2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

An adaptive Extremum Seeking scheme for non-convex optimisation

N. Mimmo and L. Marconi

Abstract— **The paper presents an extremum-seeking scheme in which the dither is adaptively tuned to deal with non-convex cost functions. The adaptation law decreases the dither when local cost function trends are easily visible from output data. Contrarily, when the cost function does not have a dominant trend, the dither is increased to enrich the output data. This adaptive scheme can give advantages in practical applications when a conservatively large dither implies unnecessary high energy to optimise a cost function corrupted by non-uniform state-dependent disturbances. Numerical comparisons confirm the superior performance of the proposed solution.**

I. INTRODUCTION

Extremum Seeking (ES) represents a well-known technique [1] to solve optimisation problems in several sectors.

In the framework of optimisation via ES, the cost function is unknown but for some high-level properties, such as the existence of a minimiser (in the case of cost optimisation). In this context, ES perturbs the optimisation variable by injecting an exciting signal called *dither*. Then, the local variations of the cost function are correlated with the injected variation of the optimisation variable to retrieve the regional cost function trend. With this information, ES moves the optimisation variable toward the direction that decreases the cost. It is intuitive that, for small-amplitude dithers, the local cost function trend is close to the local cost function gradient, thus making all the gradient-based decent methods applicable to solve the optimisation problem.

In this paper, we focus on the minimisation of non-convex cost functions via adaptive ES, *i.e.*, through the adoption of an adaptive dither. In this context, the following works are worth mentioning.

Guay and his co-workers have contributed to this topic since 2003 [2], [3], [4], [5], [6], [7], [8], [9], [10]. Li et al. have proficiently implemented the adaptive ES proposed by Guay et al. to solve photovoltaic cell optimisation problems [11]. The main idea is to design a controller based on some high-level knowledge of the system (both plant and cost function), described via a parametric model whose parameters are unknown. Then, a learning law estimates these unknowns and lets the ES scheme, which uses the parametric model, search for the optimal input. All these algorithms are demonstrated to work only for convex optimisation problems. Poveda et al. proposed in [12] a hybrid ES, named *NHESC*, in which the local gradient of the cost function is estimated via neural networks. NHESC does not solve the optimisation of non-convex costs because, relying on the concept of gradient estimation, it gets stuck in local minima. In 2018, Kebir et al. proposed an adaptive-dither ES scheme to maximise the output power of photovoltaic cells [13]. This paper adopts a neural network that estimates the actual optimiser. Unfortunately, this algorithm assumes the disturbance, making the cost function non-convex, measurable. For this reason, it cannot be applied to solve our optimisation problem (in which we only assume the disturbance is bounded). Suttner investigated an ES algorithm with adaptive dither in [14]. In the line of principle, since this algorithm tries to estimate the local descent direction of $h(x)$, one can only guarantee that it eventually converges to a local minimum. Atta and Guay proposed an adaptive-dither ES in [15]. In the context of convex optimisation, this paper suggests reducing the dither amplitude in the proximity of the minimiser. However, this logic makes the ES algorithm stuck in local minima. Yuheng et al. have recently proposed a dither amplitude adaptive law, specialised for electrical inverters [16]. It is worth noting that the ES scheme in [16] is different from ours because of the correspondence of the optimisation variable with the dither signal.

The paradigm of making the dither amplitude sufficiently small to use gradient-based methods is inapplicable to contrast uncertainties making the cost function non-convex, such as the effect of measurement noise affecting the reading of the cost function. Indeed, intuitively, instead of decreasing, one should increase the dither amplitude to make the cost function "trend" more evident. In this direction, we take advantage of our previous results [17], which, adopting the Fourier series to represent the cost function output, assess the stability of ES with dither amplitudes not necessarily small. Based on this new paradigm, we can increase and decrease the dither amplitude when required to explore larger or smaller areas of the cost function.

It is worth noting that in [17], we designed the dither as a non-necessarily small constant parameter. Conversely, we aim to make the dither adaptive with the current work.

Section II presents a formal definition of the optimisation problem and reports the main assumptions adopted to develop the proposed solution. Section III reports the results of numerical tests. In detail, Section III is divided into two subsections, the first dedicated to the investigation of the adaptive ES proposed in this paper and the second focused on comparisons with adaptive ES schemes already existing in the literature. Finally, Section IV concludes this paper with some final comments.

We denote the set of real and natural numbers greater then 0 with $\mathbb{R}$ and $\mathbb{N}$. We use lowercase and calligraphic letters to denote vectors and subspaces, *i.e.*, $x \in \mathcal{X} \subseteq \mathbb{R}^n$, with

The authors are with the Department of Electrical and Information Engineering, "Guglielmo Marconi", University of Bologna, 40126 Bologna, Italy, e-mail: {nicola.mimmo2, lorenzo.marconi}@unibo.it

$n \in \mathbb{N}$.

II. PROBLEM FORMULATION AND PROPOSED SOLUTION

The ES problem consists of optimising an unknown cost function $h : \mathbb{R} \to \mathbb{R}$ satisfying the following two assumptions.

Assumption 1: The function h is smooth and there exists $x^\star \in \mathbb{R}$ such that

$$h(x) - h(x^\star) > 0 \quad \forall\, x \in \mathbb{R} : x \neq x^\star. \qquad \square$$

Assumption 2: There exist a locally Lipschitz and *strictly quasi-convex* function $m : \mathbb{R} \to \mathbb{R}$, a class-$\mathcal{K}_\infty$ function $\alpha(\cdot)$, and a $A \geq 0$ such that

1) $|h(x) - m(x)| \leq A$ for all $x \in \mathbb{R}$
2) for all $x_1, x_2 \in \mathbb{R} : (x_1 - x^\star)(x_2 - x^\star) \geq 0$

$$|m(x_2) - m(x_1)| \geq \alpha(|x_2 - x_1|). \qquad \square$$

Roughly, Assumption 2 allows the cost function to be non-convex but with bounded local variations. The problem of *semi-global extremum seeking* can be formulated in the following way.

Problem 1 (Optimisation): Let $r > 0$. Then, design a system

$$\begin{aligned} \dot{z} &= \phi(z, y, t) \quad z(0) = z_0 \\ x &= \psi(z) \\ y &= h(\chi(z, t)) \end{aligned} \tag{1}$$

for some $\phi(\cdot,\cdot,\cdot)$, $\chi(\cdot,\cdot)$, and $\psi(\cdot)$ such that for all z_0 satisfying $|\psi(z_0) - x^\star| \leq r$ the resulting trajectories $z(t)$ are bounded. Moreover, there exists $\epsilon > 0$ such that $\limsup_{t\to\infty} |x(t) - x^\star| \leq \epsilon$. $\square$

Define $\underline{b}_1(\cdot) \in \mathcal{K}_\infty$ as

$$\underline{b}_1(\delta) = 2 \int_0^{1/2} \alpha(\delta \sin(2\pi t)) \sin(2\pi t)\, dt \tag{2a}$$

and

$$\underline{\delta}(A) := \underline{b}_1^{-1}(A), \tag{2b}$$

then, the solution to Problem 1 proposed in this paper is named "Adaptive ES" (see Figure 1) and consists of (1) in which $z := (x, \delta, b, \sigma, a)$, $\psi(z) := x$, $\chi(z, t) := x + \delta u(t)$, and

$$\phi(z, y, t) := \begin{bmatrix} -\gamma_1\gamma_2 \dfrac{\varepsilon(y, a)}{\sqrt{\epsilon_0 + \sigma/2}} u(t) \\ \gamma_1\gamma_2 \left(g(\sqrt{b^2/(\epsilon_0 + \sigma/2)}) - \delta \right) \\ \gamma_1 \left(\varepsilon(y, a) u(t) - b \right) \\ \gamma_1 \left(\varepsilon^2(y, a) - \sigma \right) \\ \gamma_1\, \varepsilon(y, a) \end{bmatrix},$$

with $\gamma_1, \gamma_2, \epsilon_0 > 0$, $\varepsilon(y, a) := y - a$, where $u(t) := \sin(2\pi t)$ is the dither signal and $z_0 \in \mathbb{R}\times\mathbb{R}^+\times\mathbb{R}\times\mathbb{R}^+\times\mathbb{R}$. Moreover, $g(\cdot) : [0, +\infty) \to \mathbb{R}^+$ is a smooth non-increasing function such that $g(0) \geq \underline{\delta}(A)$.

The remaining of this section provides an intuitive description of Adaptive ES. Quantities x and a represent the optimisation variable and a low-pass version of the output $y(t)$. These two components are typical of standard ES schemes, as detailed in [18]. Conversely, b and σ are not commonly found in classic ES schemes. Indeed, they are proportional to the local trend of the cost function and the variation around the local mean, respectively. To understand these concepts, assume γ_2 is sufficiently small such that x and δ are quasi-constant with $x \approx x_0$ and $\delta \approx \delta_0$, define the local trend as $\int_0^1 dh(x)/dx|_{x_0 + \delta_0 u(\tau)}\, d\tau$, and let the algorithm run. Then, on the one hand, $\liminf_{t\to\infty} \|b(t)\|$ is significant when the cost function has a strong local trend and gets smaller when the local trend is closer to zero. On the other hand, $\limsup_{t\to\infty} \|\sigma(t)\|$ is large when the cost function shows non-negligible variations around the mean $\int_0^1 h(x_0 + \delta_0 u(\tau))\, d\tau$. Therefore, $b^2(t)/\sigma(t)$ is related to the classic signal-to-noise ratio. Consequently, $g(\cdot)$, a non-increasing function of $b^2(t)/\sigma(t)$, decreases the dither amplitude δ when the signal-to-noise ratio is high and vice-versa. With this strategy, Adaptive ES minimises the dither amplitude when possible and assures the dither maximisation to correctly estimate the cost trend in the case of local minima.

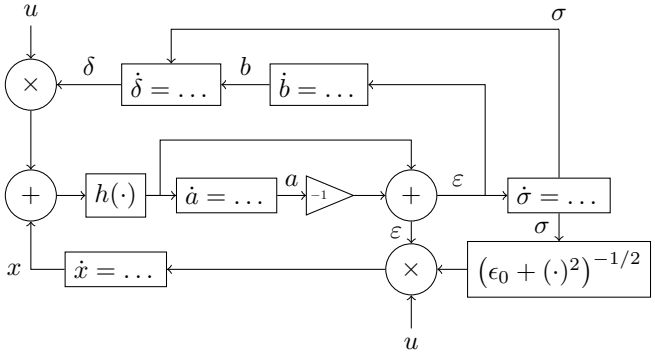

Fig. 1. Adaptive ES consists of an extension of the classic ES algorithms with the addition of b, σ, and δ dynamics.

Finally, while ε_0 is positive to make the first two entries of ϕ well-posed, parameters γ_1 and γ_2 are designed to guarantee two time-scale separations. In more detail, γ_1 is used as an *averaging* parameter while γ_2 separates fast and slow dynamics of a singularly perturbed system as described hereafter. Define $v = \texttt{col}(x, \delta)$ and $w = \texttt{col}(b, \sigma, a)$, let

$$f_v(v, w, t) := \begin{bmatrix} -\dfrac{\varepsilon(h(\chi(z,t)), a)}{\sqrt{\epsilon_0 + \sigma/2}}\, u(t), \\ -\delta + g(\sqrt{b^2/(\epsilon_0 + \sigma/2)}) \end{bmatrix}, \tag{3a}$$

$$f_w(v, w, t) := \begin{bmatrix} \varepsilon(h(\chi(z,t)), a) u(t) - b \\ \varepsilon^2(h(\chi(z,t)), a) - \sigma \\ \varepsilon(h(\chi(z,t)), a) \end{bmatrix}, \tag{3b}$$

with $z = \texttt{col}(v, w)$, and rewrite (1) as

$$\dot{v} = \gamma_1\gamma_2\, f_v(v, w, t) \qquad v(0) = v_0 \tag{4a}$$

$$\dot{w} = \gamma_1\, f_w(v, w, t) \qquad w(0) = w_0 \tag{4b}$$

in which $v_0 := \texttt{col}(x_0, \delta_0)$ and $w_0 := \texttt{col}(b_0, \sigma_0, a_0)$. The averaged system of (4) is defined as (see Appendices for

details)

$$\dot{v}_a = \gamma_1\gamma_2\, f_{va}(v_a, w_a) \qquad v_a(0) = v_0 \tag{5a}$$
$$\dot{w}_a = \gamma_1\, f_{wa}(v_a, w_a) \qquad w_a(0) = w_0. \tag{5b}$$

Let us define the *Boundary Layer* by forcing $\gamma_2 = 0$ in (5)

$$\dot{v}_a = 0 \tag{6a}$$
$$\dot{w}_a = \gamma_1 f_{wa}(v_a, w_a). \tag{6b}$$

Claim 1 (Boundary Layer): There exists $\mu(\cdot) : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}^3$ such that $f_{wa}(v_a, \mu(v_a)) = 0$ for any $v_a \in \mathbb{R} \times \mathbb{R}^+$.

Claim 1 is proved in Appendix IV-C. Now, we enforce $w_a = \mu(v_a)$ into (5) to define the *Reduced System*

$$\dot{v}_a = \gamma_1\gamma_2\, f_v(v_a, \mu(v_a)). \tag{7}$$

Intuitively, under the assumptions of asymptotic stability of (6) and (7), we use standard singular perturbation arguments to design $\gamma_2^\star > 0$ such that (5) is asymptotically stable for any $\gamma_2 \in (0, \gamma_2^\star)$. As a consequence, we use standard averaging arguments to design $\gamma_1^\star > 0$ such that the trajectories of (4) remain close to those of (5) for all $t \geq 0$. See [20] for details.

III. NUMERICAL TESTS

This section shows, through simulations, the behaviour of Adaptive ES (1). The results presented in this section have been obtained through MATLAB `ode23` function with maximum integration step of size of 0.01. The design parameters have been chosen as $\gamma_1 = \gamma_2 = 0.3$.

The following cost function has been designed, *ad-hoc*, to clearly show how δ adapts during the optimisation process. In detail, we adopted

$$\begin{aligned} d(x) &= c_1 \sin(\omega_1 x + \pi/3) + c_2 \sin(\omega_2 x + \pi/4) - 1 \\ h(x) &= c_2 \frac{x^2}{\sqrt{1+x^2}} - c_3 \cos(x) + c_2 d(x) \sin^4(\omega_3 x). \end{aligned} \tag{8}$$

Roughly, the convex cost $\dfrac{x^2}{\sqrt{1+x^2}}$ is made non-convex by $\cos(x)$ and $d(x)$ with the latter representing a high-frequency disturbance also modulated, in space, via $\sin^4(\omega_3 x)$. The cost function $h(x)$ evaluated for $x \in [-10, 50]$, $c_1 = 1$, $c_2 = 0.1$, $c_3 = 0.25$, $\omega_1 = 10$, $\omega_2 = 20$, and $\omega_3 = 0.3$, is depicted in Figure 2 and the unique global optimiser is $x^\star = 0$.

First, we analyse $h(x)$ by mapping $b_1(v)$ as in Figure 3. The background of Figure 3 represents, through a warm-colour scale, the magnitude of $b_1(v)$. In particular, the brighter the colour, the more positive is $b_1(v)$. On the same Figure, the solid white line denotes the set of v such that $b_1(v) = 0$. Therefore, for any $x > 0$, we have $b_1(v) > 0$ for any $\delta > \delta_m(x) := \sup_{\delta>0}\{\delta : b_1(v) = 0\}$. From Figure 3 we also deduce that $\underline{\delta} \approx 2.8$ within the interval $x \in [-10, 50]$. The main goal of Adaptive ES is to automatically lower $\delta(t)$ from $\underline{\delta}$ to $\delta_m(x(t))$.

As for the dither adaptation law, we adopted

$$g(s(t)) = \overline{\delta}\,(1 - \mathtt{sat}(s(t), 1))^2$$

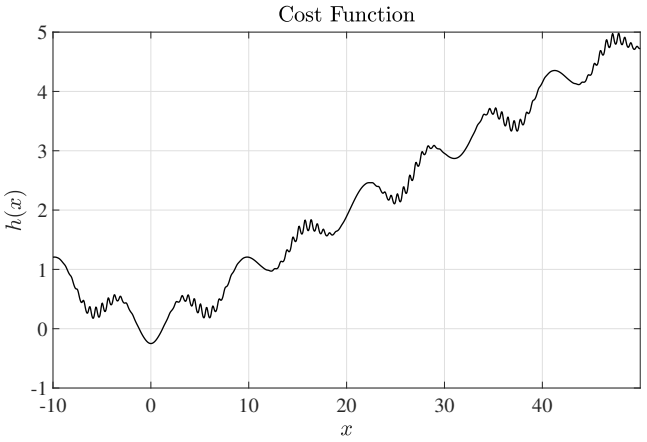

Fig. 2. Non-convex cost function exploited for simulation tests. The unique global optimiser is $x^\star = 0$.

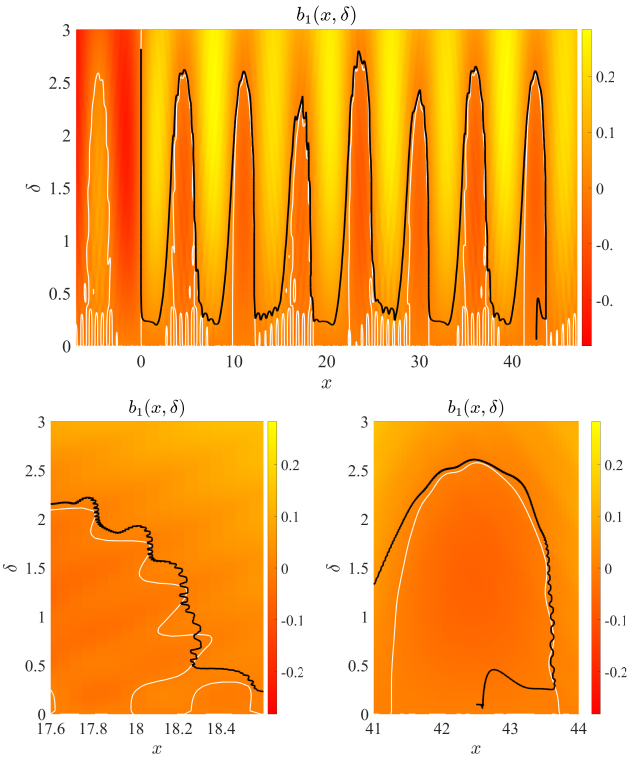

Fig. 3. Function $b_1(v)$ evaluated for $x \in [-10, 50]$ and $\delta \in [0, 3]$. The coloured background represents, via colour intensities, the values of $b_1(v)$. In details, lighter colours represent positive values of $b_1(v)$ while darker colours are associated with negative values of $b_1(v)$. A solid white line shows the set of (v) such that $b_1(v) = 0$. The black solid line represents the path of $v(t)$. (top) Overall path of $v(t)$ for $t/(2\pi) \in [0, 600]$. (bottom) Particulars highlighting the behaviour of Adaptive ES. The trajectory of $v(t)$ enters the set $\mathcal{V}_0$ from below only. Then, the trajectory moves backward in x and exist $\mathcal{V}_0$ from the nearest right boundary.

in which $\overline{\delta} = 3$ $(> \underline{\delta})$, $s(t) = \sqrt{b^2(t)/(\epsilon_0 + \sigma(t)/2)}$ with $\epsilon_0 = \mathtt{eps}$ begin the floating-point relative accuracy, and where

$$\mathtt{sat}(s, 1) := \begin{cases} s & s \leq 1 \\ 1 & \text{otherwise.} \end{cases}$$

To test the stability of Adaptive ES, we enforce un-

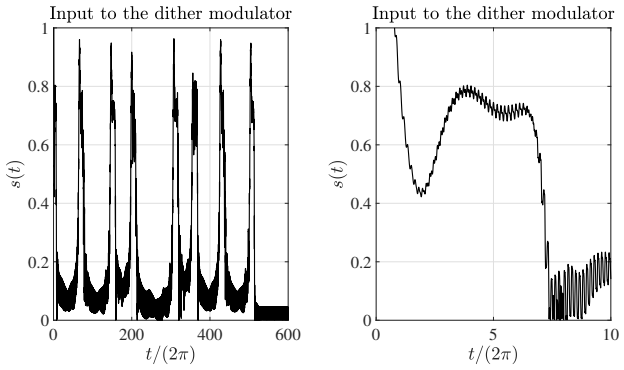

Fig. 4. Left) behaviour of $s(t)$. The signal s is high when $h(\bar{x})$, with $\bar{x} \in x + [-\delta, \delta]$, clearly shows an increasing/decreasing trend. At the opposite, s decreases when $h(\bar{x})$ is highly corrupted by the disturbances $\cos(\bar{x})$ and $d(\bar{x})$ and therefore the increasing/decreasing trend is not clear. Right) Detail for $t/(2\pi) \in [0, 10]$.

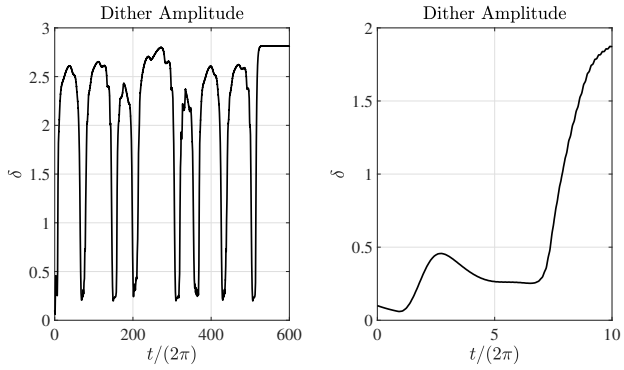

Fig. 5. Left) behaviour of $\delta(t)$. The signal δ is low when $h(\bar{x})$, with $\bar{x} \in x+[-\delta, \delta]$, clearly shows either an increasing or a decreasing trend. At the opposite, δ increases when $h(\bar{x})$ is highly corrupted by the disturbances $\cos(\bar{x})$ and $d(\bar{x})$ and therefore the increasing/decreasing trend is not clear. Right) Detail for $t/(2\pi) \in [0, 10]$.

favourable initial conditions by choosing $x_0 = 42.5$, $y_0 = h(x_0)$, $\delta_0 = 0.1$, $\sigma_0 = 0$, and $b_0 = -1$. These initial conditions make $s(t) > 1$ for all $t/(2\pi) \in [0, \tau]$, see Figure 4, where $\tau := -1/(\gamma_1 2\pi)\ln(1e\text{-}3) \approx 3.7$ represents the settling time of the *Boundary Layer* (6) to reach $99,9\%$ of the asymptotic value. Consequently, $g(s(t)) = 0$, for all $t/(2\pi) \in [0, \tau]$ and $\delta(t)$ decreases starting from δ_0, see Figure 5. Then, since $|b_1(v)| \ll 1$ in the neighbourhood of v_0, we have $\dot{x}(t) \approx 0$ for $t/(2\pi) \in [0, \tau]$, see Figure 6.

For $3.7 \lessapprox t/(2\pi) \lessapprox 3.8$, we see $s(t)$ decreasing from 1 to ≈ 0.2 thus making $\delta(t)$ increasing. Consequently, $x(t)$ starts moving where the increase of $x(t) - x^\star$ is due to the initial conditions, chose to have $b_1(v) < 0$ for $\delta \approx \delta_0$. As a consequence, $v(t)$ moves left toward the set $\mathcal{V}_0 := \{v : b_1(v) = 0\}$. As $v(t)$ approaches $\mathcal{V}_0$, $b_1(v(t))$ decreases thus making $\delta(t)$ increase while slowing down $x(t)$. Therefore, the path of $v(t)$ becomes nearly vertical as depicted by the solid black line in Figure 3.

At $t/(2\pi) \approx 8.1$, we have $b_1(v(t)) > 0$ because $\delta(t) > \delta_m(x(t))$. From that time on, the optimisation error $x(t) - x^\star$ decreases while $\delta(t)$ is modulated accordingly with $g(s(t))$, see again Figure 3.

Finally, $x(t)$ asymptotically converges to a neighbourhood of $x^\star$ around which $b_1(v(t)) \approx 0$. Therefore, $\delta(t)$ asymptotically converges to a neighbourhood of $g(0)$ such that $\dot{\delta}_a \approx 0$.

A. Comparisons with the existing literature

This section compares Adaptive ES with some of the already existing adaptation laws detailed in Section I.

The first work we compare with is [16]. We replaced the adaptation law [[16], Eq. (9)] with our $g(\cdot)$. In detail, we enforced

$$g(\sigma) = \frac{J}{2\sqrt{\sigma} + \lambda}$$

with $J = 0.3$ and $\lambda = 0.1$ to get $\|g(\sigma(t))\|_\infty = \overline{\delta}$. This new algorithm is named Algorithm 1 in what follows. Then, we run simulations with the cost function and the same initial conditions detailed in Section III, whose results are reported in Figure 7. As anticipated in Section I, in the context of non-convex optimisation, Algorithm 1 is not able to increase δ when needed. Consequently, it remains stuck on $\mathcal{V}_0$ while Adaptive ES asymptotically converges in a neighbourhood of $x^\star$.

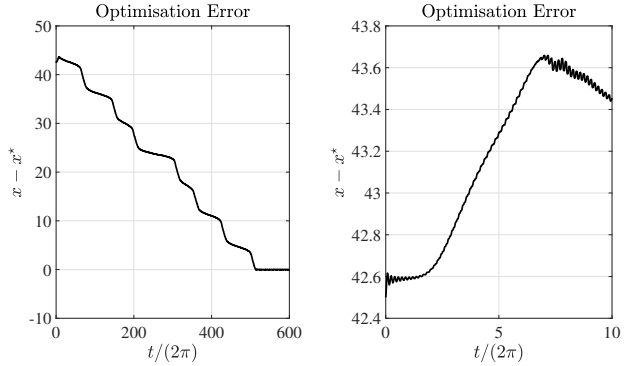

Fig. 6. Left) behaviour of $x(t) - x^\star$. The optimisation error increases, but remain bounded, within the initial time interval because of the unfavourable initial conditions. Then, after the transient of the *Boundary Layer*, $\delta(t)$ becomes sufficiently large to make $x(t) - x^\star$ decrease up to a neighbourhood of the origin. Right) Detail for $t/(2\pi) \in [0, 10]$.

The second comparison is made with the algorithm NHESC proposed in [12], hereafter recalled for completeness with the fundamental quantities renamed to be compatible, in meaning, with the notation used in this paper. In detail, NHESC is specialised as in [[12], §VI.A] with

$$\dot{x} \in \gamma_1\gamma_2 f_x(x, \hat{w})$$
$$\dot{\hat{w}} = \delta\Gamma \frac{\phi(x + \delta u(t))(\phi^\top(x + \delta u(t))\hat{w} - h(x + \delta u(t)))}{(1 + \phi^\top(x + \delta u(t))\phi(x + \delta u(t)))^2},$$

$x(0) = x_0$, $\hat{w}(0) = \hat{w}_0$, $\Gamma > 0$,

$$f_x(x, \hat{w}) = \begin{cases} 1 & \nabla J(x, \hat{w}) < 0 \\ [-1,\, 1] & \nabla J(x, \hat{w}) = 0 \\ -1 & \nabla J(x, \hat{w}) > 0 \end{cases},$$

and $\nabla J(x, \hat{w}) := \left[\frac{\partial \phi(x)}{\partial x}\right]^\top \hat{w}$. As for the basis functions,

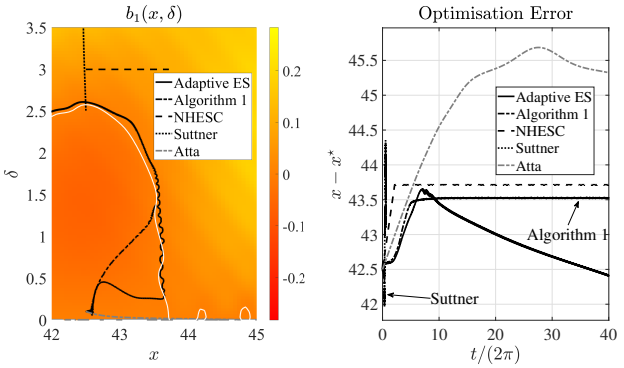

Fig. 7. Comparison of Adaptive ES, Algorithm 1 inspired by [16], and NHESC by [12]. Solid lines represents the trajectories of Adaptive ES, dashed-dotted lines depicts those of Algorithm 1, dashed lines are those of NHESC, dotted lines represents the trajectories of the algorithm by Suttner, and the grey dash-dotted lines are those of algorithm by Atta. (left) Particular of the path of $v(t)$ for the three algorithms where Algorithm 1 and NHESC get stuck on $\mathcal{V}_0$ and on a local minima. Contrarily, Adaptive ES keeps evolving and asymptotically reaches the minimiser. (right) Time-behaviour of the optimisation variable generated via Adaptive ES, Algorithm 1, NHESC,, and algorithms proposed by Suttner and Atta. This confirms that the algorithms considered in this comparison do not converge to the right minimiser while Adaptive ES does.

we assumed the structure (8) known and took

$$\phi(x) = \begin{bmatrix} x^2/\sqrt{1+x^2} \\ \cos(x) \\ \sin(\omega_1 x + \pi/3)\sin^4(\omega_3 x) \\ \sin(\omega_2 x + \pi/4)\sin^4(\omega_3 x) - \sin^4(\omega_3 x) \end{bmatrix}$$

with ω_1, ω_2, and ω_3 known. In this setup, we make the neural network estimate the non-linear combinations of coefficients c_1, c_2, and c_3 appearing in (8). We imposed $\Gamma = 500$ in agreement with [[12], §VI.A] and $u(t) = \sum_{i=1}^{3} \bar{a}_i \sin(\bar{\omega}_i t) + \bar{b}_i \cos(\bar{\omega}_i t)$ following the guidelines on [[12], §IV.B]. As anticipated in Section I, despite $\delta(t) = \bar{\delta}$ for all $t \geq 0$, NHESC gets stuck into the first encountered local minimum, see Figure 7.

The third algorithm we compare with is proposed by Suttner in [[14], Eq.s (9)-(11)], hereafter reported,

$$\begin{aligned} \dot{x} &= \gamma_1\gamma_2\delta(x,z)\sin(2\pi\Omega + 1/(z-h(x))) && x(0) = x_0 \\ \dot{z} &= -(z-h(x)) && z(0) = h(x_0)+1 \\ \dot{\Omega} &= 1/((z-h(x))^5) && \Omega(0) = 0, \end{aligned}$$

with $\delta(x,z) := \sqrt{2\pi}/((z-h(x))^2)$. Despite [[14], Theorem 1] assessing the stability of this algorithm only for quadratic $h(x)$, we found the adaptation philosophy interesting also for non-convex cost functions. When initialised at x_0, this algorithm increases $x(t)$ accordingly with the local gradient. Then, since $\partial h(x)/\partial x \approx 0$, we observe $\dot{x}(t) \approx 0$ when $x(t) \approx 43.5$. Consequently, $z(t) \approx h(x(t))$, thus making Ω and δ suddenly increase to values so large that the estimated local gradient is nearly zero (due to the symmetry of $h(x)$ around $x^\star$). Therefore, $x(t)$ is not guaranteed to converge to $x^\star$. In the simulations proposed in Figure 7, we stopped the numerical integration when $\Omega(t) > 10^6$.

As a fourth comparison, we selected the algorithm presented in [15] by Atta and Guay. This scheme adapts the dither amplitude accordingly with the following logic, customised for the case of static optimisation,

$$\begin{aligned} \dot{x} &= k_p\sigma(z) + k_i\eta && x(0) = x_0 \\ \dot{\eta} &= \sigma(z) && \eta(0) = 0 \\ \dot{a} &= \lambda(\sigma^2(z) - a) && a(0) = a_0 \\ \dot{z} &= L(t)(y - C(t)z) && z(0) = z_0 \\ \delta &= a^2 + a_0 \end{aligned}$$

where $C(t) := \begin{bmatrix} 1 & \delta\sin(2\pi t) \end{bmatrix}$, $L(t) := \begin{bmatrix} \ell_1 & \ell_2\sin(2\pi t) \end{bmatrix}^\top$, and $\sigma(z) := \tan^{-1}(2\pi z_2)/\epsilon_\sigma$, with z_2 being the second component of z. Roughly, the PI-like ES scheme composed of x and η is forced by σ, which is proportional to z_2. Then, z_2 represents the estimation of the local gradient of $h(x)$, which is approximated via a Taylor polynomial truncated at the first order as $y(x) \approx y_0(x) + y_1(x)\delta\sin(2\pi t)$, with $y_0(x)$ and $y_1(x)$ unknown. Therefore, at $x = x_0$, the estimation of the local gradient makes $x(t)$ increase to reach the nearest local minima. While approaching this critical point, $z_2(t)$ becomes smaller, thus decreasing both σ and a. The result is that the algorithm converges to the nearest local minima with a diminishing dither. The simulation reposted in Figure 7 is obtained with $k_p = 0.1$, $k_i = 10^{-4}$, $\lambda = 0.1$, $\ell_1 = 0.1$, $\ell_2 = 1$, $a_0 = 5 \cdot 10^{-3}$, $\epsilon_\sigma = 2$, $a_0 = \sqrt{0.1}$, and $z(0) = \begin{bmatrix} 0 & 0 \end{bmatrix}^\top$.

IV. CONCLUSIONS

This paper proposes an ES scheme with a dither adaptation law, which, contrary to the current literature, can be used to solve non-convex optimisation problems. In detail, adopting the Fourier series to describe the ES, at the averaging, allows the generation of meaningful information, as those embedded into b and σ. The former represents the local trend of the cost function. In contrast, the latter represents the *noise level*, *i.e.*, of how much the cost function varies in the considered interval. Therefore, the ratio b^2/σ is seen as a friend of the classic signal-to-noise ratio. Furthermore, thanks to Adaptive ES, the energy injected into the system for optimisation is reduced, which could represent an essential feature for practical applications. Future works will focus on proving the proposed scheme's stability, even with a non-unique global minimiser, and on the extension to multidimensional optimisation variables.

REFERENCES

[1] D. Dochain, M. Perrier, and M. Guay, "Extremum seeking control and its application to process and reaction systems: A survey," *Mathematics and Computers in Simulation*, vol. 82, no. 3, pp. 369 – 380, 2011. 6th Vienna International Conference on Mathematical Modelling.

[2] M. Guay and T. Zhang, "Adaptive extremum seeking control of nonlinear dynamic systems with parametric uncertainties," *Automatica*, vol. 39, no. 7, pp. 1283–1293, 2003.

[3] N. Marcos, M. Guay, and D. Dochain, "Output feedback adaptive extremum seeking control of a continuous stirred tank bioreactor with monod's kinetics," *Journal of Process Control*, vol. 14, no. 7, pp. 807–818, 2004. Dynamics, Monitoring, Control and Optimization of Biological Systems.

[4] M. Guay, D. Dochain, and M. Perrier, "Adaptive extremum seeking control of continuous stirred tank bioreactors with unknown growth kinetics," *Automatica*, vol. 40, no. 5, pp. 881–888, 2004.

[5] D. DeHaan and M. Guay, "Extremum-seeking control of state-constrained nonlinear systems," *Automatica*, vol. 41, no. 9, pp. 1567–1574, 2005.
[6] M. Guay, D. Dochain, and M. Perrier, "Adaptive extremum-seeking control of nonisothermal continuous stirred tank reactors," *Chemical Engineering Science*, vol. 60, no. 13, pp. 3671–3681, 2005.
[7] V. Adetola and M. Guay, "Adaptive output feedback extremum seeking receding horizon control of linear systems," *Journal of Process Control*, vol. 16, no. 5, pp. 521–533, 2006.
[8] V. Adetola and M. Guay, "Parameter convergence in adaptive extremum-seeking control," *Automatica*, vol. 43, no. 1, pp. 105–110, 2007.
[9] N. Hudon, M. Guay, M. Perrier, and D. Dochain, "Adaptive extremum-seeking control of convection-reaction distributed reactor with limited actuation," *Computers & Chemical Engineering*, vol. 32, no. 12, pp. 2994–3001, 2008.
[10] A. Favache, D. Dochain, M. Perrier, and M. Guay, "Extremum-seeking control of retention for a microparticulate system," *The Canadian Journal of Chemical Engineering*, vol. 86, no. 5, pp. 815–827, 2008.
[11] X. Li, Y. Li, and J. E. Seem, "Maximum power point tracking for photovoltaic system using adaptive extremum seeking control," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 6, pp. 2315–2322, 2013.
[12] J. I. Poveda, K. G. Vamvoudakis, and M. Benosman, "A neuro-adaptive architecture for extremum seeking control using hybrid learning dynamics," in *2017 American Control Conference (ACC)*, pp. 542–547, 2017.
[13] A. Kebir, L. Woodward, and O. Akhrif, "Extremum-seeking control with adaptive excitation: Application to a photovoltaic system," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 3, pp. 2507–2517, 2018.
[14] R. Suttner, "Extremum seeking control with an adaptive dither signal," *Automatica*, vol. 101, pp. 214–222, 2019.
[15] K. T. Atta and M. Guay, "Adaptive amplitude fast proportional integral phasor extremum seeking control for a class of nonlinear system," *Journal of Process Control*, vol. 83, pp. 147–154, 2019.
[16] Y. Wu, M. H. Mahmud, R. S. Krishna Moorthy, M. Chinthavali, and Y. Zhao, "Adaptive extremum seeking control based lcl filter resonant frequency online estimation," *IEEE Transactions on Power Electronics*, vol. 37, no. 1, pp. 59–64, 2022.
[17] N. Mimmo, L. Marconi, and G. Notarstefano, "Uniform quasi-convex optimisation via extremum seeking," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, pp. 6306–6311, 2022.
[18] Y. Tan, D. Nešić, and I. Mareels, "On non-local stability properties of extremum seeking control," *Automatica*, vol. 42, no. 6, pp. 889–903, 2006.
[19] H. K. Khalil and J. W. Grizzle, *Nonlinear systems*, vol. 3. Prentice hall Upper Saddle River, NJ, 2002.
[20] A. Teel, L. Moreau, and D. Nesic, "A unified framework for input-to-state stability in systems with two time scales," *IEEE Transactions on Automatic Control*, vol. 48, no. 9, pp. 1526–1544, 2003.

ACKNOWLEDGMENT

This work was partially supported by the Italian Ministry for Research in the framework of the 2020 Program for Research Projects of National Interest (PRIN), under Grant 2020RTWES4.

APPENDIX

A. AVERAGING DEFINITION

Definition 1 (Averaging [19], §10.4): Given $z \in \mathbb{R}^n$, $f(\cdot,\cdot) : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}^n$ and $f(\cdot,t) \in \mathcal{C}^1$, a small parameter $\gamma_1 > 0$ and the ordinary differential problem

$$\dot{z} = \gamma_1 f(z,t) \qquad z(0) = z_0$$

in which f is assumed T-periodic in t, then, the associated *average* system is defined as

$$\dot{z}_a = \gamma_1 f_a(z_a) \qquad z_a(0) = z_0$$

with

$$f_a(x_a) := \frac{1}{T}\int_0^T f(z_a,t)\,dt.$$

B. FOURIER COMPUTATIONS

Start by noting that $h(x_a+\delta_a u(t))$ and its time derivatives are continuous and periodic. Therefore, $h(x_a+\delta_a u(t))$ can be expressed in terms of its Fourier series as

$$\begin{aligned} h(x_a+\delta_a u(t)) &= \frac{a_0(v_a)}{2} + \\ &\sum_{k=1}^{\infty} a_k(v_a)\cos(k2\pi t) + b_k(v_a)\sin(k2\pi t) \end{aligned} \tag{9}$$

where

$$a_k(v_a) := 2\int_0^1 h(x_a+\delta_a u(t))\cos(k2\pi t)\,dt \tag{10a}$$

$$b_k(v_a) := 2\int_0^1 h(x_a+\delta_a u(t))\sin(k2\pi t)\,dt\,. \tag{10b}$$

By expanding $h(x_a+g(\sigma_a)u(\tau))$ in terms of Fourier series and by keeping in mind the definition of $\varepsilon(\cdot)$, it turns out that

$$\int_0^1 h(x_a+\delta_a u(t))\,dt = \frac{a_0(v_a)}{2} \tag{11a}$$

$$\int_0^1 \varepsilon(x_a,a_a,\delta_a,t)u(t)\,dt = \frac{b_1(v_a)}{2} \tag{11b}$$

$$\begin{aligned} \int_0^1 \varepsilon^2(x_a,a_a,\delta_a,t)\,dt &= \left(\frac{a_0(v_a)}{2}-a_a\right)^2 \\ &+ \sum_{k=1}^{\infty}\frac{a_k^2(v_a)+b_k^2(v_a)}{2} \end{aligned} \tag{11c}$$

Define

$$\begin{aligned} \rho(v_a,a_a) &= \left(\frac{a_0(v_a)}{2}-a_a\right)^2 \\ &+ \sum_{k=1}^{\infty}\frac{a_k^2(v_a)+b_k^2(v_a)}{2} \end{aligned} \tag{12}$$

and rewrite system (5) as

$$\dot{x}_a = -\gamma_1\gamma_2\,\frac{b_1(v_a)}{2\sqrt{\epsilon_0+\sigma_a/2}} \tag{13a}$$

$$\dot{\delta}_a = -\gamma_1\gamma_2\,\delta_a + \gamma_1\gamma_2\,g(\sqrt{b_1^2/(\epsilon_0+\sigma_a/2)}) \tag{13b}$$

$$\dot{b}_a = -\gamma_1\,b_a + \gamma_1\,\frac{b_1(v_a)}{2} \tag{13c}$$

$$\dot{\sigma}_a = -\gamma_1\,\sigma_a + \gamma_1\,\rho(v_a,a_a) \tag{13d}$$

$$\dot{a}_a = -\gamma_1\,a_a + \gamma_1\,\frac{a_0(v_a)}{2}. \tag{13e}$$

C. PROOF OF CLAIM 1

We take (5b) and (13c)-(13e), and define

$$\mu(v_a) = \begin{bmatrix} b_1(v_a)/2 \\ \rho(v_a,a_0(v_a)/2) \\ a_0(v_a)/2 \end{bmatrix}. \tag{14}$$

Then, $f_{wa}(v_a,\mu(v_a)) = 0$.