Extremum Seeking Over a Discrete Action Space

Michael D. Sankur and Daniel B. Arnold

***Abstract*— Extremum Seeking is a black box optimization/-control technique that utilizes perturbations to system inputs in order to optimize outputs. In this work, we propose an extension to the family of Extremum Seeking for *model-free optimization of convex functions over a discrete action space*. We refer to this method as Discrete Action Extremum Seeking (DA-ES). In this setting the DA-ES controller perturbs system inputs by visiting neighboring discrete actions, and then estimates a gradient from the resulting output signal. The DA-ES then uses the gradient estimate to select the best action to optimize the objective (e.g. the system output), and the perturbation process repeats. In this paper we outline the DA-ES algorithm, and derive convergence criteria for local minimizers of convex functions. Simulation results demonstrate the effectiveness of DA-ES in optimizing convex functions over a space of discrete actions.**

I. Introduction

Extremum Seeking (ES) is a model-free adaptive control technique which has gained popularity as a tool for black-box optimization. The scheme utilizes periodic or stochastic perturbations in system input channels to optimize system outputs [1], [2]. Unlike many meta-heuristic approaches (e.g., Particle Swarm Optimization [3]), that utilize a multitude of searchers who coordinate to optimize a given objective, ES employs a single searcher whose past actions are used to determine new search directions.

While it is common to use ES to optimize convex functions over continuously defined decision variables, little attention has been paid to the use of the technique to optimize discrete or integer-based convex mathematical programs. For these types of problems, the objective function and constraints are convex but the decision variables are discrete in nature.

Several works in literature have investigated discrete-time ES, ES with discrete or discontinuous perturbations and continuous action spaces, and online discrete optimization. In [4], the authors analyze the stability of discrete-time ES with dynamic plant, where the setpoint and control values are continuous values. The authors of [5] analyze the convergence properties of ES algorithms with perturbations that are discontinuous in time, such as square waves, triangle waves, or sawtooth waves. The setpoint and control may both take any value on a continuous number line, and thus the action space is not discrete.

The authors of [6] investigate the stability of Extremum Seeking with a variable dither amplitude that varies with the objective function value. However, all of these works consider ES where the setpoint and control may take any value within a continuous space. The authors of [7] propose online (real-time) reinforcement learning to develop optimal control policies for unknown discrete-time linear state-space models. The major limitations with the aforementioned works is the use of a continuously defined perturbation signal (e.g., $a\sin(\omega t)$, where a is the perturbation amplitude) which cannot be employed to search a discrete action space. This limitation prevents ES from being used to optimize mixed continuous and discrete (or integer-based) convex mathematical programs.

Previously, we have applied Extremum Seeking to optimize active and reactive power injections of Distributed Energy Resources (e.g., solar photovoltaic generation systems) in unbalanced electric distribution networks [8], [9]. The use of continuously defined perturbation signals prevented the use of ES to co-optimize both DER power contributions *and* voltage regulator tap positions (which are integer in nature).

To address the gap in literature examining the use of ES to optimize discrete or integer-based convex mathematical programs we propose an extension of the theory of Extremum Seeking to enable perturbations over a discrete action space, whereby both the setpoint and control of an ES algorithm may only take values from a uniformly spaced discrete set in order to optimize an unknown (convex) objective function. We refere to this scheme as Discrete Action Extremum Seeking (DA-ES). To our knowledge, this is the first work examining this problem.

II. DA-ES over a Multidimensional Discrete Action Space

This section presents the DA-ES algorithm, in which a single DA-ES algorithm operates to minimize an objective function with D decision variables. It is assumed that the action space along each dimension are independently uniform in each dimension (i.e., the discrete action step size for a single dimension is constant, but other dimensions may have different discrete action step sizes).

In this work, scalar values are given with the associated dimension of the DA-ES, and are denoted by the subscript $m \in \{1, \dots, D\}$. Vectors and matrices are denoted by bold typeface. Vectors are of dimension $D \times 1$, matrices are of dimension $D \times D$. The first entry in a double subscript denotes the DA-ES dimension. The second entry denotes the timestep k, or batch period (BP) b. The symbol $\circ$ denotes Hadamard (index-wise) multiplication, $\oslash$ denotes Hadamard (index-wise) division, $\mathbf{1} \in \mathbb{R}^D$ denotes a $D \times 1$ vector with all entries being 1, and $\mathbf{0} \in \mathbb{R}^D$ denotes a $D \times 1$ vector with

This work was supported through the Office of Energy Efficiency & Renewable Energy's Enabling Extreme Real-Time Grid Integration of Solar Energy (ENERGISE) Program in the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Michael Sankur and Daniel Arnold are with the Grid Integration Group at Lawrence Berkeley National Laboratory, Berkeley, CA, United States. msankur@lbl.gov, dbarnold@lbl.gov

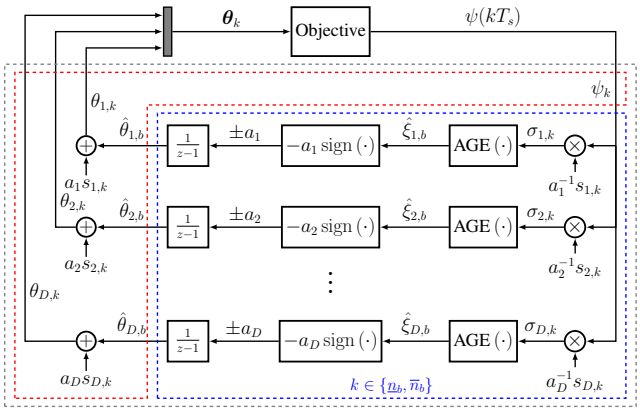

Fig. 1: Multidimensional Discrete Action Extremum Seeking algorithm with static mapping between input and objective function. The gray dashed box is the DA-ES algorithm. The red box indicates the system perturbation (probing), and the blue box encircles the demodulation, gradient estimation, and setpoint update processes that happens at the end of batch periods.

Algorithm 1 Multidimensional DA-ES Algorithm, as shown in Fig. 1.

loop
  while in batch period b: $\underline{n}_b \leq k \leq \overline{n}_b$ **do**
    - Hold setpoint constant as in (1):
      $\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_b$, $k \in \{\underline{n}_b, \ldots, \overline{n}_b\}$
    - Add perturbation to setpoint as in (3):
      $\boldsymbol{\theta}_k = \hat{\boldsymbol{\theta}}_b + \boldsymbol{A}\boldsymbol{s}_k$
    - Record and store objective function values:
      $\psi_k = \psi(kT_s)$
  end while
  if at end of batch period: $k = \overline{n}_b$ **then**
    - Demodulate objective function values with (7):
      $\boldsymbol{\sigma}_k = \boldsymbol{A}^{-1}\boldsymbol{s}_k \circ \psi_k \mathbf{1}$, $k \in \{\underline{n}_b, \ldots, \overline{n}_b\}$
    - Estimate averaged gradient for b with (8):
      $\hat{\boldsymbol{\xi}}_b = \text{AGE}\left(\boldsymbol{\sigma}_k\right)$, $k \in \{\underline{n}_b, \ldots, \overline{n}_b\}$
    - Update setpoint for $b+1$ batch period with (10):
      $\hat{\boldsymbol{\theta}}_{b+1} = \hat{\boldsymbol{\theta}}_b - \boldsymbol{A} \operatorname{sign}\left(\hat{\boldsymbol{\xi}}_b\right)$
    - Increase batch period index:
      $b \leftarrow b+1$
  end if
end loop

all entries being 0. The sign operator applied to vector $\boldsymbol{x}$ is $\operatorname{sign}(\boldsymbol{x}) = \begin{bmatrix}\operatorname{sign}(x_1), & \operatorname{sign}(x_2), & \ldots, & \operatorname{sign}(x_n)\end{bmatrix}^T$.

A. DA-ES Operation

The DA-ES algorithm is designed to minimize the value of an unknown objective function over a discrete action space that is uniform in each dimension. The DA-ES algorithm operates in discrete-time, with timesteps of length T_s, and indexed by k. The DA-ES operates over distinct batch periods (BPs), indexed by b, and consisting of timesteps $k \in \{\underline{n}_b, \ldots, \overline{n}_b\}$. System perturbation (probing) and recording objective function measurements values happen at every timestep k, as shown in red in Fig. 1. At the end of each BP b, when $k = \overline{n}_b$, the DA-ES demodulates the objective function values for BP b, computes an averaged gradient estimate for BP b, then updates its setpoint for BP $b+1$, as shown in blue in Fig. 1. Algorithm 1 outlines this operation.

We begin the discussion of DA-ES operation with by introducing pertinent variables. A local minimizer is denoted by $\boldsymbol{\theta}^* = \begin{bmatrix}\theta_1^*, & \theta_2^*, & \ldots, & \theta_D^*\end{bmatrix}^T \in \mathbb{R}^D$. The setpoint at timestep k is $\hat{\boldsymbol{\theta}}_k = \begin{bmatrix}\hat{\theta}_{1,k}, & \hat{\theta}_{2,k}, & \ldots, & \hat{\theta}_{D,k}\end{bmatrix}^T \in \mathbb{R}^D$, and the setpoint over BP b is $\hat{\boldsymbol{\theta}}_b = \begin{bmatrix}\hat{\theta}_{1,b}, & \hat{\theta}_{2,b}, & \ldots, & \hat{\theta}_{D,b}\end{bmatrix}^T \in \mathbb{R}^D$. The control at timestep k is $\boldsymbol{\theta}_k = \begin{bmatrix}\theta_{1,k}, & \theta_{2,k}, & \ldots, & \theta_{D,k}\end{bmatrix}^T \in \mathbb{R}^D$. A diagonal matrix with the discrete step sizes on its diagonal is defined by $\boldsymbol{A} = \operatorname{diag}\left(\begin{bmatrix}a_1, & a_2, & \ldots, & a_D\end{bmatrix}^T\right) \in \mathbb{R}^{D\times D}$, with $a_m > 0$, $\forall m \in \{1, \ldots, D\}$. The perturbation logic at timestep k is $\boldsymbol{s}_k = \begin{bmatrix}s_{1,k}, & s_{2,k}, & \ldots, & s_{D,k}\end{bmatrix}^T \in \mathbb{Z}^D$, with $s_{m,k} \in \mathbb{Z}$, $\forall m \in \{1, \ldots, D\}$, $\forall k$.

The DA-ES optimizes its setpoint $\hat{\boldsymbol{\theta}}_b \in \mathbb{R}^D$ to minimize the value of an unknown objective function $\psi_k\left(\boldsymbol{\theta}_k\right), \psi : \mathbb{R}^D \to \mathbb{R}$, as shown in Fig. 1, and outlined in Algorithm 1. We assume the mapping between the system input $\boldsymbol{\theta}_k$ (output of the DA-ES) and objective function value ψ_k is $\mathcal{C}^2$ and strictly convex within a neighborhood around any local minimizer.

DA-ES operation is as follows. The DA-ES holds its setpoint constant over any BP, indexed by b:

$$\begin{aligned}\hat{\theta}_{m,k} &= \hat{\theta}_{m,b},\ m \in \{1, \ldots, D\},\ k \in \{\underline{n}_b, \ldots, \overline{n}_b\},\\ \hat{\boldsymbol{\theta}}_k &= \hat{\boldsymbol{\theta}}_b,\ k \in \{\underline{n}_b, \ldots, \overline{n}_b\}.\end{aligned} \tag{1}$$

At every timestep k, The DA-ES adds perturbation $\boldsymbol{A}\boldsymbol{s}_k$ to its setpoint, forming the input to the system $\boldsymbol{\theta}_k$:

$$\begin{aligned}\theta_{m,k} &= \hat{\theta}_{m,k} + a_m s_{m,k},\ m \in \{1, \ldots, D\},\\ \boldsymbol{\theta}_k &= \hat{\boldsymbol{\theta}}_k + \boldsymbol{A}\boldsymbol{s}_k.\end{aligned} \tag{2}$$

With (1) and (2), we rewrite the input to the system in terms of the setpoint for BP b, and the perturbation at timestep k:

$$\begin{aligned}\theta_{m,k} &= \hat{\theta}_{m,b} + a_m s_{m,k},\ m \in \{1, \ldots, D\},\\ \boldsymbol{\theta}_k &= \hat{\boldsymbol{\theta}}_b + \boldsymbol{A}\boldsymbol{s}_k.\end{aligned} \tag{3}$$

The perturbation logic satisfies the following conditions, which we derive in Section II-B:

$$s_{m,k} \in \mathbb{Z},\ \forall m \in \{1, \ldots, D\},\ \forall k, \tag{4a}$$

$$\sum_{k=\underline{n}_b}^{\overline{n}_b} s_{m,k} = 0,\ \forall m \in \{1, \ldots, D\}, \tag{4b}$$

$$\sum_{k=\underline{n}_b}^{\overline{n}_b} s_{l,k} s_{m,k} = 0,\ \forall l \neq m \in \{1, \ldots, D\}, \tag{4c}$$

$$\sum_{k=\underline{n}_b}^{\overline{n}_b} s_{m,k} s_{m,k} \neq 0,\ \forall m \in \{1, \ldots, D\}, \tag{4d}$$

$$\sum_{k=\underline{n}_b}^{\overline{n}_b} s_{l,k} s_{m,k} s_{n,k} = 0,\ \forall l,m,n \in \{1,\dots,D\}. \tag{4e}$$

With appropriate choice of f_m, two examples of perturbation logic that satisfy (4) are a simple square wave:

$$s_{m,k} = \operatorname{sign}\left(\sin\left(2\pi f_m k T_s\right)\right),\ m \in \{1,\dots,D\}, \tag{5}$$

and a modified square wave:

$$s_{m,k} = \begin{cases} 1 & \text{if } 2\pi f_m k T_s \pmod{2\pi} \in \left[0, \frac{\pi}{2}\right) \\ 0 & \text{if } 2\pi f_m k T_s \pmod{2\pi} \in \left[\frac{\pi}{2}, \pi\right) \\ -1 & \text{if } 2\pi f_m k T_s \pmod{2\pi} \in \left[\pi, \frac{3\pi}{2}\right) \\ 0 & \text{if } 2\pi f_m k T_s \pmod{2\pi} \in \left[\frac{3\pi}{2}, 2\pi\right). \end{cases} \tag{6}$$

The DA-ES records objective function measurements, denoted as ψ_k.

At the end of BP b, the DA-ES demodulates the objective function measurements with its perturbation, giving $\boldsymbol{\sigma}_k = \begin{bmatrix}\sigma_{1,k} & \sigma_{2,k} & \dots & \sigma_{D,k}\end{bmatrix}^T \in \mathbb{R}^D$:

$$\begin{aligned} \sigma_{m,k} &= a_m^{-1} s_{m,k} \psi_k,\ m \in \{1,\dots,D\},\ k \in \{\underline{n}_b,\dots,\overline{n}_b\}, \\ \boldsymbol{\sigma}_k &= \boldsymbol{A}^{-1}\boldsymbol{s}_k \circ \psi_k \mathbf{1},\ k \in \{\underline{n}_b,\dots,\overline{n}_b\}. \end{aligned} \tag{7}$$

The DA-ES averages the demodulated signal over BP b to obtain the averaged gradient estimate (AGE) $\hat{\boldsymbol{\xi}}_b = \begin{bmatrix}\hat{\xi}_{1,b} & \hat{\xi}_{2,b} & \dots & \hat{\xi}_{D,b}\end{bmatrix}^T \in \mathbb{R}^D$ with the AGE operator, such that $\hat{\xi}_{m,b} = AGE\left(\sigma_{m,k}\right),\ k \in \{\underline{n}_b,\dots,\overline{n}_b\}$, and $\hat{\boldsymbol{\xi}}_b = AGE\left(\boldsymbol{\sigma}_k\right),\ k \in \{\underline{n}_b,\dots,\overline{n}_b\}$. The AGE operator is defined in (8):

$$\begin{aligned} \hat{\xi}_{m,b} &= \Gamma_{m,b} \sum_{k=\underline{n}_b}^{\overline{n}_b} \sigma_{m,k},\ m \in \{1,\dots,D\}, \\ &= \Gamma_{m,b} \sum_{k=\underline{n}_b}^{\overline{n}_b} a_m^{-1} s_{m,k} \psi_k,\ m \in \{1,\dots,D\}, \\ \hat{\boldsymbol{\xi}}_b &= \boldsymbol{\Gamma}_b \circ \sum_{k=\underline{n}_b}^{\overline{n}_b} \boldsymbol{\sigma}_k = \boldsymbol{\Gamma}_b \circ \sum_{k=\underline{n}_b}^{\overline{n}_b} \boldsymbol{A}^{-1}\boldsymbol{s}_k \circ \psi_k \mathbf{1}, \end{aligned} \tag{8}$$

where $\Gamma_{m,b}$ and $\boldsymbol{\Gamma}_b = \begin{bmatrix}\Gamma_{1,b} & \Gamma_{2,b} & \dots & \Gamma_{D,b}\end{bmatrix}^T \in \mathbb{R}^D$ remove the multiplicative scaling that otherwise comes from summing the demodulated value over BP b, and are defined by (9):

$$\Gamma_{m,b} = \left(\sum_{k=\underline{n}_b}^{\overline{n}_b} s_{m,k}^2\right)^{-1},\ \boldsymbol{\Gamma}_b = \mathbf{1} \oslash \sum_{k=\underline{n}_b}^{\overline{n}_b} \boldsymbol{s}_k \circ \boldsymbol{s}_k. \tag{9}$$

The entries of $\hat{\boldsymbol{\xi}}_b$, $\hat{\xi}_{m,b}$, are the averaged gradient estimate of the objective function with respect to the respective setpoint entry of $\hat{\boldsymbol{\theta}}_b$, $\hat{\theta}_{m,b}$

The DA-ES updates its setpoint entries based on the sign of the entries of the averaged gradient estimate and each dimension's respective discrete step size as in (10):

$$\begin{aligned} \hat{\theta}_{m,b+1} &= \hat{\theta}_{m,b} - a_m \operatorname{sign}\left(\hat{\xi}_{m,b}\right),\ m \in \{1,\dots,D\}, \\ \hat{\boldsymbol{\theta}}_{b+1} &= \hat{\boldsymbol{\theta}}_b - \boldsymbol{A} \operatorname{sign}\left(\hat{\boldsymbol{\xi}}_b\right), \end{aligned} \tag{10}$$

where the sign operator applied to $\hat{\boldsymbol{\xi}}_b$ is defined as $\operatorname{sign}\left(\hat{\boldsymbol{\xi}}_b\right) = \begin{bmatrix}\operatorname{sign}(\hat{\xi}_{1,b}) & \operatorname{sign}(\hat{\xi}_{2,b}) & \dots & \operatorname{sign}(\hat{\xi}_{D,b})\end{bmatrix}^T$. The DA-ES then progresses to the next BP, and this process repeats.

B. Derivation of DA-ES Perturbation Logic

We now derive the properties of the ES probing logic as defined in (4). We start with a second order Taylor Expansion of the objective function at the setpoint $\hat{\boldsymbol{\theta}}_b$, and neglect higher order terms:

$$\begin{aligned} \psi\left(\boldsymbol{\theta}_k\right) \approx \psi\left(\hat{\boldsymbol{\theta}}_b\right) &+ \nabla_{\boldsymbol{\theta}}\psi\left(\hat{\boldsymbol{\theta}}_b\right)\left(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_b\right) \\ &+ \frac{1}{2}\left(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_b\right)^T \nabla^2_{\boldsymbol{\theta}}\psi\left(\hat{\boldsymbol{\theta}}_b\right)\left(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_b\right), \end{aligned} \tag{11}$$

and (3) rewrite the difference between control and setpoint:

$$\begin{aligned} \psi\left(\boldsymbol{\theta}_k\right) \approx \psi\left(\hat{\boldsymbol{\theta}}_b\right) &+ \nabla_{\boldsymbol{\theta}}\psi\left(\hat{\boldsymbol{\theta}}_b\right)\boldsymbol{A}\boldsymbol{s}_k \\ &+ \frac{1}{2}\boldsymbol{s}_k^T \boldsymbol{A}^T \nabla^2_{\boldsymbol{\theta}}\psi\left(\hat{\boldsymbol{\theta}}_b\right)\boldsymbol{A}\boldsymbol{s}_k. \end{aligned} \tag{12}$$

From (8) and (12), the averaged gradient estimate is:

$$\begin{aligned} \hat{\boldsymbol{\xi}}_b \approx \boldsymbol{\Gamma}_b \circ \sum_{k=\underline{n}_b}^{\overline{n}_b} \boldsymbol{A}^{-1}\boldsymbol{s}_k \circ \Big[\psi &\left(\hat{\boldsymbol{\theta}}_b\right) + \nabla_{\boldsymbol{\theta}}\psi\left(\hat{\boldsymbol{\theta}}_b\right)\boldsymbol{A}\boldsymbol{s}_k \\ &+ \frac{1}{2}\boldsymbol{s}_k^T \boldsymbol{A}^T \nabla^2_{\boldsymbol{\theta}}\psi\left(\hat{\boldsymbol{\theta}}_b\right)\boldsymbol{A}\boldsymbol{s}_k\Big]\mathbf{1}. \end{aligned} \tag{13}$$

We seek to eliminate the constant and second order terms from the AGE and solely estimate the gradient, and therefore design $\boldsymbol{s}_k$ according to (4). Applying (4b) to (13) gives:

$$\boldsymbol{\Gamma}_b \circ \sum_{k=\underline{n}_b}^{\overline{n}_b} \boldsymbol{A}^{-1}\boldsymbol{s}_k \circ \psi\left(\hat{\boldsymbol{\theta}}_b\right)\mathbf{1} = \mathbf{0}. \tag{14}$$

Applying (4c) and (4d) to (13) gives:

$$\boldsymbol{\Gamma}_b \circ \sum_{k=\underline{n}_b}^{\overline{n}_b} \boldsymbol{A}^{-1}\boldsymbol{s}_k \circ \boldsymbol{s}_k^T \boldsymbol{A}^T \nabla^T_{\boldsymbol{\theta}}\psi\left(\hat{\boldsymbol{\theta}}_b\right)\mathbf{1} = \nabla^T_{\boldsymbol{\theta}}\psi\left(\hat{\boldsymbol{\theta}}_b\right). \tag{15}$$

Applying (4e) to (13) gives:

$$\boldsymbol{\Gamma}_b \circ \sum_{k=\underline{n}_b}^{\overline{n}_b} \boldsymbol{A}^{-1}\boldsymbol{s}_k \circ \frac{1}{2}\boldsymbol{s}_k^T \boldsymbol{A}^T \nabla^2_{\boldsymbol{\theta}}\psi\left(\hat{\boldsymbol{\theta}}_b\right)\boldsymbol{A}\boldsymbol{s}_k\mathbf{1} = \mathbf{0}. \tag{16}$$

Therefore, with the properties of the perturbation logic $\boldsymbol{s}_k$ given by (4), the AGE solely estimates the gradient, as in (17):

$$\hat{\boldsymbol{\xi}}_b \approx \nabla^T_{\boldsymbol{\theta}}\psi\left(\hat{\boldsymbol{\theta}}_b\right). \tag{17}$$

C. DA-ES Stability and Convergence Properties

We now examine the stability and convergence properties of DA-ES. First, we introduce the error coordinate $\tilde{\boldsymbol{\theta}}_b \in \mathbb{R}^D$, defined by:

$$\tilde{\boldsymbol{\theta}}_b = \hat{\boldsymbol{\theta}}_b - \boldsymbol{\theta}^*. \tag{18}$$

As the objective function is $\mathcal{C}^2$ and locally convex in a neighborhood around any local minimizer $\boldsymbol{\theta}^*$, it can be approximated by a second order Taylor Expansion:

$$\begin{aligned}\psi\left(\boldsymbol{\theta}_k\right) \approx & \psi\left(\boldsymbol{\theta}^*\right)+\nabla_{\boldsymbol{\theta}} \psi\left(\boldsymbol{\theta}^*\right)\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^*\right) \\ & +\frac{1}{2}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^*\right)^T \nabla_{\boldsymbol{\theta}}^2 \psi\left(\boldsymbol{\theta}^*\right)\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^*\right).\end{aligned} \tag{19}$$

By definition, the gradient at $\boldsymbol{\theta}^*$ is the zero vector, $\nabla_{\boldsymbol{\theta}} \psi\left(\boldsymbol{\theta}^*\right)=\mathbf{0}$. We rewrite the Hessian of the objective function evaluated at $\boldsymbol{\theta}^*$ as $\boldsymbol{Q}=\nabla_{\boldsymbol{\theta}}^2 \psi\left(\boldsymbol{\theta}^*\right) \succ 0$. Eliminating the gradient and rewriting the Hessian as $\boldsymbol{Q}$, (19) becomes: (20):

$$\psi_k \approx \psi\left(\boldsymbol{\theta}^*\right)+\frac{1}{2}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^*\right)^T \boldsymbol{Q}\left(\boldsymbol{\theta}_k-\boldsymbol{\theta}^*\right). \tag{20}$$

With (3) and (18), we rewrite ψ_k in terms of the setpoint error for BP b and the perturbation $\boldsymbol{A}\boldsymbol{s}_k$:

$$\psi_k \approx \psi\left(\boldsymbol{\theta}^*\right)+\frac{1}{2}\left(\tilde{\boldsymbol{\theta}}_b+\boldsymbol{A}\boldsymbol{s}_k\right)^T \boldsymbol{Q}\left(\tilde{\boldsymbol{\theta}}_b+\boldsymbol{A}\boldsymbol{s}_k\right), \tag{21}$$

and expand (21):

$$\begin{aligned}\psi_k \approx & \psi\left(\boldsymbol{\theta}^*\right) \ldots \\ & +\frac{1}{2}\left(\tilde{\boldsymbol{\theta}}_b^T \boldsymbol{Q} \tilde{\boldsymbol{\theta}}_b+2 \boldsymbol{s}_k^T \boldsymbol{A}^T \boldsymbol{Q} \tilde{\boldsymbol{\theta}}_b+\boldsymbol{s}_k^T \boldsymbol{A}^T \boldsymbol{Q} \boldsymbol{A} \boldsymbol{s}_k\right).\end{aligned} \tag{22}$$

With (8) and (22), the averaged gradient estimate over BP b is:

$$\begin{aligned}\hat{\boldsymbol{\xi}}_b \approx \boldsymbol{\Gamma}_b \circ \sum_{k=\underline{n}_b}^{\overline{n}_b} \boldsymbol{A}^{-1} \boldsymbol{s}_k \circ \Bigg[\psi\left(\boldsymbol{\theta}^*\right)+\frac{1}{2}\Big(\tilde{\boldsymbol{\theta}}_b^T \boldsymbol{Q} \tilde{\boldsymbol{\theta}}_b \ldots \\ +2 \boldsymbol{s}_k^T \boldsymbol{A}^T \boldsymbol{Q} \tilde{\boldsymbol{\theta}}_b+\boldsymbol{s}_k^T \boldsymbol{A}^T \boldsymbol{Q} \boldsymbol{A} \boldsymbol{s}_k\Big)\Bigg] \mathbf{1}.\end{aligned} \tag{23}$$

For clarity of presentation, we examine three parts of AGE separately. From (4b) and (23), the first part of the AGE is:

$$\boldsymbol{\Gamma}_b \circ \sum_{k=\underline{n}_b}^{\overline{n}_b} \boldsymbol{A}^{-1} \boldsymbol{s}_k \circ\left(\psi\left(\boldsymbol{\theta}^*\right)+\frac{1}{2} \tilde{\boldsymbol{\theta}}_b^T \boldsymbol{Q} \tilde{\boldsymbol{\theta}}_b\right) \mathbf{1}=\mathbf{0}. \tag{24}$$

From (4c), (4d), and (23), the second part of the AGE is:

$$\boldsymbol{\Gamma}_b \circ \sum_{k=\underline{n}_b}^{\overline{n}_b} \boldsymbol{A}^{-1} \boldsymbol{s}_k \circ \boldsymbol{s}_k^T \boldsymbol{A}^T \boldsymbol{Q} \tilde{\boldsymbol{\theta}}_b \mathbf{1}=\boldsymbol{Q} \tilde{\boldsymbol{\theta}}_b. \tag{25}$$

From (4e), and (23), the third portion of the AGE is:

$$\boldsymbol{\Gamma}_b \circ \sum_{k=\underline{n}_b}^{\overline{n}_b} \boldsymbol{A}^{-1} \boldsymbol{s}_k \circ \frac{1}{2} \boldsymbol{s}_k^T \boldsymbol{A}^T \boldsymbol{Q} \boldsymbol{A} \boldsymbol{s}_k \mathbf{1}=\mathbf{0}. \tag{26}$$

Therefore, from (24) – (26), the AGE for BP b is:

$$\hat{\boldsymbol{\xi}}_b \approx \boldsymbol{Q} \tilde{\boldsymbol{\theta}}_b, \tag{27}$$

which is the actual gradient of the objective function with respect to $\hat{\boldsymbol{\theta}}$.

The setpoint update process considers the sign of the entries of the gradient estimate, and is defined by:

$$\hat{\boldsymbol{\theta}}_{b+1}=\hat{\boldsymbol{\theta}}_b-\boldsymbol{A} \operatorname{sign}\left(\hat{\boldsymbol{\xi}}_b\right). \tag{28}$$

Subtracting $\boldsymbol{\theta}^*$ from both sides, and with (18) and (27), we rewrite (28) in terms of the error coordinate:

$$\tilde{\boldsymbol{\theta}}_{b+1} \approx \tilde{\boldsymbol{\theta}}_{\boldsymbol{b}}-\boldsymbol{A} \operatorname{sign}\left(\boldsymbol{Q} \tilde{\boldsymbol{\theta}}_b\right). \tag{29}$$

To evaluate the stability and convergence properties of the setpoint update process, we design the Lyapunov Function $V_b=\tilde{\boldsymbol{\theta}}_b^T \boldsymbol{A}^{-1} \tilde{\boldsymbol{\theta}}_b$, and examine when $V_{b+1} \leq V_b$. Setting $V_{b+1} \leq V_b$, and subtracting $\tilde{\boldsymbol{\theta}}_b^T \boldsymbol{A}^{-1} \tilde{\boldsymbol{\theta}}_b$ from both sides, gives:

$$\operatorname{sign}\left(\boldsymbol{Q} \tilde{\boldsymbol{\theta}}_b\right)^T \boldsymbol{A} \operatorname{sign}\left(\boldsymbol{Q} \tilde{\boldsymbol{\theta}}_b\right) \leq 2 \tilde{\boldsymbol{\theta}}_b^T \operatorname{sign}\left(\boldsymbol{Q} \tilde{\boldsymbol{\theta}}_b\right) \tag{30}$$

As the entries of $\operatorname{sign}\left(\boldsymbol{Q} \tilde{\boldsymbol{\theta}}_b\right) \in\{-1,0,1\}$, we can simplify this condition using the L1 norm:

$$\frac{1}{2}\left\|\boldsymbol{A} \operatorname{sign}\left(\boldsymbol{Q} \tilde{\boldsymbol{\theta}}_b\right)\right\|_1 \leq \tilde{\boldsymbol{\theta}}_b^T \operatorname{sign}\left(\boldsymbol{Q} \tilde{\boldsymbol{\theta}}_b\right). \tag{31}$$

We define $\boldsymbol{\Omega} \subset \mathbb{R}^D$ as the region containing all possible values of $\hat{\boldsymbol{\theta}}$ to which values of $\hat{\tilde{\boldsymbol{\theta}}} \notin \boldsymbol{\Omega}$ will converge toward, defined in (32).

$$\boldsymbol{\Omega}=\left\{\hat{\boldsymbol{\theta}} \left| \frac{1}{2}\right\| \boldsymbol{A} \operatorname{sign}\left(\boldsymbol{Q} \tilde{\boldsymbol{\theta}}\right)\right\|_1>\tilde{\boldsymbol{\theta}}^T \operatorname{sign}\left(\boldsymbol{Q} \tilde{\boldsymbol{\theta}}\right)\right\}. \tag{32}$$

For a one-dimensional DA-ES, (32) simplifies to $\boldsymbol{\Omega}=\left\{\hat{\theta} \left| \frac{1}{2} a>\left|\tilde{\theta}\right|\right.\right\}$.

The preceding analysis shows that DA-ES will converge to a feasible point within $\boldsymbol{\Omega}$. At this the feasible point in $\boldsymbol{\Omega}$, (32) will no longer be satisfied, and the setpoint will exit $\boldsymbol{\Omega}$ to a feasible point outside $\boldsymbol{\Omega}$.

Fig. 2 further explains the preceding analysis, using the the two-dimensional DA-ES of the experiment in Section III-C as an example. The setpoint will move toward $\boldsymbol{\Omega}$, and eventually move to a point in $\boldsymbol{\Omega}$. However the setpoint update for any point in Ω will move to a feasible value of $\hat{\boldsymbol{\theta}}$ outside $\boldsymbol{\Omega}$. The next setpoint update will return the setpoint to the previous feasible point within $\boldsymbol{\Omega}$. This can be seen by the black arrows in Fig. 2, which point to the point the value of $\hat{\boldsymbol{\theta}}$ the setpoint update will take.

The convergence rate of the DA-ES algorithm depends on the setpoint update step size, and the length of each BP.

D. Heuristic for Stopping DA-ES Setpoint Update

From the previous analysis, we develop a heuristic for stopping the probing and setpoint update process. If the gradient estimate is nonzero, and each element of the gradient estimate switches signs over the course of successive BPs, then the setpoint is assumed to oscillate between a point within $\boldsymbol{\Omega}$, and one outside. Therefore, if the gradient estimate is nonzero and switches signs over the course of N_s successive BPs, satisfying (33), and the average objective function, defined by (34), for the current BP b is less than the previous BP, then probing or setpoint update can be stopped.

$$\begin{gathered}\operatorname{sign}\left(\hat{\boldsymbol{\xi}}_{b-n}\right)=-\operatorname{sign}\left(\hat{\boldsymbol{\xi}}_{b-n-1}\right),\ n \in\{0,1, \ldots, N_s\}, \\ \operatorname{sign}\left(\hat{\boldsymbol{\xi}}_{b-n}\right) \neq \mathbf{0},\ n \in\{0,1, \ldots, N_s\}, \\ \overline{\psi}_b \leq \overline{\psi}_{b-1}.\end{gathered} \tag{33}$$

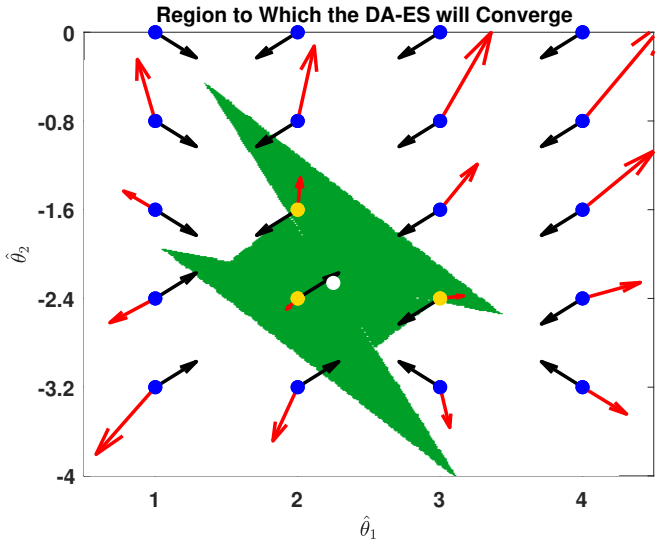

Fig. 2: Green represents $\boldsymbol{\Omega}$ as defined by (32). The white dot is the minimizer, blue dots are feasible values of $\hat{\boldsymbol{\theta}}$ outside $\boldsymbol{\Omega}$, and gold dots are feasible values of $\hat{\boldsymbol{\theta}}$ inside $\boldsymbol{\Omega}$. Red arrows plot the AGE, $\hat{\boldsymbol{\xi}}$, at $\hat{\boldsymbol{\theta}}$. Black arrows point in the direction of the setpoint update at values of $\hat{\boldsymbol{\theta}}$, defined by $-\boldsymbol{A}\,\text{sign}\left(\hat{\boldsymbol{\xi}}\right)$.

$$\overline{\psi}_b = \frac{1}{\overline{n}_b - \underline{n}_b} \sum_{k=\underline{n}_b}^{\overline{n}_b} \psi_k \tag{34}$$

E. DA-ES with Decreasing Probe Amplitude and Setpoint Update Step Size

We now consider an extension to DA-ES where the perturbation amplitude and setpoint update size decreases if the setpoint oscillates around the optimal value, and after sign of the gradient switches over several successive BPs.

Extensions similar to this have been considered in previous work. In [10], the authors propose a modification to ES in which the ES probe amplitude is proportional to the norm of system states. As the ES drives the system to an equilibrium point, the probe amplitude decreases, and for certain cases decreases to zero. In our previous work [11], the probe amplitude is dependent on the magnitude of the gradient estimate. Conversely, in this extension, the probe amplitude and setpoint update size are based on the sign of the gradient over successive batch periods, and the previous probe amplitude and setpoint update size. This extension enables the DA-ES to evaluate and traverse regions of the discrete action space faster, and over fewer BPs.

As an illustrative example, imagine a system operator can turn three knobs whose values sum to the setpoint, to find the optimal setpoint over the set of integers. The first knob moves in increments of 100, the second knob moves in increments of 10, and the third knob moves in increments of 1. First the operator finds the optimal position of the first knob, which is the closest multiple of 100 to the optimal value. Next, the operator finds the optimal position of the second knob, which is the closest multiple of 10 to the optimal value. Lastly, the operator finds the optimal position of the third knob, which is the closest multiple of 1 to the optimal value.

For this extension, we first modify (3) with the positive integer $\kappa_{m,b} \in \mathbb{Z}^+$, and diagonal matrix $\boldsymbol{\kappa}_b = \text{diag}\left(\left[\kappa_{1,b}, \quad \kappa_{2,b}, \quad \ldots, \quad \kappa_{D,b}\right]^T\right) \in \mathbb{Z}^{D\times D,+}$, such that the m^{th} perturbation takes $\kappa_{m,b}$ steps of size a_m when $|s_{m,k}| = 1$:

$$\begin{aligned} \theta_{m,k} &= \hat{\theta}_{m,b} + a_m \kappa_{m,b} s_{m,k},\ m \in \{1,\ldots,D\}, \\ \boldsymbol{\theta}_k &= \hat{\boldsymbol{\theta}}_b + \boldsymbol{A}\boldsymbol{\kappa}_b \boldsymbol{s}_k. \end{aligned} \tag{35}$$

Demodulation of the objective function is modified as:

$$\begin{aligned} \sigma_{m,k} &= (a_m \kappa_b)^{-1} s_{m,k} \psi_k,\ \ldots \\ &\qquad m \in \{1,\ldots,D\},\ k \in \{\underline{n}_b,\ldots,\overline{n}_b\}, \\ \boldsymbol{\sigma}_k &= (\boldsymbol{A}\boldsymbol{\kappa}_b)^{-1} \boldsymbol{s}_k \circ \psi_k \mathbf{1},\ k \in \{\underline{n}_b,\ldots,\overline{n}_b\}. \end{aligned} \tag{36}$$

We also modify (10) with the parameters κ_b and $\boldsymbol{\kappa}_b$, such that the setpoint update is for dimension m is $\kappa_{m,b+1}$ steps of size a_m:

$$\begin{aligned} \hat{\theta}_{m,b+1} &= \hat{\theta}_{m,b} - a_m \kappa_{m,b+1}\,\text{sign}\left(\hat{\xi}_{m,b}\right),\ m \in \{1,\ldots,D\}, \\ \hat{\boldsymbol{\theta}}_{b+1} &= \hat{\boldsymbol{\theta}}_b - \boldsymbol{A}\boldsymbol{\kappa}_{b+1}\,\text{sign}\left(\hat{\boldsymbol{\xi}}_b\right) \end{aligned} \tag{37}$$

We use the heuristic proposed in Section II-D to determine when to decrease $\kappa_{m,b}$. If (33) is not satisfied then $\boldsymbol{\kappa}_{b+1} = \boldsymbol{\kappa}_b$. When (33) is satisfied for $N_s \geq 1$, then the setpoint update size and number of discrete probe steps are decreased for one or more dimension of DA-ES, such that $\kappa_{m,b+1} \leq \kappa_{m,b}$, $\kappa_{m,b+1} \in \mathbb{Z}^+$, $m \in \{1,\ldots D\}$, and $\boldsymbol{\kappa}_{b+1} \in \mathbb{Z}^{D\times D,+}$.With slight modification of the analysis from Section II-C, the stability of this extension to DA-ES is simple to demonstrate.

III. Simulations

In this section, we present simulations in which a DA-ES optimizes its setpoint to minimize the value of an objective function.

A. DA-ES Over a Single Dimension

In our first experiment, a DA-ES optimized its setpoint to minimize the value of an unknown objective function, over the discrete set of non-negative integers, such that $a = 1$. As this DA-ES operates over a single dimension, we drop the subscript denoting the dimension. The sampling time of the DA-ES was $T_s = 0.01$. The perturbation logic s_k was given by (6) with frequency of 1 Hz. The BP length was 5 seconds, with $\overline{n}_b - \underline{n}_b = 500$. The initial setpoint was $\hat{\theta}_1 = 1$. The objective function, which was unknown to the DA-ES, was:

$$\psi_k = \exp\left(0.5\left(\theta_k - \theta^*\right)\right) + \exp\left(-0.5\left(\theta_k - \theta^*\right)\right), \tag{38}$$

where the minimizer was $\theta^* = 4.75$ for $0 \leq t < 39$ and $\theta^* = 1.2$ for $39 \leq t$.

Fig. 3 shows the DA-ES updating its setpoint toward the minimizer through successive setpoint BPs, and responding to a change in the minimizer. The dashed lines highlight how the setpoint oscillates within a neighborhood of the minimizer, with the neighborhood either being $\left(\theta^* - \frac{a}{2}, \theta^* + a\right)$, or $\left(\theta^* - a, \theta^* + \frac{a}{2}\right)$. Fig. 4 shows the DA-ES minimizing the

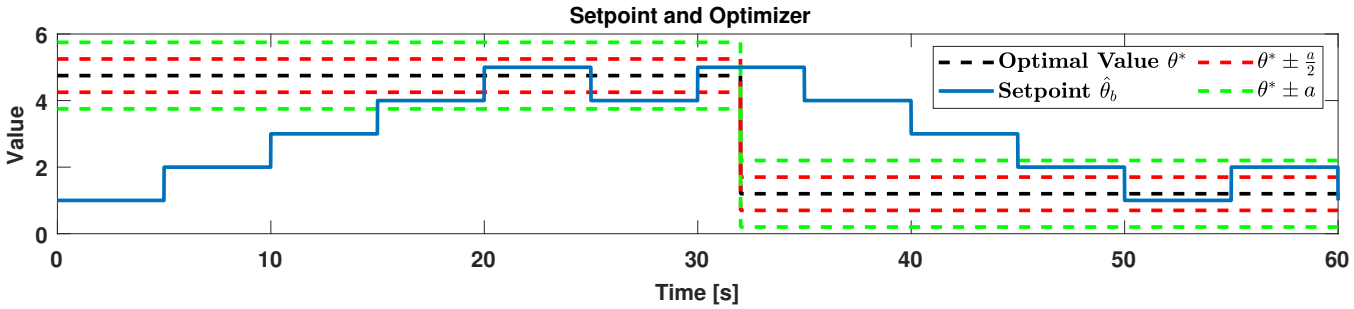

Fig. 3: Minimizer and setpoint of the DA-ES. The red dashed lines represent a region of half the probe step size around the minimizer, $\theta^* \pm \frac{a}{2}$. The green dashed lines represent a region of of the probe step size around the minimizer, $\theta^* \pm a$.

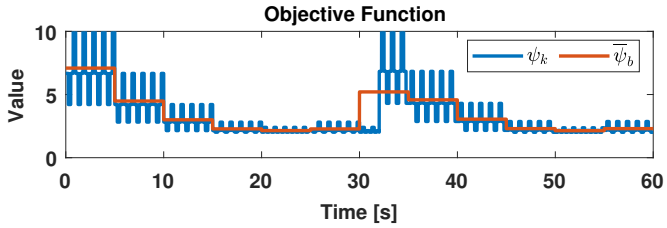

Fig. 4: Objective function, and averaged value over BPs.

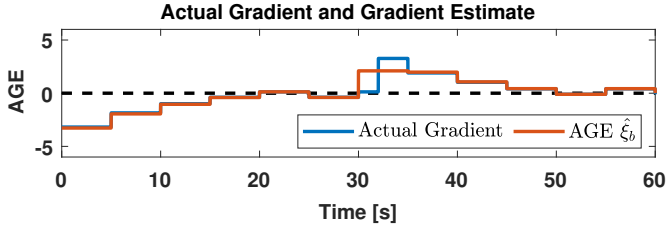

Fig. 5: Averaged gradient estimate over BPs.

the objective function value for this simulation. Fig. 5 plots the actual gradient of the objective function evaluated at the setpoint and the averaged gradient estimate for each BP.

B. DA-ES with Decreasing Probe Amplitude and Setpoint Update Step Size

In our second experiment, a DA-ES optimized its setpoint to minimize the value of an unknown objective function, and decreased its perturbation step size and setpoint update step size when its setpoint converged to within a neighborhood around the minimizer, according to the analysis in Section II-E. As this DA-ES operates over a single dimension, we drop the subscript denoting the dimension. The discrete action space was the set of integer multiples of 0.01, with $a = 0.01$. The sampling time of the DA-ES was $T_s = 0.01$. The perturbation logic s_k was given by (6) with frequency of 1 Hz. The BP length was 5 seconds, with $\overline{n}_b - \underline{n}_b = 500$. The initial setpoint was $\hat{\theta}_1 = 1$.

The initial number of discrete perturbation steps was $\kappa_1 = 100$. When (33) was first satisfied for $N_s = 3$, $\kappa_{b+1} \leftarrow 10$. The second time (33) was satisfied for $N_s = 3$, $\kappa_{b+1} \leftarrow 1$. The objective function of the this simulation was:

$$\psi_k = (\theta_k - \theta^*)^2 , \tag{39}$$

with minimizer $\theta^* = 2.74$.

Fig. 6 plots the minimizer and the DA-ES setpoint and clearly shows the DA-ES updating its setpoint toward the minimizer, along with the probe amplitude decay at 25 seconds and 45 seconds. Fig. 7 plots the objective function averaged over each BP. The objective function greatly decays as smaller values of s_k allow the setpoint to converge much closer to the minimizer. Fig. 8 plots the probe amplitude and setpoint update step size, $a\kappa_b$, and shows it twice decreasing by a factor of 10.

C. DA-ES Over a Multi-Dimensional Action Space

In our third experiment, a DA-ES optimizes its setpoint across a two dimensional discrete action space (grid) to minimize the value of an unknown objective function. The discrete action step size for the first dimension was uniform with steps of $a_1 = 1$, and the discrete action step size for the second dimension was uniform with steps of $a_2 = 0.8$. The sampling time of the DA-ES was $T_s = 0.01$. The perturbation logic for the first dimension was defined by (6), with $f_1 = 1$ Hz and $f_2 = 1.2$ Hz. The BP length was 5 seconds, with $\overline{n}_b - \underline{n}_b = 500$. The objective function was:

$$\psi_k = (\boldsymbol{\theta}_k - \boldsymbol{\theta}^*)^T \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix} (\boldsymbol{\theta}_k - \boldsymbol{\theta}^*) , \tag{40}$$

with minimizer $\boldsymbol{\theta}^* = [2.25, -2.26]^T$. We consider two initial setpoints; the first initial setpoint was $\hat{\boldsymbol{\theta}}_1 = [0, 0]^T$, and the second initial setpoint was $\hat{\boldsymbol{\theta}}_1 = [0, 0.8]^T$.

Fig. 9 plots the setpoint of the DA-ES for the first initial conditions and shows that the setpoint converges toward the minimizer. However, the first dimension of the setpoint does not oscillate around its optimal value, whereas the second dimension of the setpoint does. After converging to $\hat{\boldsymbol{\theta}}_b = [2, -1.6]^T \in \boldsymbol{\Omega}$, the setpoint oscillates between it and $\hat{\boldsymbol{\theta}}_b = [1, -2.4]^T \notin \boldsymbol{\Omega}$. Fig. 10 plots the objective function and averaged objective function for the first initial condition, and Fig. 11 plots the AGE for both channels of the DA-ES.

Fig. 12 plots the setpoint of the DA-ES for the second initial conditions and shows that setpoint converges toward the minimizer. In this case, both dimensions of the setpoint oscillate around their respective optimal values. After converging to $\hat{\boldsymbol{\theta}}_b = [2, -2.4]^T \in \boldsymbol{\Omega}$, the setpoint oscillates between it and $\hat{\boldsymbol{\theta}}_b = [2, -2.4]^T \notin \boldsymbol{\Omega}$. Fig. 13 plots the objective function and averaged objective function for the second initial condition, and Fig. 14 plots the AGE for both

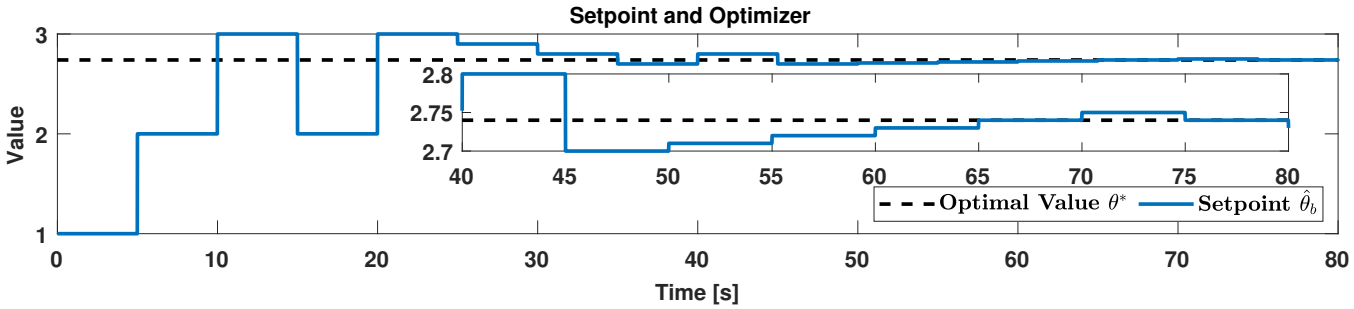

Fig. 6: Minimizer and DA-ES setpoint. The probe amplitude change from $a\kappa_b = 1$ to $a\kappa_b = 0.1$ can be seen at 25 seconds, and the change from $a\kappa_b = 0.1$ to $a\kappa_b = 0.01$ can be seen at 45 seconds.

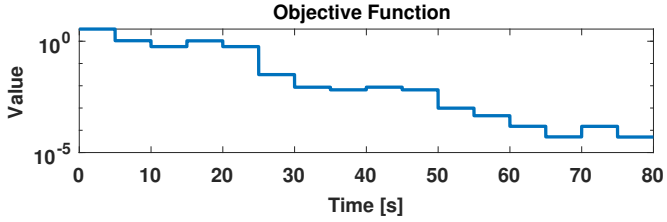

Fig. 7: Averaged objective function over BPs.

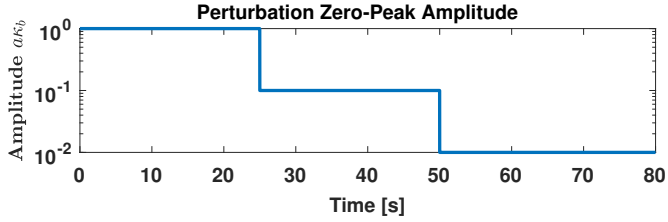

Fig. 8: Probe amplitude and setpoint update step size, $a\kappa_b$.

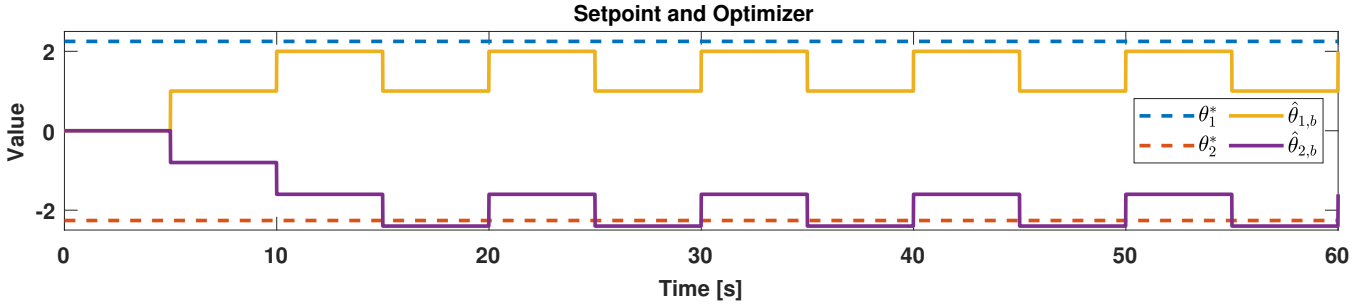

Fig. 9: Minimizer and DA-ES setpoint.

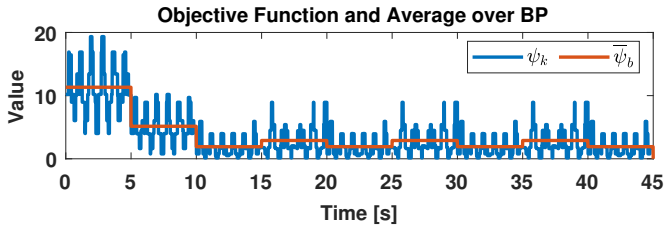

Fig. 10: Objective function, and averaged value over BPs.

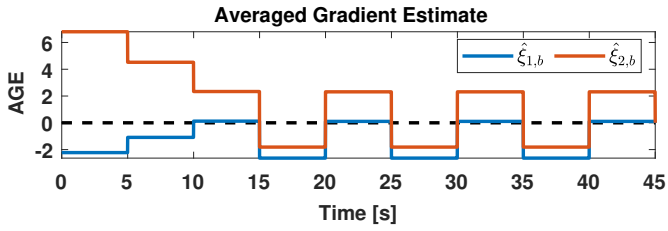

Fig. 11: Averaged gradient estimate over BPs.

channels of the DA-ES.

This simulation highlights both the importance of the initial condition of the DA-ES algorithm, and the limitations of the algorithm, namely as the setpoint update is constrained to fixed length discrete steps in each dimension, the setpoint update is not likely to align with the negative gradient.

Finally, we modify this experiment by implementing the heuristic for decreasing the probe step size and setpoint update size. For this modification the discrete step sizes were changed to $a_1 = 0.25$ and $a_2 = 0.2$. The initial number of step sizes was $\kappa_1 = 4$ and $\kappa_2 = 4$. When (33) was satisfied for $N_s = 3$, $\kappa_1 \leftarrow 1$ and $\kappa_2 \leftarrow 1$. The initial setpoint was $\hat{\boldsymbol{\theta}}_1 = [0, 0]^T$. Simulation results are plotted in Fig. 15, which shows the probe step size decay at 30 seconds, and the setpoint converge to values closer to the minimizer than in the original experiment.

IV. Conclusion

In this work, we propose an Extremum Seeking algorithm that operates over discrete action space, which we call discrete action Extremum Seeking (DA-ES). We first introduce the DA-ES algorithm, and describe its operation, in Section II. Section II-C details the stability and convergence properties of DA-ES. Next, we discuss an extensions to DA-ES, such as one that implements a heuristic to decrease the discrete step size amplitude as the setpoint converges to the optimal value. Section III presents three experiments in which an DA-ES algorithm optimizes its setpoint over a discrete action space and minimizes an the value of a convex objective function. These simulations show the efficacy of the

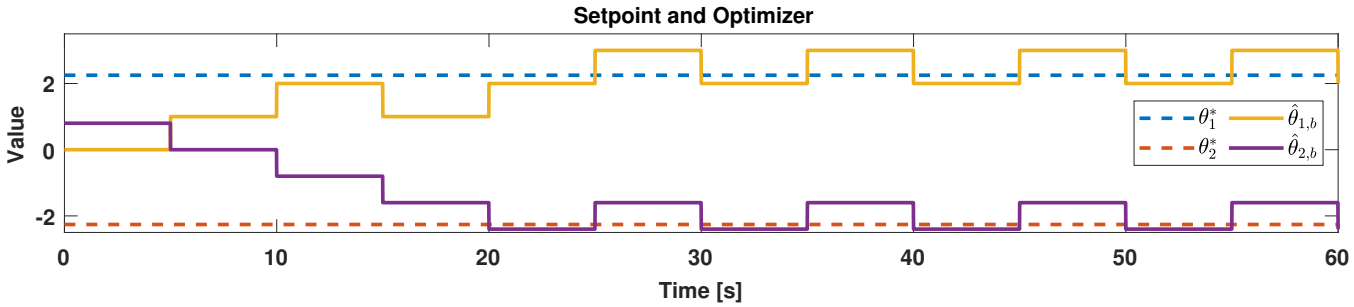

Fig. 12: Minimizer and DA-ES setpoint.

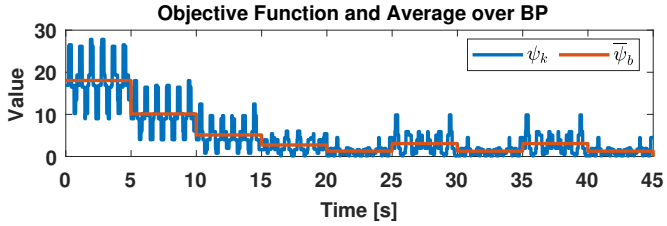

Fig. 13: Objective function, and averaged value over BPs.

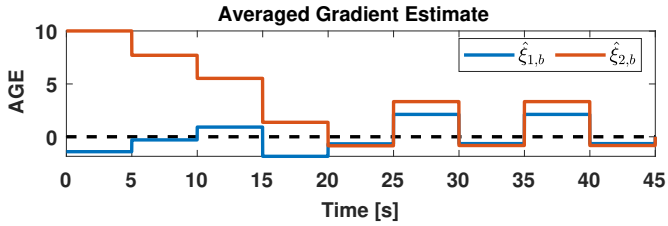

Fig. 14: Averaged gradient estimate over BPs.

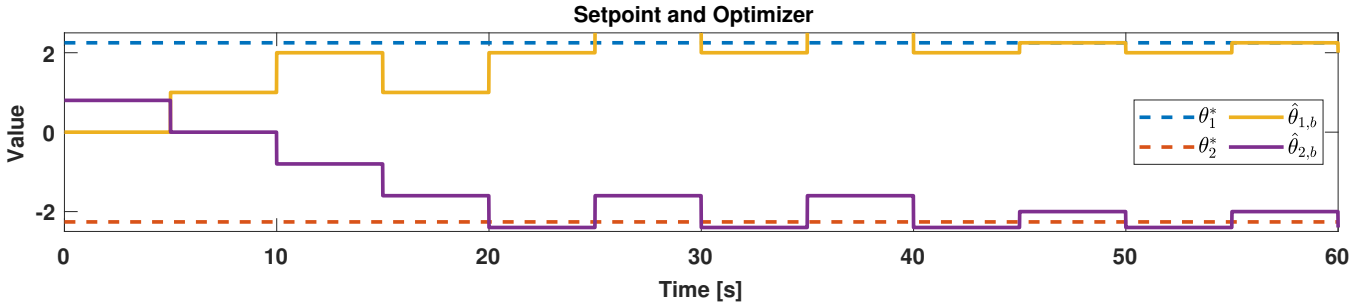

Fig. 15: Minimizer and DA-ES setpoint for experiment with decreasing probe amplitude and setpoint update step size.

DA-ES approach for single and multidimensional discrete optimization.

We plan to continue and extend this work in several key areas. First, we plan to examine DA-ES which considers both the magnitude and sign of the gradient estimate in its setpoint update process. We plan to examine DA-ES over single and multidimensional nonuniform discrete action spaces. We then plan to examine the interaction between multiple separate asynchronous DA-ES operating in parallel to minimize a common objective function. Finally, we plan to examine DA-ES operating in conjunction with ES operating over continuous action spaces to optimize mixed-integer convex programs.

REFERENCES

[1] M. Krstic and H. Wang, "Stability of extremum seeking feedback for general nonlinear dynamic systems," *Automatica*, vol. 36, no. 4, pp. 595 – 601, 2000.

[2] 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. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378475410003290

[3] S. Sengupta, S. Basak, and R. Peters, "Particle swarm optimization: A survey of historical and recent developments with hybridization perspectives," *Machine Learning and Knowledge Extraction*, vol. 1, no. 1, pp. 157–191, Oct 2018. [Online]. Available: http://dx.doi.org/10.3390/make1010010

[4] J.-Y. Choi, M. Krstic, K. B. Ariyur, and J. S. Lee, "Extremum seeking control for discrete-time systems," *IEEE Transactions on automatic control*, vol. 47, no. 2, pp. 318–323, 2002.

[5] A. Scheinker and D. Scheinker, "Bounded extremum seeking with discontinuous dithers," *Automatica*, vol. 69, pp. 250–257, 2016.

[6] L. Wang, S. Chen, and K. Ma, "On stability and application of extremum seeking control without steady-state oscillation," *Automatica*, vol. 68, pp. 18–26, 2016.

[7] J. Lai, J. Xiong, and Z. Shu, "Model-free optimal control of discrete-time systems with additive and multiplicative noises," *arXiv preprint arXiv:2008.08734*, 2020.

[8] D. B. Arnold, M. D. Sankur, M. Negrete-Pincetic, and D. S. Callaway, "Model-free optimal coordination of distributed energy resources for provisioning transmission-level services," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 817–828, 2017.

[9] M. D. Sankur, R. Dobbe, A. von Meier, and D. B. Arnold, "Model-free optimal voltage phasor regulation in unbalanced distribution systems," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 884–894, 2019.

[10] A. Scheinker and M. Krstić, "Non-c2 lie bracket averaging for non-smooth extremum seekers," *Journal of Dynamic Systems, Measurement, and Control*, vol. 136, no. 1, 2014.

[11] M. Sankur and D. Arnold, "Extremum seeking control of distributed energy resources with decaying dither and equilibrium-based switching," in *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.