sinusoidal, to disturb the parameters being tuned [1]–[7]. This approach quantifies the effects of the parameters on the output of the nonlinear map, then uses that information to generate the search of the optimal values.

In recent years, there have been a lot of advances in theory and applications of ES. This list includes the proof of local [1]–[3] or semi-global [6] stability properties of the search algorithm even in the presence of local extrema [7], its extension to the multi-variable case [8] and advances in parameter convergence and performance improvement [4], [5], [9], [11]. The book [13] also presents stochastic versions of the algorithm with filtered noise perturbation signals.

Despite the large number of publications on delay compensation via predictor feedback [14], to the best of our knowledge, there is no work in the literature which rigorously concerns ES in the presence of time delays. The motivation for this study is that there are applications in which post-processing of the plant's measured output translates into a considerable delay in generating the control input to be applied to the plant. Such is the case with the image processing that takes place in laser-based light sources for photolithography in semiconductor manufacturing [15], [16], or in various chemical and biochemical processes where analysis of samples takes place. For instance, the phase lag observed in batch cultures applied to bioreactors [17], [18] illustrates the delay phenomenon occurring in the biological optimizing process. These delays are typically known, constant, and relatively large.

In this paper, we propose a solution to the problem of designing multi-variable ES algorithms for delayed systems via predictors. Predictor feedback requires a known model. In our problem the model (and, most notably, the Hessian of the map) is unknown. So we present two approaches to construct a predictor based on perturbation-based estimates of the model. One of our approaches is based on gradient optimization where we estimate the Hessian [8], [20] for the purpose of implementing a predictor that compensates the delay. Our other approach is based on the Newton optimization where we estimate the Hessian's *inverse* for the purpose of making the convergence rate independent of the unknown parameters of the map and, as a bonus, obviate the predictor design because the average plant for which the predictor is designed becomes, essentially, an integrator with a known gain.

We employ a semi-model-based approach due to the treatment of the delay. While we assume a known delay, our predictor construction follows a model-free approach. We do not estimate the map parametrically as was done in [4]–[7], [9]–[12] for ES in the absence of delays—and, in particular, we do not estimate the Hessian parametrically. We estimate the Hessian with the same demodulating signal used to seek the extremum, but with a different additive perturbation. So we employ a predictor that is "partially model-based": the delay is known but the plant model is unknown and is estimated using perturbations.

In order to compensate multiple distinct delays in multivariable ES, we have had to develop an extension, from scratch, of the predictor feedback approach for the multi-input case with distinct delays. We developed an elegant and explicit solution to this problem through the introduction of a new successive-backstepping transformation [19].

Our paper also presents an innovation in the analysis methodology. Our analysis presents a carefully constructed sequence of analytical steps, a predictor-based infinite-dimensional

Manuscript received November 26, 2015; revised November 30, 2015 and April 14, 2016; accepted April 20, 2016. Date of publication May 9, 2016; date of current version March 27, 2017. T. R. Oliveira thanks CAPES, CNPq and FAPERJ for the financial support. Recommended by Associate Editor L. Zaccarian.

T. R. Oliveira is with Department of Electronics and Telecommunication Engineering, State University of Rio de Janeiro (UERJ), Rio de Janeiro, RJ 20550-900, Brazil (e-mail: tiagoroux@uerj.br).

M. Krstić is with Department of Mechanical and Aerospace Engineering, University of California-San Diego (UCSD), La Jolla, CA 92093-0411, USA (e-mail: krstic@ucsd.edu).

D. Tsubakino is with Department of Aerospace Engineering, Graduate School of Engineering, Nagoya University, Nagoya, 464-8603, Japan (e-mail: tsubakino@nuae.nagoya-u.ac.jp).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2016.2564958

backstepping transformation, a synthesist of a Lyapunov functional, and a computation of a Lyapunov estimate, for the overall infinite-dimensional system with nonlinearities, periodic perturbations, and distributed delays. The analysis process involving so many steps has a large number of possible permutations—all of which *but one* would be wrong. We show how to properly sequence the steps of averaging, backstepping, and Lyapunov functional analysis, to prove stability. This new "analysis pathway" will serve the needs of future researchers who deal with ES under delays.

Inspired by the reference [21], numerical results for an autonomous vehicle target tracking problem without position measurements and under actuator/sensor delays is portrayed.

Notation and Norms: The 2-norm of a finite-dimensional (ODE) state vector $X(t)$ is denoted by single bars, $|X(t)|$. In contrast, norms of functions (of x) are denoted by double bars. By default, $\|\cdot\|$ denotes the spatial $L_2[0,D]$ norm, i.e., $\|\cdot\| = \|\cdot\|_{L_2[0,D]}$. For example, the $L_2[0,D]$ norm of the PDE state variable $u(x,t)$ in $x\in[0,D]$ is $\|u(t)\| = \left(\int_0^D u^2(x,t)dx\right)^{1/2}$. Now, consider a generic nonlinear system $\dot{x} = f(t,x,\epsilon)$, where $x\in\mathbb{R}^n$, $f(t,x,\epsilon)$ is periodic in t with period T, i.e., $f(t+T,x,\epsilon) = f(t,x,\epsilon)$. Thus, for $\epsilon>0$ sufficiently small, we can obtain its average model given by $\dot{x}_{\rm av} = f_{\rm av}(x_{\rm av})$, with $f_{\rm av}(x_{\rm av}) = 1/T\int_0^T f(\tau,x_{\rm av},0)d\tau$, where $x_{\rm av}(t)$ denotes the average version of the state $x(t)$ [22]. As defined in [22], a vector function $f(t,\epsilon)\in\mathbb{R}^n$ is said to be of order $\mathcal{O}(\epsilon)$ over an interval $[t_1,t_2]$ if there exist positive constants k and ϵ^* such that $|f(t,\epsilon)|\le k\epsilon$, $\forall\,\epsilon\in[0,\epsilon^*]$ and $\forall\,t\in[t_1,t_2]$.

II. Problem Statement

Multi-parameter or multi-variable ES considers applications in which the goal is to maximize (or minimize) the scalar output $y\in\mathbb{R}$ of an unknown and convex nonlinear static map $y = Q(\theta)$ by varying the input vector $\theta = [\theta_1\ \theta_2\ ,\ \dots,\ \theta_n]^T$.

A. Basic Idea of ES Algorithms Free of Delays

In maximum seeking problem, there exists $\theta^*\in\mathbb{R}^n$ such that

$$\frac{\partial Q(\theta^*)}{\partial\theta} = 0 \tag{1}$$

$$\frac{\partial^2 Q(\theta^*)}{\partial\theta^2} = H < 0, \quad H = H^T \tag{2}$$

where θ^* and H are considered unknown. Referring to the Taylor series expansion of the nonlinear map around the peak θ^*, we have

$$y = y^* + \frac{1}{2}(\theta-\theta^*)^T H(\theta-\theta^*) + R(\theta-\theta^*) \tag{3}$$

where $R(\theta-\theta^*)$ stands for higher order terms in $\theta-\theta^*$ and $y^* = Q(\theta^*)$ is the extremum.

The basic Gradient ES algorithm measures the scalar signal $y(t)$ and with the help of the dither signals

$$S(t) = [a_1\sin(\omega_1 t)\ ,\ \dots,\ a_n\sin(\omega_n t)]^T \tag{4}$$

$$M(t) = \left[\frac{2}{a_1}\sin(\omega_1 t)\ ,\ \dots,\ \frac{2}{a_n}\sin(\omega_n t)\right]^T \tag{5}$$

construct $G(t) = M(t)y(t)$ to estimate the unknown gradient $\partial Q(\theta)/\partial\theta$ of the nonlinear map $Q(\theta)$. The actual input $\theta(t) := \hat{\theta}(t) + S(t)$ is based on the real-time estimate $\hat{\theta}(t)$ of θ^*, but is perturbed by $S(t)$. The estimate $\hat{\theta}$ is generated with the integrator $\hat{\theta} = (K/s)G$ which locally approximates the gradient update law $\dot{\hat{\theta}}(t) = KH(\hat{\theta}(t)-\theta^*)$, tuning $\hat{\theta}(t)$ to θ^*, if $\hat{\theta}(0)$ is close of θ^*. The adaptation gain (diagonal matrix $K>0$) controls the speed of estimation, but it cannot be arbitrarily increased *a priori* due to the limitations on the averaging analysis [1].

To guarantee convergence, the user should choose appropriate frequencies $\omega_i\neq\omega_j$ and nonzero small amplitudes a_i. The former is a key condition that differentiates the multi-input case [8] from the single-input case [1]. The sinusoid feature of (4) and (5) is only one choice for the dither signals—many other perturbations, from square waves to stochastic noise, can be used in lieu of it, provided they are of zero mean [13], [23].

However, the convergence rate depends on the unknown Hessian H. This weakness of the Gradient-based ES algorithm is removed with the Newton-based ES and will be detailed later on in the paper. Briefly, a multiplicative excitation denoted by $N(t)$ is introduced to generate the estimate of the Hessian H as $\hat{H}(t) = N(t)y(t)$ [20]. According to [8], then a Riccati differential equation inverts this Hessian's estimate and cancels out the term H from the convergence rate, making it user-assignable. A fair overview of Gradient- and Newton-based versions of ES free of delays can be found in [3].

B. Input-Output Delays

The main contribution of the present paper is to additionally consider that the nonlinear map to be optimized in real time is subject to delays in the actuator path and/or measurement system. In order to start to formulate the problem, we can initially assume the simplest case of delays in the scalar output. In this case, there exists a constant delay denoted by $D_{\rm out}\ge 0$ in the measurement system such that the output is given by

$$y(t) = Q\left(\theta(t-D_{\rm out})\right). \tag{6}$$

However, it is not difficult to show that input delays could be handled in the same way noting that input delays, denoted by $D_{\rm in}\ge 0$, can be moved to the output of the static map. The restriction here is that the delay must be the same in each individual input channel. The scenario when input and output delays occur simultaneously could also be coped with assuming that the total delay to be counteract would be $D_{\rm total} = D_{\rm in} + D_{\rm out}$.

In a more general framework, we can assume the following input-output delay representation

$$y(t) = Q\left(\theta(t-D)\right) = e^{-Ds}\left[Q\left(\theta(t)\right)\right] \tag{7}$$

where the constant delay matrix $D = {\rm diag}\{D_1,\dots,D_n\}$ must have the same component delays D_i in each individual input channel of $\theta(t)\in\mathbb{R}^n$, i.e., $D_1\equiv D_2\equiv\cdots\equiv D_n\ge 0$ for non distinct input delays (or output delays). Nevertheless, the main advantage of this representation is the possibility of including multiple and distinct input delays. In this case, without loss of generality, we assume that the inputs have distinct delays which are ordered so that

$$D = {\rm diag}\{D_1,\dots,D_n\}, \quad 0\le D_1\le D_2\le\cdots\le D_n. \tag{8}$$

In addition, we consider that the constants D_i must be *known* for all $i\in\{1,2,\dots,n\}$. We mix the time and frequency domains in (7) by using the brackets $[\cdot]$ to denote that the transfer function acts as an operator on a time-domain function.

C. Locally Quadratic Maps With Delays

Without loss of generality, let us consider the maximum seeking problem such that the maximizing value of θ is denoted by θ^*, satisfying (1) and (2). For the sake of simplicity, we assume that the nonlinear map (7) is at least locally quadratic

$$Q(\theta) = y^* + \frac{1}{2}(\theta - \theta^*)^T H(\theta - \theta^*) \tag{9}$$

within a neighborhood of the unknown extremum point (θ^*, y^*), where $\theta^* \in \mathbb{R}^n$, $y^* \in \mathbb{R}$ and $H = H^T < 0$ is the $n \times n$ unknown Hessian matrix of this static map.

By plugging (9) into (7), we obtain the *locally quadratic static map with delay*:

$$y(t) = y^* + \frac{1}{2}\left(\theta(t-D) - \theta^*\right)^T H \left(\theta(t-D) - \theta^*\right). \tag{10}$$

In the general case of multiple and distinct delays in the control channels, the delayed input vector can be represented by

$$\theta(t-D) := \begin{bmatrix} \theta_1(t-D_1) \\ \theta_2(t-D_2) \\ \vdots \\ \theta_n(t-D_n) \end{bmatrix}. \tag{11}$$

D. System and Signals

Let $\hat{\theta}$ be the estimate of θ^* and

$$\tilde{\theta}(t) = \hat{\theta}(t) - \theta^* \tag{12}$$

be the *estimation error*. Moreover, let us define

$$G(t) = M(t)y(t), \quad \theta(t) = \hat{\theta}(t) + S(t) \tag{13}$$

where the vector dither signals are given by

$$S(t) = \left[a_1 \sin\left(\omega_1(t+D_1)\right), \ldots, \; a_n \sin\left(\omega_n(t+D_n)\right)\right]^T \tag{14}$$

$$M(t) = \left[\frac{2}{a_1}\sin(\omega_1 t), \ldots, \; \frac{2}{a_n}\sin(\omega_n t)\right]^T \tag{15}$$

with nonzero perturbation amplitudes a_i.

The elements of the $n \times n$ demodulating matrix $N(t)$ to construct the signal

$$\hat{H}(t) = N(t)y(t) \tag{16}$$

are given by:

$$N_{i,i}(t) = \frac{16}{a_i^2}\left(\sin^2(\omega_i t) - \frac{1}{2}\right) \tag{17}$$

$$N_{i,j}(t) = \frac{4}{a_i a_j}\sin(\omega_i t)\sin(\omega_j t), \quad i \neq j. \tag{18}$$

The probing frequencies ω_i's can be selected as

$$\omega_i = \omega_i' \omega = \mathcal{O}(\omega), \quad i \in 1, 2, \ldots, n \tag{19}$$

where ω is a positive constant and ω_i' is a rational number. One possible choice is given in [8] as

$$\omega_i' \notin \left\{\omega_j', \; \frac{1}{2}\left(\omega_j' + \omega_k'\right), \; \omega_j' + 2\omega_k', \; \omega_j' + \omega_k' \pm \omega_l'\right\} \tag{20}$$

for all distinct i, j, k and l.

Notice that (14) is different from the additive dither signal (4) used in standard ES algorithms free of delays. As it will be shown, the input/output delays can always be transferred for analysis purposes to the integrator output of the estimation error dynamics (or, equivalently, to its input), thus, the phase shift $+\omega_i D_i$ is applied to compensate the delay effect in the dither signal $S(t)$.

E. Basic Averaging Properties

In [8], the following two averaging properties were proved:

$$\frac{1}{\Pi}\int_0^{\Pi} N(\sigma) y d\sigma = H \tag{21}$$

$$\frac{1}{\Pi}\int_0^{\Pi} M(\sigma) y d\sigma = H\tilde{\theta}_{\rm av} \tag{22}$$

if a quadratic map as in (9) is considered, which are still valid even in the presence of delays. In other words, we obtain the average signals

$$\hat{H}_{\rm av} = (Ny)_{\rm av} = H \tag{23}$$

$$G_{\rm av}(t) = (My)_{\rm av} = H\tilde{\theta}_{\rm av}(t-D) \tag{24}$$

for $\hat{H}(t)$ and $G(t)$ with

$$\tilde{\theta}_{\rm av}(t-D) = \left[\tilde{\theta}_1^{\rm av}(t-D_1), \ldots, \; \tilde{\theta}_n^{\rm av}(t-D_n)\right]^T \tag{25}$$

and Π defined as

$$\Pi = 2\pi \times \text{LCM}\left\{\frac{1}{\omega_i}\right\}, \quad \forall i \in \{1, 2, \ldots, n\} \tag{26}$$

where LCM stands for the least common multiple.

Basically, we can say that the signals $\hat{H}(t)$ and $G(t)$ provide averaging-based estimates for the Hessian and the delayed Gradient of the nonlinear map (9) and (10). These properties are fundamental for the predictor design and stability analysis of the Gradient and Newton-based ES schemes developed in the next sections.

III. Gradient-Based ES With Output Delays

For the sake of clarity, we start considering output delays (or same delay in the input channels) for Gradient ES before moving on to the more complex case of multiple and distinct input delays. Therefore, the variables D_i in (10) and (11) are equal along of this section and $D > 0$ can be faced as a single *known* constant. The proposed Gradient ES under output delays is shown in the block diagram of Fig. 1.

A. Averaging Analysis Without Predictor Compensation

If the filtered predictor based controller was not applied in Fig. 1, but the standard Gradient ES [1] feedback law $U(t) = KG(t)$, one could write $\dot{\hat{\theta}}(t) = \dot{\tilde{\theta}}(t) = KG(t)$, where $K > 0$ is a $n \times n$ positive diagonal matrix. From (13), the closed-loop system equation would be written as:

$$\dot{\tilde{\theta}}(t) = KM(t)y(t) \tag{27}$$

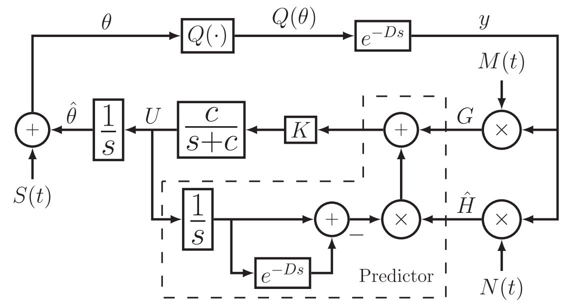


Fig. 1. Block diagram of the basic prediction scheme for output-delay compensation in multi-variable Gradient-based ES, where the delay $D \geq 0$ is a simple scalar. The predictor with a perturbation-based estimate of the Hessian obeys equation (38), and the additive-multiplicative dither signals are given by: $S(t) = [a_1 \sin(\omega_1(t+D)), \ldots, a_n \sin(\omega_n(t+D))]^T$ and $M(t) = [(2/a_1)\sin(\omega_1 t), \ldots, (2/a_n)\sin(\omega_n t)]^T$. The demodulating matrix $N(t)$ is computed by (17) and (18).

and using the identity (22) to average (27), we would obtain the following average model:

$$\frac{d\tilde{\theta}_{\rm av}(t)}{dt} = KH\tilde{\theta}_{\rm av}(t-D). \tag{28}$$

From (28), it is clear that the equilibrium $\tilde{\theta}^e_{\rm av} = 0$ of the average system is not necessarily stable for arbitrary values of the delay D. This reinforces the necessity of applying the prediction $U(t) = KG(t+D)$, $\forall t \geq 0$, to stabilize the system.

B. Predictor Feedback With a Perturbation-Based Estimate of the Hessian

From Fig. 1, the *error dynamics* of (12) is written as

$$\dot{\tilde{\theta}}(t-D) = U(t-D). \tag{29}$$

The average version of the vector signal (13) is given by (24). Hence, from (29), the following average models can be obtained

$$\dot{\tilde{\theta}}_{\rm av}(t-D) = U_{\rm av}(t-D) \tag{30}$$
$$\dot{G}_{\rm av}(t) = HU_{\rm av}(t-D) \tag{31}$$

where $U_{\rm av} \in \mathbb{R}^n$ denotes the resulting average control for $U \in \mathbb{R}^n$. Given the stabilizing diagonal matrix $K > 0$ for the undelayed system, our wish is to have a controller that achieves

$$U_{\rm av}(t) = KG_{\rm av}(t+D), \quad \forall t \geq 0 \tag{32}$$

and it appears to be nonimplementable since it requires future values of the state. However, by applying the variation of constants formula to (31) we can express the future state as

$$G_{\rm av}(t+D) = G_{\rm av}(t) + H\int_t^{t+D} U_{\rm av}(\tau - D)d\tau \tag{33}$$

where the current state $G_{\rm av}(t)$ is the initial condition. Shifting the time variable under the integral in (33), we obtain

$$G_{\rm av}(t+D) = G_{\rm av}(t) + H\int_{t-D}^{t} U_{\rm av}(\tau)d\tau \tag{34}$$

which gives the future state $G_{\rm av}(t+D)$ in terms of the average control signal $U_{\rm av}(\tau)$ from the past window $[t-D, t]$. It yields the following feedback law

$$U_{\rm av}(t) = K\left[G_{\rm av}(t) + H\int_{t-D}^{t} U_{\rm av}(\tau)d\tau\right]. \tag{35}$$

Hence, from (34) and (35), the average feedback law (32) can be obtained indeed as desired. Consequently,

$$\dot{\tilde{\theta}}_{\rm av}(t) = KG_{\rm av}(t+D), \quad \forall t \geq 0. \tag{36}$$

Therefore, from (24), one has

$$\frac{d\tilde{\theta}_{\rm av}(t)}{dt} = KH\tilde{\theta}_{\rm av}(t), \quad \forall t \geq D \tag{37}$$

with an exponentially attractive equilibrium $\tilde{\theta}^e_{\rm av} = 0$, since $KH < 0$. It means that the delay is perfectly compensated in D seconds, namely, the system evolves as if the delay were absent after D seconds.

The feedback law (35) seems to be implicit since $U_{\rm av}$ is present on both sides. However, the input memory $U_{\rm av}(\tau)$, where $\tau \in [t-D, t]$, is part of the state of an infinite-dimensional system, and thus the control law is effectively a complete-state-feedback controller. However, the analysis sketched above in the spirit of "finite spectrum assignment" does not capture the entire cascade system consisting of the ODE in (29) and the infinite-dimensional (PDE) subsystem of the input delay.

Another difficulty arises in the application of *Averaging Theorem* to infinite dimensions (see Appendix). For the class of functional differential equations (FDEs) studied here, there is no "off the shelf" averaging theorem result oriented for input-output delays. In general, the theory applies only to state-delay systems. This fact lead us to propose a simple modification of the basic predictor-based controllers, which employs a low-pass filter [24], to achieve our control objectives.

Thus, the averaging-based predictor feedback used in order to compensate output delays is redefined by

$$U(t) = \frac{c}{s+c}\left\{K\left[G(t) + \hat{H}(t)\int_{t-D}^{t} U(\tau)d\tau\right]\right\} \tag{38}$$

where $c > 0$ is sufficiently large. The predictor feedback is of the form of a low-pass filtered of the non-average version of (35). With some abuse of notation, now we mix again the time and frequency domains in (38) by using the braces $\{\cdot\}$ to denote that the lag transfer function acts as an operator on a time-domain function.

The predictor (38) is infinite-dimensional because the integral involves the control history over the interval $[t-D, t]$. Furthermore, it is averaging-base (perturbation-based) because $\hat{H}$ is updated according to the estimate (16) of the unknown Hessian H, satisfying the averaging property (21).

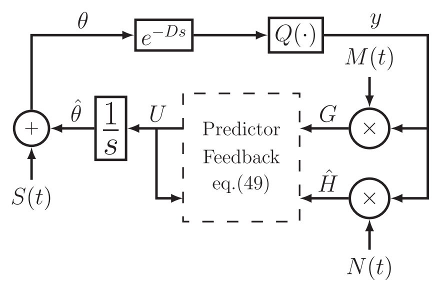


Fig. 2. Block diagram of Gradient-based ES for multiple-input delay compensation, where $D = \text{diag}\{D_1, \ldots, D_n\}$. The predictor feedback is implemented according to (49). The additive dither signal is $S(t) = [a_1 \sin(\omega_1(t + D_1)) \ , \ldots, \ a_n \sin(\omega_n(t + D_n))]^T$ and the demodulating signals $M(t)$ and $N(t)$ are given by $M(t) = [(2/a_1)\sin(\omega_1 t) \ , \ldots, \ (2/a_n)\sin(\omega_n t)]^T$ and (17), (18), respectively.

IV. Gradient-Based ES With Multiple and Distinct Input Delays

Now, let us consider the case where the input delays D_i are distinct and satisfy (8). From Fig. 2 and (12), we can easily write

$$\dot{\hat{\theta}}(t-D) = \begin{bmatrix} U_1(t-D_1) \\ U_2(t-D_2) \\ \vdots \\ U_n(t-D_n) \end{bmatrix}, \quad \dot{\hat{\theta}}_i(t-D_i) = U_i(t-D_i) \tag{39}$$

and the average model below by using (24):

$$\dot{G}_{\text{av}}(t) = \sum_{i=1}^{n} H_i U_i^{\text{av}}(t-D_i) = H \begin{bmatrix} U_1^{\text{av}}(t-D_1) \\ U_2^{\text{av}}(t-D_2) \\ \vdots \\ U_n^{\text{av}}(t-D_n) \end{bmatrix} \tag{40}$$

since

$$\dot{\hat{\theta}}_{\text{av}}(t-D) = \begin{bmatrix} U_1^{\text{av}}(t-D_1) \\ U_2^{\text{av}}(t-D_2) \\ \vdots \\ U_n^{\text{av}}(t-D_n) \end{bmatrix}. \tag{41}$$

In (40), $U_i^{\text{av}}(t) \in \mathbb{R}$ for all $i \in \{1, 2, \ldots, n\}$ are the elements of the average control $U_{\text{av}}(t) \in \mathbb{R}^n$ and the Hessian matrix $H = (H_1, H_2, \ldots, H_n) \in \mathbb{R}^{n\times n}$. In this case, there always exists a positive diagonal matrix $K = (K_1, K_2, \ldots, K_n)^T \in \mathbb{R}^{n\times n}$ such that HK is Hurwitz. For the sake of clarity, we say that H_i are column vectors of H and K_i^T are row vectors of $K, i = 1, \ldots, n$.

Henceforth, the purpose of this section is to find a control feedback which has to perform prediction of the cross-coupling of the channels in (40). By applying the variation of constants formula to (40) gives

$$G_{\text{av}}(t+D_i) = G_{\text{av}}(t) + \sum_{j=1}^{n} \int_{t-D_j}^{t-(D_j-D_i)} H_j U_j^{\text{av}}(\tau) d\tau \tag{42}$$

such that the predictor feedback realizes

$$U_i^{\text{av}}(t) = K_i^T G_{\text{av}}(t+D_i). \tag{43}$$

From (8), each control input U_i^{av} arrives at the plant at a different time. Since the system is causal, values of U_j^{av} with $j > i$ on the interval $(t-(D_j-D_i), t)$ provide no information for prediction of $G_{\text{av}}(t+D_i)$. On the other hand, values of U_j^{av} with $j < i$ on the interval $(t, t+D_i-D_j)$ are necessary to calculate (42), but they are unavailable future values. For $i, j \in \{1, \ldots, n\}$ with $i < j$, we have

$$\int_{t-D_j}^{t-(D_j-D_i)} (*) d\tau = \int_{t-D_j}^{t} (*) d\tau + \int_{t}^{t+D_i-D_j} (*) d\tau. \tag{44}$$

Differently from the results in Section III, the variation of constants formula contains an extra term $\int_t^{t+D_i-D_j} (*) d\tau$.

In what follows, we will provide a predictor based controller that is more consistent with (42). To derive it, we have had to extend the backstepping approach [14] introducing a new successive backstepping-like transformation [19]. The explicit equation of this transformation of the delay state is given in the proof of the main theorem.

First of all, let us define the following notation:

$$A_i := \sum_{j=1}^{i} H_j K_j^T, \quad i \in \{0, 1, 2, \ldots, n\} \tag{45}$$

being obvious that $A_0 = 0_{n\times n}$ and $A_n = HK$. In addition, the matrix-valued function Φ can be represented as [19]

$$\Phi(x,\zeta) = e^{A_{i-1}(x-D_{i-1})} e^{A_{i-2}(D_{i-1}-D_{i-2})} \cdots e^{A_{j-1}(D_j-\zeta)}, \quad D_{i-1} \le x < D_i, \quad D_{j-1} \le \zeta < D_j \tag{46}$$

for any $i, j \in \{1, 2, \ldots, n\}$ satisfying $i > j$, and

$$\Phi(x,\zeta) = e^{A_{i-1}(x-\zeta)}, \quad D_{i-1} \le \zeta \le x \le D_i \tag{47}$$

$\forall\, i \in \{1, 2, \ldots, n\}$, where we need to treat D_0 as 0.

In a few words, Φ can be seen as the state-transition matrix of the time varying system $\dot{X}(t) = A(t)X(t), \ \forall\, t \ge 0$, where $A(t) \in \mathbb{R}^{n\times n}$ is a piecewise constant function defined by

$$A(t) = \begin{cases} 0_{n\times n}, & t \le D_1, \\ A_i, & D_i < t \le D_{i+1}, \quad i = 1, \ldots, n-1, \\ A_n, & t > D_n. \end{cases} \tag{48}$$

As it will be shown in the next section, the following predictor-based controller

$$U_i(t) = \frac{c}{s+c}\Bigg\{ K_i^T \hat{\Phi}(D_i, 0) G(t) + K_i^T \Bigg(\sum_{j=1}^{i} \int_{t-D_j}^{t} \hat{\Phi}(D_i, \tau - t + D_j) \hat{H}_j(t) U_j(\tau) d\tau + \sum_{j=i+1}^{n} \int_{t-D_i}^{t} \hat{\Phi}(D_i, \tau - t + D_i) \hat{H}_j(t) U_j(\tau - (D_j - D_i)) d\tau \Bigg) \Bigg\} \tag{49}$$

guarantees exponential stability for the closed-loop system, with $\hat{\Phi}$ defined as in (46) and (47) but replacing A_i in (45) by $\hat{A}_i(t) := \sum_{j=1}^{i} \hat{H}_j(t)K_j^T$, the vector signal $G(t)$ given in (13) and $\hat{H}_j$ being the columns of the Hessian estimate $\hat{H} = (\hat{H}_1, \hat{H}_2, \ldots, \hat{H}_n) \in \mathbb{R}^{n\times n}$ given by (16). For ω in (19) sufficiently large and a_i in (14)–(16) sufficiently small, the average version of the predictor-feedback form (49) can be numerically approximated by

$$U_i^{\rm av}(t) = \frac{c}{s+c}\Bigg\{K_i^T\Phi(D_i,0)G_{\rm av}(t) + K_i^T\left(\sum_{j=1}^{i}\int_{t-D_j}^{t}\Phi(D_i,\tau-t+D_j)H_jU_j^{\rm av}(\tau)d\tau + \sum_{j=i+1}^{n}\int_{t-D_i}^{t}\Phi(D_i,\tau-t+D_i)H_jU_j^{\rm av}\left(\tau-(D_j-D_i)\right)d\tau\right)\Bigg\}. \quad (50)$$

From (48), it is possible to show the term between braces in (50) by itself, if applied to (40), is enough to conclude that the closed-loop system $\dot{G}_{\rm av}(t) = A_nG_{\rm av}(t)$ is totally delay-compensated $\forall\, t \geq D_n$, since $A_n = HK$ is Hurwitz. However, due to technical limitations involving averaging results in infinite dimensions, we must include a low-pass filter $c/(s+c)$ in the predictor control loop, as was done in (49) and (50).

V. Stability Analysis for Gradient-Based ES Under Time Delays

The exponential stability estimate in L_2-norm of the closed-loop infinite-dimensional system is stated in the next theorem for Gradient ES subject to multiple and distinct input delays.

Theorem 1 (Multiple and Distinct Input Delays): Consider the control system in Fig. 2 with multiple and distinct input delays according to (7)–(11) and locally quadratic nonlinear map (9). There exists $c^* > 0$ such that, $\forall\, c \geq c^*$, $\exists \omega^*(c) > 0$ such that, $\forall\, \omega > \omega^*$, the closed-loop delayed system (39) and (49) with state $\tilde{\theta}_i(t-D_i)$, $U_i(\tau)$, $\forall\, \tau \in [t-D_i, t]$ and $\forall\, i \in \{1, 2, \ldots, n\}$, has a unique locally exponentially stable periodic solution in t of period Π, denoted by $\tilde{\theta}_i^{\Pi}(t-D_i)$, $U_i^{\Pi}(\tau)$, $\forall\, \tau \in [t-D_i, t]$, satisfying, $\forall\, t \geq 0$:

$$\left(\sum_{i=1}^{n}\left[\tilde{\theta}_i^{\Pi}(t-D_i)\right]^2 + \left[U_i^{\Pi}(t)\right]^2 + \int_{t-D_i}^{t}\left[U_i^{\Pi}(\tau)\right]^2 d\tau\right)^{\frac{1}{2}} \leq \mathcal{O}\left(\frac{1}{\omega}\right). \quad (51)$$

Furthermore,

$$\limsup_{t\to+\infty} |\theta(t)-\theta^*| = \mathcal{O}\left(|a| + 1/\omega\right) \quad (52)$$

$$\limsup_{t\to+\infty} |y(t)-y^*| = \mathcal{O}\left(|a|^2 + 1/\omega^2\right) \quad (53)$$

where $a = [a_1,\ a_2\ , \ldots,\ a_n]^T$.

Proof: The demonstration follows the **Steps 1** to **7**.

Step 1: *Transport PDE for Delay Representation*

Each individual delay D_i in equation (39) can be represented using a transport PDE as

$$\dot{\hat{\theta}}_i(t-D_i) = u_i(0,t) \quad (54)$$

$$\partial_t u_i(x,t) = \partial_x u_i(x,t), \quad x \in [0, D_i] \quad (55)$$

$$u_i(D_i,t) = U_i(t), \quad i = 1, 2, \ldots, n \quad (56)$$

where the solution of (55), (56) is

$$u_i(x,t) = U_i(t+x-D_i) \quad (57)$$

and $u(x,t) = [u_1(x,t), \ldots, u_n(x,t)]^T$ is the state of the total delay infinite-dimensional subsystem.

Step 2: *Average Model of the Closed-loop System*

From (54)–(56), we can rewrite (40) as

$$\dot{G}_{\rm av}(t) = \sum_{i=1}^{n} H_iu_i^{\rm av}(0,t) \quad (58)$$

$$\partial_t u_i^{\rm av}(x,t) = \partial_x u_i^{\rm av}(x,t), \quad x \in [0, D_i] \quad (59)$$

$$u_i^{\rm av}(D_i,t) = U_i^{\rm av}(t), \quad i = 1, 2, \ldots, n \quad (60)$$

where the solution of (59), (60) is

$$u_i^{\rm av}(x,t) = U_i^{\rm av}(t+x-D_i) \quad (61)$$

and the PDE state is $u_{\rm av}(x,t) = [u_1^{\rm av}(x,t), \ldots, u_n^{\rm av}(x,t)]^T$.

By representing the integrand in (50) using the transport PDE state, one has the average control law

$$U_i^{\rm av}(t) = \frac{c}{s+c}\left\{K_i^T\left(\Phi(D_i,0)G_{\rm av}(t) + \sum_{j=1}^{n}\int_0^{\phi_j(D_i)}\Phi(D_i,\sigma)H_ju_j^{\rm av}(\sigma,t)d\sigma\right)\right\} \quad (62)$$

with $\phi_j : [0, D_n] \to [0, D_j],\ j \in \{1, 2, \ldots, n\}$ being the function defined by

$$\phi_j(x) = \begin{cases} x, & 0 \leq x \leq D_j \\ D_j, & D_j < x < D_n. \end{cases} \quad (63)$$

Finally, substituting (62) into (60), we have

$$\dot{G}_{\rm av}(t) = \sum_{i=1}^{n} H_iu_i^{\rm av}(0,t) \quad (64)$$

$$\partial_t u_i^{\rm av}(x,t) = \partial_x u_i^{\rm av}(x,t), \quad x \in [0, D_i] \quad (65)$$

$$\frac{d}{dt}u_i^{\rm av}(D_i,t) = -cu_i^{\rm av}(D_i,t) + cK_i^T\left(\Phi(D_i,0)G_{\rm av}(t) + \sum_{j=1}^{n}\int_0^{\phi_j(D_i)}\Phi(D_i,\sigma)H_ju_j^{\rm av}(\sigma,t)d\sigma\right). \quad (66)$$

Step 3: *Successive Backstepping-like transformation, its inverse and the target system*

Consider the new infinite-dimensional backstepping-like transformation [19] of the delay state

$$w_i(x,t) = u_i^{\mathrm{av}}(x,t) - K_i^T\left(\Phi(x,0)G_{\mathrm{av}}(t)+\sum_{j=1}^{n}\int_0^{\phi_j(x)}\Phi(x,\sigma)H_j u_j^{\mathrm{av}}(\sigma,t)d\sigma\right) \quad (67)$$

which maps the system (64)–(66) into the target system:

$$\dot{G}_{\mathrm{av}}(t) = A_n G_{\mathrm{av}}(t) + \sum_{i=1}^{n} H_i w_i(0,t) \quad (68)$$

$$\partial_t w_i(x,t) = \partial_x w_i(x,t) - \sum_{j=1}^{i-1}\lambda_{ij}(x)w_j(D_j,t),\ x\in[0,D_i] \quad (69)$$

$$w_i(D_i,t) = -\frac{1}{c}\partial_t u_i^{\mathrm{av}}(D_i,t), \quad i=1,2,\ldots,n \quad (70)$$

where $A_n = HK$ and the coefficients $\lambda_{ij}:[0,D_i]\to\mathbb{R}$ are

$$\lambda_{ij}(x) = \begin{cases}0, & 0\le x\le D_j\\ K_i^T\Phi(x,D_j)H_j, & D_j < x \le D_i.\end{cases} \quad (71)$$

Note that the PDE for w_i is not a simple transport equation unless w_i vanishes at the right boundary. Using (67) for $x = D_i$ and the fact that $u_i^{\mathrm{av}}(D_i,t) = U_i^{\mathrm{av}}(t)$, we can directly obtain (66) and (62) from (70).

On the other hand, the inverse of (67) is given by

$$u_i^{\mathrm{av}}(x,t) = w_i(x,t) + K_i^T\left(e^{A_n x}G_{\mathrm{av}}(t) + \sum_{j=1}^{n}\int_0^{\phi_j(x)} e^{A_n(x-\sigma)}H_j w_j(\sigma,t)d\sigma\right). \quad (72)$$

For later use, we find an expression for $\partial_t w_i(D_i,t)$. Differentiating (72) with respect to $x\in(D_{i-1},D_i)$, gives

$$\partial_x u_i^{\mathrm{av}}(x,t) = \partial_x w_i(x,t) + \sum_{j=i}^{n} K_i^T H_j w_j(x,t) + K_i^T A_n\left(e^{A_n x}G_{\mathrm{av}}(t)+\sum_{j=1}^{n}\int_0^{\phi_j(x)} e^{A_n(x-\sigma)}H_j w_j(\sigma,t)dy\right). \quad (73)$$

In light of (65), (66) and (69)–(71), we arrive at

$$\begin{aligned}\partial_t w_i(D_i,t) = &-c w_i(D_i,t) - \sum_{j=1}^{i} K_i^T\Phi(D_i,D_j)H_j w_j(D_j,t)\\ &- \sum_{j=i+1}^{n} K_i^T H_j w_j(D_i,t) - \gamma_i(0)^T G_{\mathrm{av}}(t)\\ &- \sum_{j=1}^{n}\int_0^{\phi_j(D_i)}\gamma_i(\sigma)^T H_j w_j(\sigma,t)d\sigma \end{aligned} \quad (74)$$

where $\gamma_i(x) := e^{A_n^T(D_i-x)}A_n^T K_i$ for each $i\in\{1,2,\ldots,n\}$.

Note that the right-hand side contains $w_j(D_i,t)$ for each j greater than i, which is not a boundary value of w_j. For this reason, a key feature of the Lyapunov functional is the necessity of breaking the domain of integration for the terms $(1+x)w_i(x,t)^2$, as shown in the next step.

Step 4: *Lyapunov-Krasovskii Functional*

Let V be the candidate of Lyapunov function defined by

$$V(t) = G_{\mathrm{av}}(t)^T P G_{\mathrm{av}}(t) + \sum_{i=1}^{n}\sum_{j=1}^{i}\frac{\bar{a}_j}{2}\int_{D_{j-1}}^{D_j}(1+x)w_i(x,t)^2dx + \frac{1}{2}\sum_{i=1}^{n}w_i(D_i,t)^2 \quad (75)$$

where $P=P^T\in\mathbb{R}^{n\times n}$ is the solution of the Lyapunov equation $PA_n + A_n^T P = -Q$ for some $Q = Q^T > 0$. The real constant $\bar{a}_1 = 4\lambda_{\max}(PHH^TP)/\lambda_{\min}(Q)$. The other constants $\bar{a}_2,\ldots,\bar{a}_n$ are arbitrary real numbers satisfying $\bar{a}_1 < \bar{a}_2 < \cdots < \bar{a}_n$. To shorten notation, we define a function $w:[0,1]\times[0,+\infty)\to\mathbb{R}^n$ by

$$w(\xi,t) = (w_1(D_1\xi,t),\ w_2(D_2\xi,t),\ldots,\ w_n(D_n\xi,t))^T \quad (76)$$

for $0\le\xi\le 1$. In addition, we omit the dependence on the temporal variable t for simplicity. For instance, we write V and $w_i(x)$ instead of $V(t)$ and $w_i(x,t)$. Differentiating V with respect to t yields

$$\begin{aligned}\dot{V} = &-G_{\mathrm{av}}^T Q G_{\mathrm{av}} + 2G_{\mathrm{av}}^T PHw(0)\\ &+\sum_{i=1}^{n}\sum_{j=1}^{i}\frac{\bar{a}_j}{2}\int_{D_{j-1}}^{D_j}(1+x)2w_i(x)\partial_t w_i(x)dx\\ &+ w(1)^T\partial_t w(1)\\ = &-G_{\mathrm{av}}^T Q G_{\mathrm{av}} + 2G_{\mathrm{av}}^T PHw(0) - \frac{\bar{a}_1}{2}w(0)^T w(0)\\ &-\sum_{i=2}^{n}\sum_{j=1}^{i-1}\frac{\alpha_j}{2}w_i(D_j)^2 + w(1)^T\partial_t w(1)\\ &-\sum_{i=1}^{n}\sum_{\ell=1}^{i-1}\sum_{j=\ell+1}^{i}\bar{a}_j w_\ell(D_\ell)\\ &\times\int_{D_{j-1}}^{D_j}(1+x)K_i^T\Phi(x,D_\ell)H_\ell w_i(x)dx\\ &+\frac{1}{2}w(1)^T\Delta w(1) - \sum_{i=1}^{n}\sum_{j=i}^{n}\frac{\bar{a}_i}{2}\int_{D_{i-1}}^{D_i}w_j(x)^2dx \end{aligned} \quad (77)$$

where $\alpha_j > 0$ and $\Delta\in\mathbb{R}^{n\times n}$ are defined by

$$\alpha_j = (\bar{a}_{j+1}-\bar{a}_j)(1+D_j), \quad j\in\{1,2,\ldots,n-1\} \quad (78)$$

$$\Delta = \mathrm{diag}\,\{\bar{a}_1(1+D_1),\bar{a}_2(1+D_2),\ldots,\bar{a}_n(1+D_n)\}. \quad (79)$$

After lengthy calculations omitted here due to space limitations, we can conclude

$$\dot{V} \le -\frac{1}{4}G_{\mathrm{av}}^T Q G_{\mathrm{av}} - \frac{\bar{a}_1}{4}\sum_{i=1}^{n}\int_0^{D_i}w_i(x)^2dx - w(1)^T(cI_{n\times n}+\bar{R})w(1) \quad (80)$$

where $\bar{R}$ is an appropriate matrix with elements depending on H, K and D. Hence, if $c > \lambda_{\min}(R)$, there exists $\mu > 0$ such that

$$\dot{V} \leq -\mu V. \tag{81}$$

Thus, the closed-loop system is exponentially stable in the sense of the full state norm

$$\left(G_{\rm av}(t)^T G_{\rm av}(t) + \sum_{i=1}^{n} \int_0^{D_i} w_i(x,t)^2 dx + w_i(D_i,t)^2 \right)^{\frac{1}{2}} \tag{82}$$

i.e., in the transformed variable $(G_{\rm av}, w)$.

Step 5: *Exponential Stability Estimate (in L_2 norm) for the Average System* (64)–(66)

To obtain exponential stability in the sense of the norm

$$\Upsilon(t) \triangleq \left(|G_{\rm av}(t)|^2 + \sum_{i=1}^{n} \int_0^{D_i} \left[u_i^{\rm av}(x,t)\right]^2 dx + \left[u_i^{\rm av}(D_i,t)\right]^2 \right)^{\frac{1}{2}} \tag{83}$$

we must show there exist positive α_1 and α_2 such that

$$\alpha_1 \Upsilon(t)^2 \leq V(t) \leq \alpha_2 \Upsilon(t)^2. \tag{84}$$

This is straightforward to establish by using (67), (72), (75) and employing the Cauchy-Schwartz inequality and other calculations, as in the proof of [14, Theorem 2.1].

Hence, with (81), we get

$$\begin{aligned} &|G_{\rm av}(t)|^2 + \sum_{i=1}^{n} \int_0^{D_i} \left[u_i^{\rm av}(x,t)\right]^2 dx + \left[u_i^{\rm av}(D_i,t)\right]^2 \\ &\leq \frac{\alpha_2}{\alpha_1} e^{-\mu t} \left(|G_{\rm av}(0)|^2 + \sum_{i=1}^{n} \int_0^{D_i} \left[u_i^{\rm av}(x,0)\right]^2 dx + \left[u_i^{\rm av}(D_i,0)\right]^2 \right) \end{aligned} \tag{85}$$

which completes the proof of exponential stability in the original variable $(G_{\rm av}, u_{\rm av})$.

Step 6: *Invoking Averaging Theorem*

First, note that the closed-loop system (39) and (49) can be rewritten as:

$$\dot{\tilde{\theta}}_i(t-D_i) = U_i(t-D_i), \quad i = 1,\ldots,n \tag{86}$$

$$\begin{aligned} \dot{U}_i(t) = &-cU_i(t) + cK_i^T \Big\{ \hat{\Phi}(D_i,0)G(t) \\ &+ \sum_{j=1}^{i} \int_{t-D_j}^{t} \hat{\Phi}(D_i, \tau - t + D_j) \hat{H}_j(t) U_j(\tau) d\tau \\ &+ \sum_{j=i+1}^{n} \int_{t-D_i}^{t} \hat{\Phi}(D_i, \tau - t + D_i) \hat{H}_j(t) \\ &\times U_j\left(\tau - (D_j - D_i)\right) d\tau \Big\} \end{aligned} \tag{87}$$

where $\eta(t) = [\tilde{\theta}(t-D)^T, U(t)^T]^T$ is the state vector. Moreover, from the definitions of $\hat{\Phi}$ in (49), $G(t)$ in (13), $\hat{H}(t)$ in (16), and (19), one can conclude they are implicit functions of ωt such that (86) and (87) can be given in the next compact form

$$\dot{\eta}(t) = f(\omega t, \eta_t) \tag{88}$$

where $\eta_t(\Theta) = \eta(t+\Theta)$ for $-D_n \leq \Theta \leq 0$ and f is an appropriate continuous functional. Thus the averaging theorem by [25], [26] in Appendix can be applied with $\omega = 1/\epsilon$.

From (85), the origin of the average closed-loop system (64)–(66) with transport PDE for delay representation is locally exponentially stable. Then, from (24) and (25), we can conclude the same results in the norm

$$\left(\sum_{i=1}^{n} \left[\tilde{\theta}_i^{\rm av}(t-D_i)\right]^2 + \int_0^{D_i} \left[u_i^{\rm av}(x,t)\right]^2 dx + \left[u_i^{\rm av}(D_i,t)\right]^2 \right)^{\frac{1}{2}}$$

since H is non-singular, i.e., $|\tilde{\theta}_i^{\rm av}(t-D_i)| \leq |H^{-1}||G_{\rm av}(t)|$.

Thus, there exist positive constants α and β such that all solutions satisfy $\Psi(t) \leq \alpha e^{-\beta t}\Psi(0)$, $\forall\, t \geq 0$, where $\Psi(t) \triangleq \sum_{i=1}^{n}[\tilde{\theta}_i^{\rm av}(t-D_i)]^2 + \int_0^{D_i}[u_i^{\rm av}(x,t)]^2 dx + [u_i^{\rm av}(D_i,t)]^2$, or equivalently,

$$\Psi(t) \triangleq \sum_{i=1}^{n} \left[\tilde{\theta}_i^{\rm av}(t-D_i)\right]^2 + \int_{t-D_i}^{t} \left[U_i^{\rm av}(\tau)\right]^2 d\tau + \left[U_i^{\rm av}(t)\right]^2 \tag{89}$$

using (61). Then, according to the averaging theorem by [25], [26] in Appendix, for ω sufficiently large, (54)–(56) or (39) and (49), has a unique locally exponentially stable periodic solution around its equilibrium (origin) satisfying (51).

Step 7: *Asymptotic Convergence to the Extremum* (θ^*, y^*)

By using the change of variables $\tilde{\vartheta}_i(t) := \tilde{\theta}_i(t-D_i)$ and then integrating both sides of (54) within $[t, \sigma + D_i]$, we have:

$$\tilde{\vartheta}_i(\sigma + D_i) = \tilde{\vartheta}_i(t) + \int_t^{\sigma + D_i} u_i(0,s) ds, \qquad i = 1,\ldots,n. \tag{90}$$

From (57), we can rewrite (90) in terms of U, namely

$$\tilde{\vartheta}_i(\sigma + D_i) = \tilde{\vartheta}_i(t) + \int_{t-D_i}^{\sigma} U_i(\tau) d\tau. \tag{91}$$

Now, note that

$$\tilde{\theta}_i(\sigma) = \tilde{\vartheta}_i(\sigma + D_i), \quad \forall\, \sigma \in [t-D_i, t]. \tag{92}$$

Hence,

$$\tilde{\theta}_i(\sigma) = \tilde{\theta}_i(t-D_i) + \int_{t-D_i}^{\sigma} U_i(\tau) d\tau, \quad \forall\, \sigma \in [t-D_i, t]. \tag{93}$$

Applying the supremum norm to both sides of (93), we have

$$\begin{aligned} &\sup_{t-D_i \leq \sigma \leq t} \left|\tilde{\theta}_i(\sigma)\right| \\ &\leq \sup_{t-D_i \leq \sigma \leq t} \left|\tilde{\theta}_i(t-D_i)\right| + \sup_{t-D_i \leq \sigma \leq t} \int_{t-D_i}^{t} |U_i(\tau)|\, d\tau \\ &\leq \left|\tilde{\theta}_i(t-D_i)\right| + \int_{t-D_i}^{t} |U_i(\tau)|\, d\tau \ \text{(by Cauchy-Schwarz)} \\ &\leq \left|\tilde{\theta}_i(t-D_i)\right| + \sqrt{D_i} \left(\int_{t-D_i}^{t} U_i^2(\tau) d\tau \right)^{\frac{1}{2}}. \end{aligned} \tag{94}$$

Now, it is easy to check

$$\left|\tilde{\theta}_i(t-D_i)\right| + \sqrt{D_i}\left(\int_{t-D_i}^{t} U_i^2(\tau)d\tau\right)^{\frac{1}{2}}$$
$$\leq (1+\sqrt{D_i})\left(\left|\tilde{\theta}_i(t-D_i)\right|^2 + \int_{t-D_i}^{t} U_i^2(\tau)d\tau\right)^{\frac{1}{2}}. \quad (95)$$

From (94), it is straightforward to conclude that

$$\left|\tilde{\theta}_i(t)\right| \leq (1+\sqrt{D_i})\left(\left|\tilde{\theta}_i(t-D_i)\right|^2 + \int_{t-D_i}^{t} U_i^2(\tau)d\tau\right)^{\frac{1}{2}}. \quad (96)$$

Inequality (96) can be given in terms of the periodic solution $\tilde{\theta}_i^{\Pi}(t-D_i)$, $U_i^{\Pi}(\tau)$, $\forall\, \tau \in [t-D_i, t]$ as follows

$$\left|\tilde{\theta}_i(t)\right| \leq (1+\sqrt{D_i})\left(\left|\tilde{\theta}_i(t-D_i) - \tilde{\theta}_i^{\Pi}(t-D_i) + \tilde{\theta}_i^{\Pi}(t-D_i)\right|^2\right.$$
$$\left. + \int_{t-D_i}^{t} \left[U_i(\tau) - U_i^{\Pi}(\tau) + U_i^{\Pi}(\tau)\right]^2 d\tau\right)^{\frac{1}{2}}. \quad (97)$$

By applying Young's inequality and some algebra, the right-hand side of (97) and $|\tilde{\theta}_i(t)|$ can be majorized by

$$\left|\tilde{\theta}_i(t)\right| \leq \sqrt{2}(1+\sqrt{D_i})$$
$$\times\left(\left|\tilde{\theta}_i(t-D_i) - \tilde{\theta}_i^{\Pi}(t-D_i)\right|^2 + \left|\tilde{\theta}_i^{\Pi}(t-D_i)\right|^2\right.$$
$$\left. + \int_{t-D_i}^{t} \left[U_i(\tau) - U_i^{\Pi}(\tau)\right]^2 d\tau + \int_{t-D_i}^{t} \left[U_i^{\Pi}(\tau)\right]^2 d\tau\right)^{\frac{1}{2}}. \quad (98)$$

According to the averaging theorem by [25], [26], we can conclude that the actual state converges exponentially to the periodic solution, i.e., $\tilde{\theta}_i(t-D_i) - \tilde{\theta}_i^{\Pi}(t-D_i) \to 0$ and $\int_{t-D_i}^{t} \left[U_i(\tau) - U_i^{\Pi}(\tau)\right]^2 d\tau \to 0$, exponentially. Hence,

$$\limsup_{t\to+\infty} |\tilde{\theta}_i(t)| = \sqrt{2}\left(1+\sqrt{D_i}\right)$$
$$\times\left(\left|\tilde{\theta}_i^{\Pi}(t-D_i)\right|^2 + \int_{t-D_i}^{t} \left[U_i^{\Pi}(\tau)\right]^2 d\tau\right)^{\frac{1}{2}}.$$

Then, from (51), we can write $\limsup_{t\to+\infty} |\tilde{\theta}(t)| = \mathcal{O}(1/\omega)$. From (12) and reminding that $\theta(t) = \hat{\theta}(t) + S(t)$ with $S(t)$ in (14), one has that $\theta(t) - \theta^* = \tilde{\theta}(t) + S(t)$. Since the first term in the right-hand side is ultimately of order $\mathcal{O}(1/\omega)$ and the second term is of order $\mathcal{O}(|a|)$, then we state (52). From (10) and (52), we get (53). ■

Corollary 1 (Gradient ES Under Output Delays): It is easy to show that the controller (49) becomes (38) in the case of output delays or equal inputs delays. Hence, the local stability/convergence results of the multi-parameter Gradient ES in Fig. 1 with delayed output (10) and $D \geq 0$ being simply a scalar can be directly stated for the closed-loop delayed system (29) and (38) from Theorem 1 (for more details, see [37]).

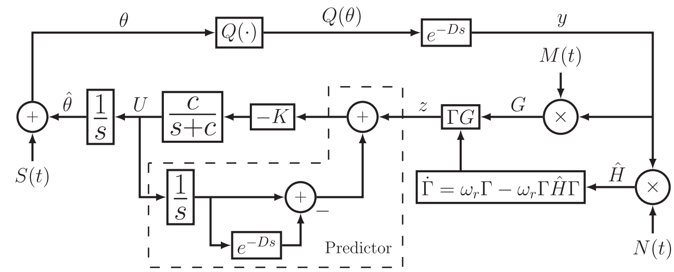


Fig. 3. Block diagram of the basic prediction scheme for output-delay compensation in multi-variable Newton-based ES, where $D \geq 0$ is a simple scalar. The predictor feedback with a perturbation-based estimate of the Hessian's inverse obeys equation (116). The dither vector signals are given by $S(t) = [a_1 \sin(\omega_1(t+D)) \ , \ldots, \ a_n \sin(\omega_n(t+D))]^T$ and $M(t) = [(2/a_1)\sin(\omega_1 t) \ , \ldots, \ (2/a_n)\sin(\omega_n t)]^T$. The demodulating signal $N(t)$ is computed by (17) and (18).

VI. Newton-Based ES With Output Delays

The convergence rate of the Gradient algorithms introduced above are severely dependent of the parameters K and H according to (37), for instance. Since the elements of the diagonal matrix K are of order $\mathcal{O}(1)$, then, the speed of the response is governed by the unknown Hessian matrix H. Once more, we start to consider the case where $D \geq 0$ is simply a scalar and in Section VII we move on to the more involved case of multiple and distinct input delays.

A. Hessian's Inverse Estimation

In [8], the authors presented a multi-variable version of the Newton ES algorithm (free of delays), which also ensures that its convergence rate can be user-assignable, rather than being dependent on the Hessian of the static map. In Fig. 3, we introduce our generalization of such a Newton-based ES in the presence of output delays.

As proved in [8], the Riccati differential equation

$$\dot{\Gamma} = \omega_r \Gamma - \omega_r \Gamma \hat{H} \Gamma \quad (99)$$

where $\omega_r > 0$ is a design constant, generates an estimate of the Hessian's inverse, avoiding inversions of the Hessian estimates that may be zero during the transient phase. The estimation error of the Hessian's inverse can be defined as

$$\tilde{\Gamma}(t) = \Gamma(t) - H^{-1} \quad (100)$$

and its dynamic equation can be written from (99), (100) by

$$\dot{\tilde{\Gamma}} = \omega_r[\tilde{\Gamma} + H^{-1}] \times \left[I_{n\times n} - \hat{H}(\tilde{\Gamma} + H^{-1})\right]. \quad (101)$$

For the quadratic map (9), and in the absence of the prediction action, it is easy to conclude that the average model in the error variables $\tilde{\theta}$ and $\tilde{\Gamma}$ would be

$$\frac{d\tilde{\theta}_{\rm av}(t)}{dt} = -K\tilde{\theta}_{\rm av}(t-D) - K\tilde{\Gamma}_{\rm av}(t)H\tilde{\theta}_{\rm av}(t-D) \tag{102}$$

$$\frac{d\tilde{\Gamma}_{\rm av}(t)}{dt} = -\omega_r\tilde{\Gamma}_{\rm av}(t) - \omega_r\tilde{\Gamma}_{\rm av}(t)H\tilde{\Gamma}_{\rm av}(t). \tag{103}$$

From (102), it is clear the importance of using a delay compensation. Differently from (38), this time we introduce a prediction scheme using (99) to estimate H^{-1}.

B. Predictor Feedback via Hessian's Inverse Estimation

First of all, let us define the measurable vector signal

$$z(t) = \Gamma(t)G(t). \tag{104}$$

Using averaging, we can verify from (13) and (104) that

$$z_{\rm av}(t) = \frac{1}{\Pi}\int_0^{\Pi} \Gamma M(\lambda)y d\lambda = \Gamma_{\rm av}(t)H\tilde{\theta}_{\rm av}(t-D). \tag{105}$$

From (100), equation (105) can be written in terms of $\tilde{\Gamma}_{\rm av}(t) = \Gamma_{\rm av}(t) - H^{-1}$ as

$$z_{\rm av}(t) = \tilde{\theta}_{\rm av}(t-D) + \tilde{\Gamma}_{\rm av}(t)H\tilde{\theta}_{\rm av}(t-D). \tag{106}$$

The second term in the right side of (106) is quadratic in $(\tilde{\Gamma}_{\rm av}, \tilde{\theta}_{\rm av})$, thus, the linearization of $\Gamma_{\rm av}(t)$ at H^{-1} and $\tilde{\theta}_{\rm av}(t)$ at zero results in the linearized version of (105) given by

$$z_{\rm av}(t) = \tilde{\theta}_{\rm av}(t-D). \tag{107}$$

From Fig. 3 and (12), we can repeat the error dynamics (29)

$$\dot{\tilde{\theta}}(t-D) = U(t-D) \tag{108}$$

and obtain the following average models by using (107), (108):

$$\dot{\tilde{\theta}}_{\rm av}(t-D) = U_{\rm av}(t-D) \tag{109}$$

$$\dot{z}_{\rm av}(t) = U_{\rm av}(t-D) \tag{110}$$

where $U_{\rm av} \in \mathbb{R}^n$ is the resulting average control for $U \in \mathbb{R}^n$.

In order to motivate the predictor feedback design, the idea again is to compensate for the delay by feeding back the future state $z(t+D)$, or $z_{\rm av}(t+D)$ in the equivalent average system. To obtain it with the variation of constants formula to (110), the future state is written as

$$z_{\rm av}(t+D) = z_{\rm av}(t) + \int_{t-D}^{t} U_{\rm av}(\tau)d\tau \tag{111}$$

in terms of the average control signal $U_{\rm av}(\tau)$ from the past window $[t-D, t]$. Given the same diagonal matrix $K > 0$ used before, the average control would be given by

$$U_{\rm av}(t) = -K\left[z_{\rm av}(t) + \int_{t-D}^{t} U_{\rm av}(\tau)d\tau\right] \tag{112}$$

resulting in the average control feedback

$$U_{\rm av}(t) = -Kz_{\rm av}(t+D), \quad \forall t \geq 0 \tag{113}$$

as desired. Hence, the average system would be, $\forall t \geq D$:

$$\frac{d\tilde{\theta}_{\rm av}(t)}{dt} = -K\tilde{\theta}_{\rm av}(t) - K\tilde{\Gamma}_{\rm av}(t+D)H\tilde{\theta}_{\rm av}(t) \tag{114}$$

$$\frac{d\tilde{\Gamma}_{\rm av}(t)}{dt} = -\omega_r\tilde{\Gamma}_{\rm av}(t) - \omega_r\tilde{\Gamma}_{\rm av}(t)H\tilde{\Gamma}_{\rm av}(t). \tag{115}$$

Since $K\tilde{\Gamma}_{\rm av}H\tilde{\theta}_{\rm av}$ is quadratic in $(\tilde{\Gamma}_{\rm av}, \tilde{\theta}_{\rm av})$ and $\omega_r\tilde{\Gamma}_{\rm av}H\tilde{\Gamma}_{\rm av}$ is quadratic in $\tilde{\Gamma}_{\rm av}$, the linearization of the system (114) and (115) has all its eigenvalues determined by $-K$ and $-\omega_r$. Therefore, the (local) exponential stability of the algorithm could be guaranteed with a convergence rate which is independent of the unknown Hessian H, being user-assignable.

As was done for the Gradient case, we propose a predictor feedback in the form of a low-pass filtered [24] of the non-average version of (112), given by

$$U(t) = \frac{c}{s+c}\left\{-K\left[z(t) + \int_{t-D}^{t} U(\tau)d\tau\right]\right\} \tag{116}$$

where $c > 0$ is sufficiently large. Recall the low pass filtering is particularly required in the stability analysis when the averaging theorem in infinite dimensions [25], [26] is invoked.

The predictor feedback control (116) is averaging-base because z in (104) is updated according to the estimate $\Gamma(t)$ for the unknown Hessian's inverse H^{-1} given by (99), with $\hat{H}(t)$ in (16) satisfying the averaging property (21).

VII. Newton-Based ES With Multiple and Distinct Input Delays

From Fig. 4 and (12), we can write

$$\dot{\tilde{\theta}}(t-D) = \begin{bmatrix} U_1(t-D_1) \\ U_2(t-D_2) \\ \vdots \\ U_n(t-D_n) \end{bmatrix}, \quad \dot{\tilde{\theta}}_i(t-D_i) = U_i(t-D_i) \tag{117}$$

and its average model

$$\dot{\tilde{\theta}}_i^{\rm av}(t-D_i) = U_i^{\rm av}(t-D_i). \tag{118}$$

Analogously to the averaging steps performed before to obtain (105)–(107), we still verify that

$$z_{\rm av}(t) = \Gamma_{\rm av}(t)H\tilde{\theta}_{\rm av}(t-D) \tag{119}$$

and its linearized version given by

$$z_{\rm av}(t) = \tilde{\theta}_{\rm av}(t-D) \tag{120}$$

with $\tilde{\theta}_{\rm av}(t-D) = [\tilde{\theta}_1^{\rm av}(t-D_1)\ ,\ldots,\ \tilde{\theta}_n^{\rm av}(t-D_n)]^T$. Thus, from (118) and (120), the following average model with state $z_{\rm av}(t)$ can be obtained:

$$\dot{z}_{\rm av}(t) = \begin{bmatrix} \dot{z}_1^{\rm av}(t) \\ \dot{z}_2^{\rm av}(t) \\ \vdots \\ \dot{z}_n^{\rm av}(t) \end{bmatrix} = \begin{bmatrix} U_1^{\rm av}(t-D_1) \\ U_2^{\rm av}(t-D_2) \\ \vdots \\ U_n^{\rm av}(t-D_n) \end{bmatrix}. \tag{121}$$

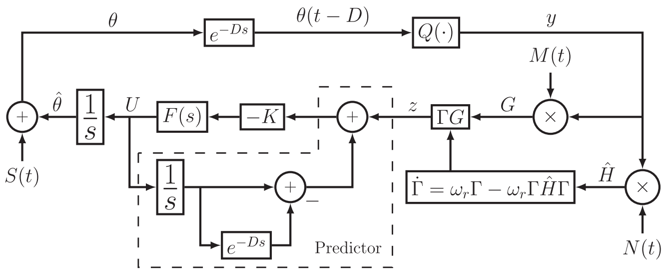


Fig. 4. Block diagram of the Newton-based ES for multiple-input delay compensation, where $D = \mathrm{diag}\{D_1, \ldots, D_n\}$. The predictor feedback (123) is implemented in its vector form with $F(s) = \mathrm{diag}\{c_1/(s+c_1), \ldots, c_n/(s+c_n)\}$ and $K = \mathrm{diag}\{k_1, \ldots, k_n\}$. The additive dither is modified to be $S(t) = [a_1 \sin(\omega_1(t+D_1)), \ldots, a_n \sin(\omega_n(t+D_n))]^T$ and the demodulating signals $M(t)$ and $N(t)$ are defined as in standard Newton-based ES [8], i.e., given by $M(t) = [(2/a_1)\sin(\omega_1 t), \ldots, (2/a_n)\sin(\omega_n t)]^T$ and (17), (18), respectively.

Since the input channels in system (121) are totally decoupled, we can apply a predictor of the form

$$U_i^{\mathrm{av}}(t) = \frac{c_i}{s+c_i}\left\{-k_i\left[z_i^{\mathrm{av}}(t) + \int_{t-D_i}^{t} U_i^{\mathrm{av}}(\tau)d\tau\right]\right\} \quad (122)$$

to each individual subsystem in (121) in order to fully stabilize it. Motivated by the control law (122) for the average model (121), we propose the following predictor feedback to compensate the delays in the non-average model (117)

$$U_i(t) = \frac{c_i}{s+c_i}\left\{-k_i\left[z_i(t) + \int_{t-D_i}^{t} U_i(\tau)d\tau\right]\right\} \quad (123)$$

for all $k_i > 0,\ c_i > 0,\ i = 1, 2, \ldots, n$.

VIII. Stability Analysis for Newton-Based ES Under Time Delays

Due to its decoupling property, (121) can be represented as a set of n cascades of first order ODE plus PDE equations. Exponential stability of the closed-loop infinite-dimensional system can be guaranteed according to the following theorem.

Theorem 2 (Multiple and Distinct Input Delays): Consider the control system in Fig. 4 with multiple and distinct input delays (7)–(11) and locally quadratic nonlinear map (9). There exists $c^* > 0$ such that, $\forall\, c_i \geq c^*$, $\exists \omega^*(c_i) > 0$ such that, $\forall\, \omega > \omega^*$, the closed-loop delayed system (101), (117) and (123) with state $\tilde{\Gamma}(t)$, $\tilde{\theta}_i(t-D_i)$, $U_i(\tau)$, $\forall\, \tau \in [t-D_i, t]$ and $\forall\, i \in \{1, 2, \ldots, n\}$, has a unique locally exponentially stable periodic solution in t of period Π, denoted by $\tilde{\Gamma}^{\Pi}(t)$, $\tilde{\theta}_i^{\Pi}(t-D)$, $U_i^{\Pi}(\tau)$, $\forall\, \tau \in [t-D_i, t]$, satisfying, $\forall\, t \geq 0$:

$$\left(\left|\tilde{\Gamma}^{\Pi}(t)\right|^2 + \sum_{i=1}^{n}\left[\tilde{\theta}_i^{\Pi}(t-D_i)\right]^2 + \left[U_i^{\Pi}(t)\right]^2 + \int_{t-D_i}^{t}\left[U_i^{\Pi}(\tau)\right]^2 d\tau\right)^{\frac{1}{2}} \leq \mathcal{O}\left(\frac{1}{\omega}\right). \quad (124)$$

Furthermore,

$$\limsup_{t\to+\infty} |\theta(t) - \theta^*| = \mathcal{O}\left(|a| + 1/\omega\right) \quad (125)$$

$$\limsup_{t\to+\infty} |y(t) - y^*| = \mathcal{O}\left(|a|^2 + 1/\omega^2\right) \quad (126)$$

where $a = [a_1,\ a_2\ , \ldots,\ a_n]^T$.

Proof: The demonstration is organized in order to follow the steps presented in the proof of Theorem 1. In the following we highlight the main differences.

Step 1: *PDE Representation for the Closed-Loop System*

After representing the delay in (117) using a transport PDE as was done in (54)–(56) and representing the integrand in (123) using the transport PDE state, one has

$$\dot{\tilde{\theta}}_i(t-D_i) = u_i(0,t) \quad (127)$$

$$\partial_t u_i(x,t) = \partial_x u_i(x,t), \quad x \in [0, D_i] \quad (128)$$

$$u(D_i,t) = \frac{c_i}{s+c_i}\left\{-k_i\left[z_i(t) + \int_0^{D_i} u_i(\sigma,t)d\sigma\right]\right\}. \quad (129)$$

Step 2: *Average Model of the Closed-loop System*

From (120) and denoting

$$\tilde{\vartheta}(t) = \tilde{\theta}(t-D) \quad (130)$$

we have $\tilde{\vartheta}_{\mathrm{av}}(t) = z_{\mathrm{av}}(t) = \tilde{\theta}_{\mathrm{av}}(t-D)$ and the linearized average version of system (127)–(129) can be obtained:

$$\dot{\tilde{\vartheta}}_i^{\mathrm{av}}(t) = u_i^{\mathrm{av}}(0,t) \quad (131)$$

$$\partial_t u_i^{\mathrm{av}}(x,t) = \partial_x u_i^{\mathrm{av}}(x,t), \quad x \in [0, D_i] \quad (132)$$

$$\frac{d}{dt}u_i^{\mathrm{av}}(D,t) = -c_i u_i^{\mathrm{av}}(D_i,t) - c_i k_i\left[\tilde{\vartheta}_i^{\mathrm{av}}(t) + \int_0^{D_i} u_i^{\mathrm{av}}(\sigma,t)d\sigma\right] \quad (133)$$

where the solution of the transport PDE (132), (133) is

$$u_i^{\mathrm{av}}(x,t) = U_i^{\mathrm{av}}(t+x-D_i). \quad (134)$$

On the other hand, by using (21), the average model for the Hessian's inverse estimation error in (101) can be deduced as in (115). Then, its linearized version at $\tilde{\Gamma}_{\mathrm{av}} = 0_{n\times n}$ is given by

$$\frac{d\tilde{\Gamma}_{\mathrm{av}}(t)}{dt} = -\omega_r \tilde{\Gamma}_{\mathrm{av}}(t). \quad (135)$$

Step 3: *Standard Backstepping Transformation, its inverse and the target system*

Consider the infinite-dimensional backstepping transformation of the delay state

$$w_i(x,t) = u_i^{\mathrm{av}}(x,t) + k_i\left(\tilde{\vartheta}_i^{\mathrm{av}}(t) + \int_0^{x} u_i^{\mathrm{av}}(\sigma,t)d\sigma\right) \quad (136)$$

which maps the system (131)–(133) into the target system:

$$\dot{\tilde{\vartheta}}_i^{\mathrm{av}}(t) = -k_i\tilde{\vartheta}_i^{\mathrm{av}}(t) + w_i(0,t) \quad (137)$$

$$\partial_t w_i(x,t) = \partial_x w_i(x,t), \quad x \in [0, D_i] \quad (138)$$

$$w_i(D_i,t) = -\frac{1}{c_i}\partial_t u_i^{\mathrm{av}}(D_i,t), \quad i = 1, 2, \ldots, n. \quad (139)$$

Using (136) for $x = D_i$ and the fact that $u_i^{\rm av}(D_i,t) = U_i^{\rm av}(t)$, from (139) we get (133), i.e.,

$$U_i^{\rm av}(t) = \frac{c_i}{s+c_i}\left\{-k_i\left[\tilde{\vartheta}_i^{\rm av}(t) + \int_0^{D_i} u_i^{\rm av}(\sigma,t)d\sigma\right]\right\}. \quad (140)$$

Let us now consider $w_i(D_i,t)$. It is easily seen that

$$\partial_t w_i(D_i,t) = \partial_t u_i^{\rm av}(D_i,t) + k_i u_i^{\rm av}(D_i,t) \quad (141)$$

where $\partial_t u_i^{\rm av}(D_i,t) = \dot{U}_i^{\rm av}(t)$. The inverse of (136) is

$$u_i^{\rm av}(x,t) = w_i(x,t) - k_i\left[e^{-k_i x}\tilde{\vartheta}_i^{\rm av}(t) + \int_0^x e^{-k_i(x-\sigma)}w_i(\sigma,t)d\sigma\right]. \quad (142)$$

Plugging (139) and (142) into (141), we get

$$\partial_t w_i(D_i,t) = -c_i w_i(D_i,t) + k_i w_i(D_i,t) - k_i^2\left[e^{-k_i D_i}\tilde{\vartheta}_i^{\rm av}(t) + \int_0^{D_i} e^{-k_i(D_i-\sigma)}w_i(\sigma,t)d\sigma\right]. \quad (143)$$

Step 4: *Lyapunov-Krasovskii Functional*

The exponential stability of the overall system is established with the Lyapunov functional

$$V(t) = \sum_{i=1}^{n} V_i(t), \quad i = 1,\ldots,n \quad (144)$$

where $V_i(t)$ are functionals

$$V_i(t) = \frac{1}{2}\left[\tilde{\vartheta}_i^{\rm av}(t)\right]^2 + \frac{\bar{a}_i}{2}\int_0^{D_i}(1+x)w_i^2(x,t)dx + \frac{1}{2}w_i^2(D_i,t) \quad (145)$$

for each subsystem in (137)–(139) and $\bar{a}_i > 0$ being appropriate constants to be chosen later. Thus, we have

$$\begin{aligned}\dot{V}_i(t) =& -k_i\left[\tilde{\vartheta}_i^{\rm av}(t)\right]^2 + \tilde{\vartheta}_i^{\rm av}(t)w_i(0,t) \\ &+ \bar{a}_i\int_0^{D_i}(1+x)w_i(x,t)\partial_x w_i(x,t)dx \\ &+ w_i(D_i,t)\partial_t w_i(D_i,t) \\ =& -k_i\left[\tilde{\vartheta}_i^{\rm av}(t)\right]^2 + \tilde{\vartheta}_i^{\rm av}(t)w_i(0,t) \\ &+ \frac{\bar{a}_i(1+D_i)}{2}w_i^2(D_i,t) - \frac{\bar{a}_i}{2}w_i^2(0,t) \\ &- \frac{\bar{a}_i}{2}\int_0^{D_i} w_i^2(x,t)dx + w_i(D_i,t)\partial_t w_i(D_i,t) \\ \le& -k_i\left[\tilde{\vartheta}_i^{\rm av}(t)\right]^2 + \frac{1}{2\bar{a}_i}\left[\tilde{\vartheta}_i^{\rm av}(t)\right]^2 - \frac{\bar{a}_i}{2}\int_0^{D_i} w_i^2(x,t)dx \\ &+ w_i(D_i,t)\left[\partial_t w_i(D_i,t) + \frac{\bar{a}_i(1+D_i)}{2}w_i(D_i,t)\right].\end{aligned}$$

Reminding that $k_i > 0$, let us choose $\bar{a}_i = 1/k_i$. Then

$$\begin{aligned}\dot{V}_i(t) \le& -\frac{1}{2\bar{a}_i}\left[\tilde{\vartheta}_i^{\rm av}(t)\right]^2 - \frac{\bar{a}_i}{2}\int_0^{D_i} w_i^2(x,t)dx \\ &+ w_i(D_i,t)\left[\partial_t w_i(D_i,t) + \frac{\bar{a}_i(1+D_i)}{2}w_i(D_i,t)\right]. \quad (146)\end{aligned}$$

Now we consider (146) along with (143). With a completion of squares, we obtain

$$\begin{aligned}\dot{V}_i(t) \le& -\frac{1}{4\bar{a}_i}\left[\tilde{\vartheta}_i^{\rm av}(t)\right]^2 - \frac{\bar{a}_i}{4}\int_0^{D_i} w_i^2(x,t)dx \\ &+ \bar{a}_i\left|k_i^2 e^{-k_i D_i}\right|^2 w_i^2(D_i,t) \\ &+ \frac{1}{\bar{a}_i}\left\|k_i^2 e^{-k_i(D_i-\sigma)}\right\|^2 w_i^2(D_i,t) \\ &+ \left[\frac{\bar{a}_i(1+D_i)}{2} + k_i\right]w_i^2(D_i,t) - c_i w_i^2(D_i,t)\end{aligned} \quad (147)$$

where the spatial variable $\sigma \in [0, D_i]$ and $\|\cdot\|$ denotes the L_2 norm in σ. Then, from (147), we arrive at

$$\dot{V}_i(t) \le -\frac{1}{4\bar{a}_i}\left[\tilde{\vartheta}_i^{\rm av}(t)\right]^2 - \frac{\bar{a}_i}{4}\int_0^{D_i} w_i^2(x,t)dx - (c_i - c_i^*)\,w^2(D,t) \quad (148)$$

where

$$c_i^* = \frac{\bar{a}_i(1+D_i)}{2} + k_i + \bar{a}_i\left|k_i^2 e^{-k_i D_i}\right|^2 + \frac{1}{\bar{a}_i}\left\|k_i^2 e^{-k_i(D_i-\sigma)}\right\|^2. \quad (149)$$

Hence, from (148), if c_i is chosen such that $c_i > c_i^*$, we obtain

$$\dot{V}_i(t) \le -\mu_i V_i(t) \quad \text{or} \quad \dot{V}(t) \le -\mu V(t) \quad (150)$$

for some $\mu_i > 0$ and $\mu = n\mu_i$. Thus, the closed-loop system is exponentially stable in the sense of the full state norm

$$\left(\left|\tilde{\vartheta}_{\rm av}(t)\right|^2 + \sum_{i=1}^{n}\int_0^{D_i} w_i^2(x,t)dx + w_i^2(D_i,t)\right)^{\frac{1}{2}} \quad (151)$$

i.e., in the transformed variable $(\tilde{\vartheta}_{\rm av}, w)$.

Step 5: *Exponential Stability Estimate (in L_2 norm) for the Average System* (131)–(133)

To obtain exponential stability in the sense of the norm

$$\Upsilon(t) \triangleq \left(|\tilde{\vartheta}_{\rm av}(t)|^2 + \sum_{i=1}^{n}\int_0^{D_i}\left[u_i^{\rm av}(x,t)\right]^2 dx + \left[u_i^{\rm av}(D_i,t)\right]^2\right)^{\frac{1}{2}} \quad (152)$$

we show from (136), (142) and (144), (145) that there exist positive numbers α_1 and α_2 such that $\alpha_1\Upsilon(t)^2 \le V(t) \le \alpha_2\Upsilon(t)^2$. Hence, with (150), we get

$$\Upsilon(t) \le \frac{\alpha_2}{\alpha_1}e^{-\mu t}\Upsilon(0) \quad (153)$$

which completes the proof of exponential stability in the original variable $(\tilde{\vartheta}_{\rm av}, u_{\rm av})$.

Step 6: *Invoking Averaging Theorem*

First, note that the closed-loop system (117), (123) and (101) can be rewritten as:

$$\dot{\tilde{\theta}}_i(t-D_i)=U_i(t-D_i), \quad i=1,\ldots,n \tag{154}$$

$$\dot{U}_i(t)=-c_iU_i(t)-c_ik_i\left[z_i(t)+\int_{t-D_i}^{t}U_i(\tau)d\tau\right] \tag{155}$$

$$\dot{\tilde{\Gamma}}(t)=\omega_r\left[\tilde{\Gamma}(t)+H^{-1}\right]\times\left[I_{n\times n}-\hat{H}(t)\left(\tilde{\Gamma}(t)+H^{-1}\right)\right] \tag{156}$$

where $\chi(t)=[\tilde{\Gamma}(t)^T,\tilde{\theta}(t-D)^T,U(t)^T]^T$ is the state vector. From the definitions of $\hat{H}(t)$ in (16) and $z(t)$ in (104), one has they are implicit functions of ωt such that the averaging theorem by [25], [26] in Appendix can be directly applied considering $\omega=1/\epsilon$ and $\chi_t(\Theta)=\chi(t+\Theta)$ for $-D_n\leq\Theta\leq 0$.

From (153), the origin of the average closed-loop system (131)–(133) with transport PDE for delay representation is locally exponentially stable. In addition, we can conclude that the equilibrium $\tilde{\Gamma}_{\rm av}(t)=0$ of the linearized error system (135) is also exponentially stable since $\omega_r>0$. Thus, using (130) and (134), there exist constants $\alpha,\beta>0$ such that all solutions satisfy $\Psi(t)\leq\alpha e^{-\beta t}\Psi(0)$, $\forall\, t\geq 0$, in the sense of the norm

$$\Psi(t)\triangleq\left|\tilde{\Gamma}_{\rm av}(t)\right|^2+\sum_{i=1}^{n}\left[\tilde{\theta}_i^{\rm av}(t-D_i)\right]^2+\int_{t-D_i}^{t}\left[U_i^{\rm av}(\tau)\right]^2d\tau+\left[U_i^{\rm av}(t)\right]^2. \tag{157}$$

Then, by applying the averaging theorem [25], [26] in Appendix, for ω sufficiently large, we can conclude (124).

Step 7: *Asymptotic Convergence to the Extremum* (θ^*,y^*)

Finally, to obtain (125) and (126), we only have to use the change of variables (130) and then integrate both sides of (127) within the interval $[t,\sigma+D_i]$ to get

$$\tilde{\vartheta}_i(\sigma+D_i)=\tilde{\vartheta}_i(t)+\int_{t}^{\sigma+D_i}u_i(0,s)ds, \qquad i=1,\ldots,n. \tag{158}$$

After that, we have to reproduce exactly the developments presented in **Step 7** of the proof of Theorem 1. ∎

Corollary 2 (Newton ES Under Output Delays): Analogously to Corollary 1, local exponential stability of the multivariable Newton-based ES in Fig. 3 with delayed output (10) and $D\geq 0$ being a simple scalar can be obtained from Theorem 2 regarding the closed-loop delayed system (101), (108) and the predictor based control law (116). For more details, see reference [37].

IX. Source Seeking Application

In this example, multi-variable ES is used for finding a source of a signal (chemical, acoustic, electromagnetic, etc.) of unknown concentration field as in (9) with Hessian and its inverse given by:

$$H=\begin{bmatrix}-2 & -2\\ -2 & -4\end{bmatrix}, \quad H^{-1}=\begin{bmatrix}-1 & 0.5\\ 0.5 & -0.5\end{bmatrix}.$$

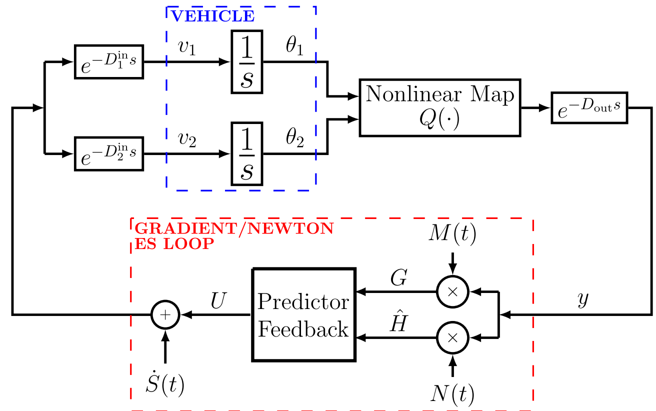


Fig. 5. Source seeking under delays for velocity-actuated point mass with additive dither $\dot{S}(t)=[a_1\omega_1\cos(\omega_1(t+D_1))\quad a_2\omega_2\cos(\omega_2(t+D_2))]^T$. The signals $M(t)$ and $N(t)$ are chosen according to (15) and (17), (18). The predictor-based controllers (49) or (123) are used in the ES loop to compensate the total delay $D_1=D_1^{\rm in}+D_{\rm out}$ and $D_2=D_2^{\rm in}+D_{\rm out}$ in Gradient and Newton schemes, respectively.

The strength of this field decays with the distance and has a local maximum at $y^*=1$ and unknown maximizer $\theta^*=(\theta_1^*,\theta_2^*)=(0,1)$. This is achieved without the measurement of the position vector $\theta=(\theta_1,\theta_2)$ and using only the measurement of the output scalar signal y with delay $D_{\rm out}=5$ s. The two actuator paths of the vehicle are also under distinct delays $D_1^{\rm in}=10$ s and $D_2^{\rm in}=20$ s. Thus, the total delays to be compensated by the predictors are $D_1=15$ s and $D_2=25$ s.

The proposed schemes are slightly modified for the stated task in Fig. 5 by observing that the integrator, a key adaptation element, is already present in vehicle models where the primary forces or moments acting on the vehicle are those that provide thrust/propulsion [21]. Thus, an application of our result for single and double integrators in control of autonomous vehicles modeled as point mass is possible. However, due to lack of space, we consider the simplest case of a velocity-actuated point mass only, where the additive dither in (14) is changed by $\dot{S}(t)$ since the integrator of the vehicle dynamics can be moved to the ES loop for analysis purposes. For the double integrator case, it would be needed to replace the lag filters used in (49) and (123) by lead compensators of the form $sc_i/(s+c_i)$, whose role is to recover some of the phase in feedback loop lost due to the addition of the second integrator [21].

We show next that the predictor feedback based ES controllers drive the autonomous vehicle modeled by

$$\dot{\theta}_1=v_1, \qquad \dot{\theta}_2=v_2 \tag{159}$$

to (θ_1^*,θ_2^*), whereas the ES automatically tunes v_1, v_2 to lead the vehicle to the peak of $Q(\theta)$.

We initially present numerical simulations for the Newton case with predictor (123), where $c_1=c_2=20$, z is given by (104) with G in (13) and Γ in (99). The control gain matrix was set to $K=\text{diag}\{1,1\}$ and $\omega_r=0.1$. We perform our tests with the parameters: $a_1=a_2=0.05$, $\omega=5$, $\omega_1=7\omega$, $\omega_2=5\omega$, $\hat{\theta}(0)=(-1,2)$ and $\Gamma(0)=-1/200\text{diag}\{2,1\}$.

Fig. 6 shows the system output $y(t)$ in 3 situations: (a) free of delays, (b) in the presence of input-output delays but without any delay compensation and (c) with input-output delays and predictor based compensation. Fig. 7 presents relevant variables for ES. It is clear the remarkable evolution of the new prediction

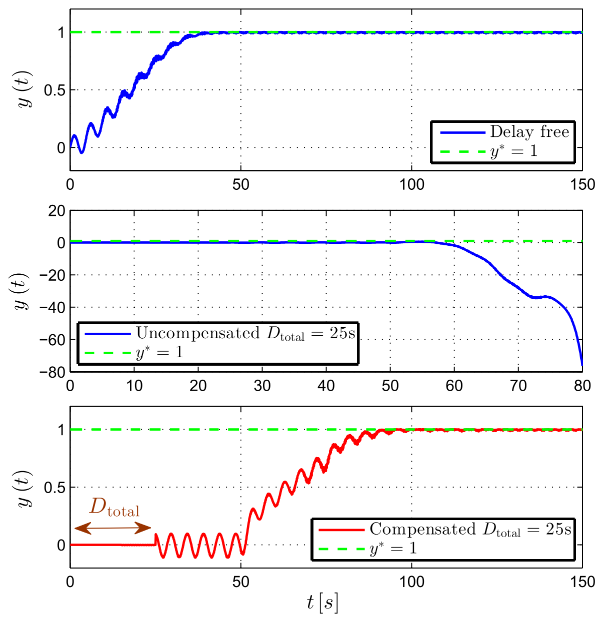


Fig. 6. Newton-based ES under input-output delays (time response of $y(t)$): (a) basic ES works well without delays; (b) ES goes unstable in the presence of delays ($D_{\text{total}} = D_2 = D_2^{\text{in}} + D_{\text{out}}$ is the longest delay); (c) predictor fixes this.

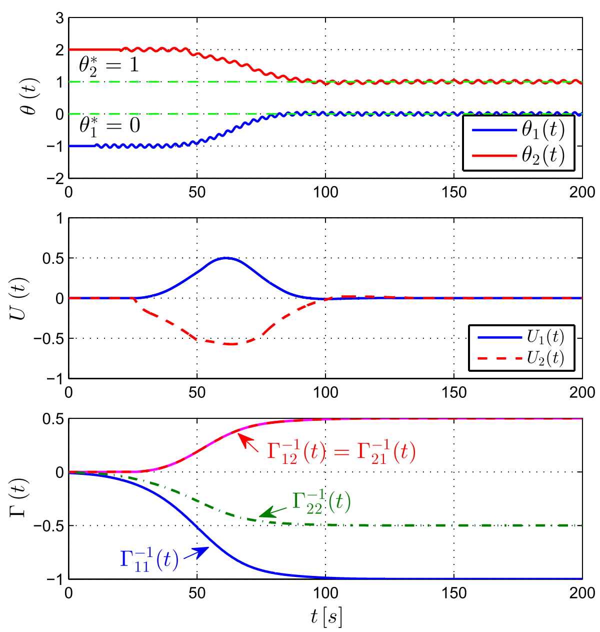


Fig. 7. Newton-based ES under input-output delays: (a) parameter $\theta(t)$; (b) the control signal $U(t)$; (c) Hessian's inverse estimate $\Gamma(t)$. The elements of $\Gamma(t)$ converge to the unknown elements of H^{-1}.

scheme in searching the maximum and the Hessian's inverse H^{-1}. This exact estimation allows us to cancel the Hessian H and thus guarantee convergence rates that can be arbitrarily assigned by the user.

In order to make a fair comparison with the Gradient ES, all common parameters are chosen the same, except for the gain matrix K. According to [8], we should select $K_G = K_N \Gamma(0)$, where K_G and K_N denote here the gain matrices for Gradient-based ES and Newton-based ES, respectively. Hence, we apply the predictor (49) for Gradient ES with averaging-based estimate of the Hessian $\hat{H}$ in (16), function $\hat{\Phi}$ given as (45)–(47)

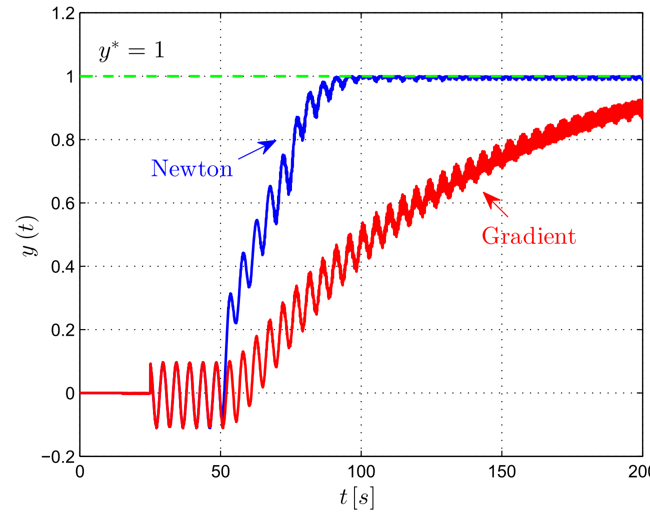


Fig. 8. Newton-based ES versus Gradient ES: Time response of the output $y(t)$ subject to input-output delays.

for $\hat{A}_i(t)$ rather than A_i, gain matrix $K = 10^{-2}\text{diag}\{1, 0.5\}$ and $c = 20$. For the case $n = 2$ (two control inputs), equation (49) is simply:

$$U_1(t) = \frac{c}{s+c}\left\{K_1^T\left(G(t) + \int_{t-D_1}^{t} \hat{H}_1 U_1(\tau)d\tau + \int_{t-D_1}^{t} \hat{H}_2 U_2\left(\tau - (D_2 - D_1)\right)d\tau\right)\right\} \tag{160}$$

$$\begin{aligned} U_2(t) = \frac{c}{s+c}\Bigg\{K_2^T\Bigg(& e^{\hat{A}_1(D_2-D_1)}G(t) \\ &+ e^{\hat{A}_1(D_2-D_1)}\int_{t-D_1}^{t} \hat{H}_1 U_1(\tau)d\tau \\ &+ e^{\hat{A}_1(D_2-D_1)}\int_{t-D_1}^{t} \hat{H}_2 U_2(\tau-(D_2-D_1))d\tau \\ &+ \int_{t-(D_2-D_1)}^{t} e^{\hat{A}_1(t-\tau)}\hat{H}_2 U_2(\tau)d\tau\Bigg)\Bigg\}. \end{aligned} \tag{161}$$

The issue of robustness to the approximation of the integral terms for prediction was raised by Mondié and Michiels in [27]. It was subsequently showed by Mirkin [28] that it is simply the result of a poor choice of the approximation scheme for the integral. Furthermore, numerical approximation schemes that are robust have been provided recently in [29].

Fig. 8 shows the estimate of the maximum for both approaches. As expected, the Newton ES converges faster than the Gradient ES, even in the presence of delays.

X. Conclusion

New extremum seeking controllers were developed for multivariable real-time optimization in the presence of actuator and sensor delays. The control scheme introduced here for output delay compensation uses prediction feedback with perturbation-based estimate of the Hessian or its inverse associated with an adequate tune of the dither signals. Two different predictor designs for Gradient-based and Newton-based extremum seeking control are formulated. The generalization for multi-input-single-output maps with distinct input delays is also addressed. While the Newton based version for distinct input delays can be obtained via conventional backstepping method, the Gradient extremum seeking is based on a novel successive backstepping-like state transformation. The predictor in the Newton case is much simpler than in the Gradient case. The Newton algorithm effectively "diagonalizes" the map and allows "decentralized" predictors for each control channel, whereas the Gradient algorithm has to perform prediction of the cross-coupling of the channels.

Results are given for static plants and the extension to include general nonlinear dynamics seems to be an open problem. We have found many interesting results in the literature of *singular perturbations for FDEs and delayed systems*. However, we have not found any theorem that is directly applicable to our problem and which would allow us to prove stability in the presence of plant dynamics, as done in the absence of delays in [1] and in papers by Tan, Nesić, and coworkers [6], [7], [20]. We need singular perturbation results that apply to distributed delays due to the the presence of a (filtered) predictor feedback in our design. In general, the existing works consider a point (fixed) delay, which must be small [30]. There are cases of time-varying or even state-dependent delays but restricted to lower order (scalar or second order) systems [31]. In some cases, the delay is distributed but it is also scaled by the small parameter and occurs in the fast dynamics [32]. There do exist results for linear systems with distributed delays in the slow dynamics, but they are not adequate since they assume that the slow/reduced model has an equilibrium at the origin when the small parameter is set to zero, i.e., that the reduced model is perturbation-free [33]. As a reminder from [1], the averaging step comes second in the stability analysis, which means that, after the singular perturbation approximation, the resulting slow/reduced model must still have the sinusoidal perturbation AND the delayed input, along with the predictor feedback. On the other hand, reference [34] solves the ES problem of dynamic plants where delays are not directly taken into account, but they may be addressed as a sufficiently large dwell-time in the algorithm. However, it would require waiting D time-delay units or longer for each function evaluation, making the transient last for many D's, whereas our transient lasts only D plus whatever the exponential transient of the delay-free ES is.

Even though we can easily compensate known time-varying delays, and unknown time-varying delays that vary sufficiently slowly, the averaging result for FDEs (Appendix) assumes the delay to be constant. In addition, we can apply our approach adaptively to unknown constant delays, as in [35]. However, because of the nonlinear parametrization of the delay, the result would necessarily be local (the initial estimate of the delay would have to be close to the actual delay). This would offer little advantage over the existing robustness of the predictor to small perturbations in the delay, which is proved in [36].

Appendix

Theorem 3 (Averaging for FDEs [25], [26])

Consider the delay system

$$\dot{x}(t) = f\left(\frac{t}{\epsilon}, x_t\right), \quad \text{for } t > 0 \tag{162}$$

where ϵ is a real parameter, $x_t(\Theta) = x(t+\Theta)$ for $-r \leq \Theta \leq 0$, and $f : \mathbb{R} \times \Omega \to \mathbb{R}^n$ is a continuous functional from a neighborhood Ω of 0 of the supremum-normed Banach space $X = C([-r, 0]; \mathbb{R}^n)$ of continuous functions from $[-r, 0]$ to $\mathbb{R}^n$. Assume that $f(t, \varphi)$ is periodic in t uniformly with respect to φ in compact subsets of Ω and that f has a continuous Fréchet derivative $\partial f(t, \varphi)/\partial\varphi$ in φ on $\mathbb{R} \times \Omega$. If $y = y_0 \in \Omega$ is an exponentially stable equilibrium for the average system $\dot{y}(t) = f_0(y_t)$, for $t > 0$, where $f_0(\varphi) = \lim_{T\to\infty}(1/T)\int_0^T f(s, \varphi)ds$, then, for some $\epsilon_0 > 0$ and $0 < \epsilon \leq \epsilon_0$, there is a unique periodic solution $t \mapsto x^*(t, \epsilon)$ of (162) with the properties of being continuous in t and ϵ, satisfying $|x^*(t, \epsilon) - y_0| \leq \mathcal{O}(\epsilon)$, for $t \in \mathbb{R}$, and such that there is $\rho > 0$ so that if $x(\cdot; \varphi)$ is a solution of (162) with $x(s) = \varphi$ and $|\varphi - y_0| < \rho$, then $|x(t) - x^*(t, \epsilon)| \leq Ce^{-\gamma(t-s)}$, for $C > 0$ and $\gamma > 0$.

References

[1] M. Krstić and H. H. Wang, "Stability of extremum seeking feedback for general nonlinear dynamic systems," *Automatica*, vol. 36, pp. 595–601, 2000.

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

[3] M. Krstić, "Extremum Seeking Control," in *Encyclopedia of Systems and Control* T. Samad and J. Baillieul, Eds. Berlin, Germany: Springer, 2014.

[4] M. Guay and T. Zhang, "Adaptive extremum seeking control of nonlinear dynamic systems with parametric uncertainties," *Automatica*, vol. 39, pp. 1283–1293, 2003.

[5] V. Adetola and M. Guay, "Parameter convergence in adaptive extremum-seeking control," *Automatica*, vol. 43, no. 1, pp. 105–110, 2007.

[6] Y. Tan, D. Nesić, and I. Mareels, "On non-local stability properties of extremum seeking control," *Automatica*, vol. 42, pp. 889–903, 2006.

[7] Y. Tan, D. Nesić, I. M. Y. Mareels, and A. Astolfi, "On global extremum seeking in the presence of local extrema," *Automatica*, vol. 45, pp. 245–251, 2009.

[8] A. Ghaffari, M. Krstić, and D. Nesić, "Multivariable Newton-based extremum seeking," *Automatica*, vol. 48, pp. 1759–1767, 2012.

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

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

[11] D. Nesić, A. Mohammadi, and C. Manzie, "A framework for extremum seeking control of systems with parameter uncertainties," *IEEE Trans. Automat. Control*, vol. 58, no. 2, pp. 435–448, Feb. 2013.

[12] A. Mohammadi, C. Manzie, and D. Nesić, "Online optimization of spark advance in alternative fueled engines using extremum seeking control," *Control Eng. Practice*, vol. 29, pp. 201–211, 2014.

[13] S.-J. Liu and M. Krstić, *Stochastic Averaging and Stochastic Extremum Seeking*, Berlin, Germany: Springer-Verlag, 2012.

[14] M. Krstić, *Delay Compensation for Nonlinear, Adaptive, PDE Systems*, Boston, MA, USA: Birkhaüser, 2009.

[15] D. J. Riggs and O. Haugan, "Advanced laser wavelength control," U. S. Patent No. US 2011/0116522 A1, May 19, 2011.

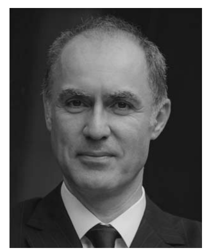
[16] H. Butler, "Adaptive feedforward for a wafer stage in a lithographic tool," *IEEE Trans. Control Syst. Technol.*, vol. 21, pp. 875–881, 2013.

[17] I. A. M. Swinnen, K. Bernaerts, E. J. J. Dens, A. H. Geeraerd, and J. F. Van Impe, "Predictive modelling of the microbial lag phase: A review," *Int. J. Food Microbiol.*, vol. 94, pp. 137–159, 2004.

[18] J. Lin, S.-M. Lee, H.-J. Lee, and Y.-M. Koo, "Modeling of typical microbial cell growth in batch culture," *Biotechnol. Bioprocess Eng.*, vol. 5, pp. 382–285, 2000.
[19] D. Tsubakino, T. R. Oliveira, and M. Krstić, "Predictor-feedback for multi-input LTI systems with distinct delays," in *Proc. Amer. Control Conf.*, Chicago, IL, USA, 2015, pp. 571–576.
[20] D. Nesić, Y. Tan, W. H. Moase, and C. Manzie, "A unifying approach to extremum seeking: Adaptive schemes based on the estimation of derivatives," in *Proc. IEEE Conf. Decision Control*, Atlanta, GA, USA, 2010, pp. 4625–4630.
[21] C. Zhang, A. Siranosian, and M. Krstić, "Extremum seeking for moderately unstable systems and for autonomous vehicle target tracking without position measurements," *Automatica*, vol. 43, pp. 1832–1839, 2007.
[22] H. K. Khalil, *Nonlinear Syst.* 3rd ed. Princeton: Prentice Hall, 2002.
[23] Y. Tan, D. Nesić, and I. Mareels, "On the choice of dither in extremum seeking systems: A case study," *Automatica*, vol. 44, pp. 1446–1450, 2008.
[24] M. Krstić, "Lyapunov tools for predictor feedbacks for delay systems: Inverse optimality and robustness to delay mismatch," *Automatica*, vol. 44, pp. 2930–2935, 2008.
[25] J. K. Hale and S. M. V. Lunel, "Averaging in infinite dimensions," *J. Integral Eq. Appl.*, vol. 2, pp. 463–494, 1990.
[26] B. Lehman, "The influence of delays when averaging slow and fast oscillating systems: Overview," *IMA J. Mathemat. Control Inform.*, vol. 19, pp. 201–215, 2002.
[27] S. Mondié and W. Michiels, "Finite spectrum assignment of unstable time-delay systems with a safe implementation," *IEEE Trans. Automat. Control*, vol. 48, no. 12, pp. 2207–2212, 2003.
[28] L. Mirkin, "On the approximation of distributed-delay control laws," *Syst. Control Lett.*, vol. 51, pp. 331–342, 2004.
[29] I. Karafyllis and M. Krstić, "Numerical schemes for nonlinear predictor feedback," *Math. Control Signals Syst.*, vol. 26, pp. 519–546, 2014.
[30] E. Fridman, "Effects of small delays on stability of singularly perturbed systems," *Automatica*, vol. 38, pp. 897–902, 2002.
[31] P. D. Paraskevopoulos, "A singular perturbation analysis with applications to delay differential equations," *J. Dynamics Differential Eq.*, vol. 7, no. 2, pp. 263–285, 1995.
[32] Z. Artstein and M. Slemrod, "On singularly perturbed retarded functional differential equations," *J. Differential Eq.*, vol. 171, pp. 88–109, 2001.
[33] V. Dragan and A. Ionita, "Exponential stability for singularly perturbed systems with state delays," *Electron. J. Qual. Theory Differential Eq.*, vol. 6, pp. 1–8, 2000.
[34] A. R. Teel and D. Popović, "Solving smooth and nonsmooth multivariable extremum seeking problems by the methods of nonlinear programming," in *Proc. Amer. Control Conf.*, Arlington, TX, USA, 2001, pp. 2394–2399.
[35] D. Bresch-Pietri and M. Krstić, "Delay-adaptive predictor feedback for systems with unknown long actuator delay," *IEEE Trans. Automat. Control*, vol. 55, no. 9, pp. 2106–2112, 2010.
[36] N. Bekiaris-Liberis and M. Krstić, "Robustness of nonlinear predictor feedback laws to time- and state-dependent delay perturbations," *Automatica*, vol. 49, pp. 1576–1590, 2013.
[37] T. R. Oliveira, M. Krstić, and D. Tsubakino, "Multiparameter extremum seeking with output delays," in *Proc. Amer. Control Conf.*, Chicago, IL, USA, 2015, pp. 152–158.

Tiago Roux Oliveira (M'09) joined State University of Rio de Janeiro (UERJ), Brazil, as an Associate Professor in 2010. He served as a Visiting Scholar at the University of California, San Diego (UCSD), USA, in 2014. His interests include nonlinear adaptive control, extremum seeking and time-delay systems.

He received from the Brazilian agency CAPES the 2011 National Award of Best Thesis in Electrical Engineering, and from FAPERJ the 2012 and 2015 Rio de Janeiro Young Researcher Award.



Miroslav Krstić (F'00) holds the Alspach chair and serves as director of the Cymer Center for Control Systems and Dynamics and as Associate Vice Chancellor for Research at University of California, San Diego (UCSD), USA. He has coauthored eleven books on adaptive, nonlinear, and stochastic control, extremum seeking, control of PDE systems including turbulent flows, and control of delay systems.

Dr. Krstic is a Fellow of IFAC, ASME, SIAM, and IET (UK), and Associate Fellow of AIAA.

Daisuke Tsubakino (M'12) received the B.E. degree from Nagoya University, Japan, in 2005, the M.E. and the Ph.D. degrees from the University of Tokyo, Japan, in 2007 and 2011, respectively.

He was an Assistant Professor at Hokkaido University from 2011 to 2015 and served as a Visiting Scholar at the University of California, San Diego, La Jolla, CA, USA, in 2013 and 2014. He is currently a lecturer at Nagoya University. His research interests include control of systems described by PDEs and hierarchical optimal control of large-scale dynamical systems.