

RESEARCH ARTICLE

WILEY

Robust data-driven dynamic optimization using a set-based gradient estimator

Ali Kashani¹  | Shirin Panahi¹ | Ankush Chakrabarty² | Claus Danielson¹

¹Department of Mechanical Engineering,
University of New Mexico, Albuquerque,
New Mexico, USA

²Mitsubishi Electric Research
Laboratories, Cambridge, Massachusetts,
USA

Correspondence

Ali Kashani, Department of Mechanical
Engineering, University of New Mexico,
Albuquerque, NM, USA.

Email: kashani@unm.edu

Funding information

National Science Foundation,
Grant/Award Number: CMMI-2105631

Abstract

This article presents an extremum-seeking control (ESC) algorithm for unmodeled nonlinear systems with known steady-state gain and generally non-convex cost functions with bounded curvature. The main contribution of this article is a novel gradient estimator, which uses a polyhedral set that characterizes all gradient estimates consistent with the collected data. The gradient estimator is posed as a quadratic program, which selects the gradient estimate that provides the best worst-case convergence of the closed-loop Lyapunov function. We show that the polyhedral-based gradient estimator ensures the stability of the closed-loop system formed by the plant and optimization algorithm. Furthermore, the estimated gradient provably produces the optimal robust convergence. We demonstrate our ESC controller through three benchmark examples and one practical example, which shows our ESC has fast and robust convergence to the optimal equilibrium.

KEYWORDS

data-driven control, extremum seeking control, gradient descent, real-time optimization, robust control

1 | INTRODUCTION

Extremum-seeking controller (ESC) is a century-old¹ form of adaptive control for real-time optimization wherein the closed-loop dynamic system is driven towards an optimal equilibrium which is learned in real-time from online data. Typically, ESC is considered a form of *model-free* adaptive control. This advantageous property of ESC allows its application to a wide range of unknown systems with unknown costs. However, in many applications, the system and costs are not completely unknown. For instance, in adaptive optics, the dynamics of the deformable mirror and the *structure* of the power-density function are known.² Likewise, a drone using ESC to search for the source of a chemical leak will have known dynamics.³ In this article, we consider ESC for systems with known steady-state gains. Furthermore, we will assume that the cost is strictly a function of the measured system outputs for example, a Hammerstein-Wiener model (although we consider nonlinear dynamics). Both of these assumptions are consistent with the applications described above for example, the concentration of pollutant will be a function of the position of the drone, and power-in-fiber will depend on the configuration of the deformable mirror. Furthermore, these assumptions are consistent with many closed-loop systems designed for reference tracking.^{4–7} Although model-free ESC can be applied to these systems, the performance of the algorithm can be improved by exploiting this partial knowledge.

ESC has other beneficial properties that are important for applications with partial system and cost knowledge. While the structure of the cost function may be known, it could be parameterized by unknown exogenous signals. For instance,

in adaptive optics, the optimal configuration of the deformable mirror depends on unknown (and typically unmeasurable) atmospheric conditions.⁵ For leak detection, the pollutant concentration depends on the unknown location of the leak.⁶ Thus, these applications can benefit from ESC's ability to perform real-time data-driven optimization. Furthermore, ESC accounts for the fact that the data is collected from a dynamic system that is not necessarily in equilibrium. Indeed, our ESC algorithm will explicitly account for tracking errors in the gradient estimation using the aforementioned measured outputs. Finally, ESC accounts for the feedback-loop created by interconnecting a dynamic system with an iterative optimization algorithm.

Over the past two decades, numerous ESC algorithms have been introduced, including perturbation-based, sliding mode, fractional order, Newton-based, and gradient-based ESC algorithms.⁸ These algorithms have been effectively applied in various fields such as air conditioning systems,⁹ braking systems,¹⁰ wind and solar energy systems,¹¹ photovoltaics,¹² plasma control,¹³ and biochemical processes.¹⁴ The convergence rate of these algorithms varies significantly. Several accelerators for ESC are proposed for systems with unknown dynamics, in Reference 15 by high-frequency dither signals, in Reference 16 by unknown Hessian, in Reference 17 by event-triggered mechanisms, and in Reference 18 by dither-free methods. Other different approaches include ESC for affine nonlinear systems,¹⁹ and Newton-based Hessian-free ESC for convex cost functions.^{20,21} Studies suggest that gradient-based ESC potentially increases the convergence rate of perturbation-based ESC.²²

The standard gradient descent ESC for example, Reference 23 is a discrete-time integrator and the set-point in classical ESC for example, References 24 and 25 is the continuous-time integral of the estimated gradient. However, this integral action can potentially destabilize a system that is initially stable (see the illustrative example in Reference 3). Therefore, ensuring stability is a significant concern in ESC, often addressed using methods such as proportional-integral method.²³ Stability in proportional-integral ESC is typically maintained either by reducing the controller's aggressiveness or by improving the accuracy of gradient descent estimation. Decreasing aggressiveness in stable open-loop systems often necessitates slower dynamics in the closed-loop system. Meanwhile, enhancing gradient estimation accuracy can be achieved through methods like dither amplitude control,²⁶ which has been extended to dither adaptation and higher-order sliding modes.²⁷ Formal guarantees of asymptotic stability of ESC is studied using decaying dither.^{28,29} In this article, a joint Lyapunov function for the plant and our ESC controller is employed to ensure input-to-state stable (ISS) stability.

The proposed ESC has an explore-exploit structure wherein there are distinct modes for gathering data and using this data to improve the cost. This explore-exploit structure is common in ESC algorithms.^{3,4,30} However, this structure is typically a heuristic where the ESC algorithm is explicitly designed to gather data for a fixed amount of time before exploiting the data. In contrast, our explore-exploit structure is an emergent property of our game-theoretic analysis of our joint Lyapunov function. The ESC only enters the exploitation mode when the data is sufficiently informative to confidently decrease the joint Lyapunov function.

One of the main contributions of this article is a novel set-based gradient estimation algorithm. While dither adaptation for efficiently probing gradient data has been thoroughly examined in previous studies,^{26,27,29} this data is processed by generic estimation algorithms, such as batch least-squares (BLS), to estimate the gradient. We prove that for costs with bounded curvature, the set of possible gradients that are consistent with the gathered data forms a polyhedron. For a persistently exciting (PE) dither, this set is bounded that is, a polytope instead of a polyhedron. Our estimator selects the estimated gradient from this set that optimizes the worst-case convergence of the joint Lyapunov function. This is in contrast to typical gradient estimators (e.g., least-squares,^{2,18,23,31–33} ESC filtering methods for example, References 34 and 35) whose tacit objective is to minimize the estimator errors. This ignores the impact of those errors for example, some small errors can have disproportionately large impacts on convergence, whereas some large errors can actually be beneficial to convergence. Our gradient estimator solves a game that optimizes the worst-case convergence of the joint Lyapunov function. We show that this game can be re-posed as a quadratic programming (QP), allowing for efficient online computation.

This article makes several novel **contributions**:

1. We characterize the set of gradients consistent with the gathered data for cost functions with bounded curvature; this set of gradients is proven to form a polytope.
2. We present a novel optimization-based gradient estimator that optimizes the worst-case convergence of the ESC; this estimator is implemented efficiently by casting it as a quadratic program (QP).

3. We present novel mathematical analysis that proves the convergence of our novel ESC algorithm. We use a game-theoretic analysis to quantify when the ESC has sufficient data to confidently decrease the cost.
4. We formally prove in Corollary 1 that a similar, previous approach³ is more conservative than the novel ESC presented in this article.
5. Using three benchmark numerical examples, we demonstrate that the presented ESC algorithm has superior performance over the previous approach³ despite very loose assumptions on the knowledge of the curvature bounds.

The remainder of the article is organized as follows. In Section 2, we formally define our ESC problem. In Section 3, we present our novel ESC algorithm and prove its convergence. This involves proving that our gradient set characterizes the set of all gradients consistent with the data and proving that our gradient estimator selects the gradient estimate that provides the best worst-case performance. In Section 4, we present numerical examples that demonstrate the efficacy of our ESC algorithm.

1.1 | Notation and definitions

For a vector $v \in \mathbb{R}^n$ and positive definite matrix $M \in \mathbb{R}^{n \times n}$, $\|v\|_M = \sqrt{v^T M v}$ is the weighted 2-norm where the subscript is omitted for the identity matrix $\|v\| = \sqrt{v^T v}$. For a square matrix $M \in \mathbb{R}^{n \times n}$, $\underline{\lambda}(M)$ and $\bar{\lambda}(M)$ denote its smallest and largest eigenvalues respectively and $\|M\| = \sup\{\|Mx\| : \|x\| \leq 1\}$ is the induced 2-norm. A function f is in C^n if the derivatives $f^{(1)}, \dots, f^{(n)}$ exist and are continuous. A function $\alpha : [0, \infty) \rightarrow [0, \infty)$ is class- $\mathcal{K}$, denoted by $\alpha \in \mathcal{K}$, if $\alpha(0) = 0$ and it is strictly increasing. A function $\beta : [0, \infty)^2 \rightarrow [0, \infty)$ is class- $\mathcal{KL}$, denoted by $\beta \in \mathcal{KL}$, if $\beta(\cdot, t) \in \mathcal{K} \forall t > 0$ and $\beta(r, \cdot)$ is continuous and strictly decreasing $\forall r > 0$. A system $x_{t+1} = f(x_t, u_t)$ is ISS if $\|x_t\| \leq \beta(\|x_0\|, t) + \zeta(\sup_t \|u_t\|)$ where $\beta \in \mathcal{KL}$ and $\zeta \in \mathcal{K}$. A signal $u_t \in \mathbb{R}^m$ is PE³⁶ in the interval $t \in [\tau, \tau + T]$, with $T \geq m$, if for every unit vector $w \in \mathbb{R}^m$ there exists an instance $\hat{t} \in [\tau, \tau + T]$ and scalar $\varepsilon > 0$ such that $|w^T u_{\hat{t}}| > \varepsilon \|u_{\hat{t}}\|$. For notational simplicity, θ is used for the gradient $\nabla \mathcal{J}$ of cost $\mathcal{J}$ and time index t is omitted in some places.

2 | DYNAMIC OPTIMIZATION PROBLEM

Our plant is an unmodeled nonlinear discrete-time system,

$$x_{t+1} = f(x_t, u_t), \quad (1a)$$

$$y_t = h(x_t, u_t), \quad (1b)$$

where $x_t \in \mathbb{R}^n$ is the state, $y_t \in \mathbb{R}^m$ is the output, and $u_t \in \mathbb{R}^m$ is the control input at time index $t \in \mathbb{N}$. We make the following assumptions about the plant (1).

Assumption 1 (Plant).

- (a) The plant (1) is controllable, observable, and Lipschitz continuous. Furthermore, each constant input $u_t = \bar{r}$ corresponds to a unique ISS equilibrium state $\bar{x} = \pi(\bar{r})$, where π is Lipschitz continuous.
- (b) The input and output have the same dimension m , and output (1b) asymptotically tracks constant input $u_t = \bar{r}$; that is $y_t \rightarrow \bar{r}$ as $t \rightarrow \infty$ for all $\bar{r} \in \mathbb{R}^m$.

Assumption 1 reflects assumptions made in recent ESC literature^{3,31,32} and is consistent with many industrial applications, where ESC is applied to a closed-loop system with a well-designed tracking controller and sufficient instrumentation. In other words, ESC is implemented as *add-on* to find the optimal reference $r_t \rightarrow \bar{r}^*$ for a system that tracks $y_t \rightarrow r_t$ when commanded $u_t = \bar{r}$ to any reference set point $r_t = \bar{r}$. Here, we are interested in ESC for its beneficial stability and data-driven optimization properties that can be integrated with systems closed-loop by model-free controllers for example, References 37–39.

Although the cost J is considered unknown, we require the following knowledge about J .

Assumption 2 (Cost).

- (a) The bound $\|\nabla J(\bar{r})\| \geq \kappa_3(\|\bar{r} - \bar{r}^*\|)$ holds where $\kappa_3 \in \mathcal{K}$.
- (b) The bounds $\kappa_1(\|\bar{r} - \bar{r}^*\|) \leq J(\bar{r}) - J(\bar{r}^*) \leq \kappa_2(\|\bar{r} - \bar{r}^*\|)$ hold where $\kappa_1, \kappa_2 \in \mathcal{K}$.
- (c) The curvature $\nabla^2 J$ of $J \in \mathcal{C}^2$ is bounded $\underline{H} \leq \nabla^2 J \leq \bar{H}$ by some known $\underline{H}$ and $\bar{H}$.

Assumption 2a,b characterize a class of generally non-convex functions for which gradient descent can be used to find the unique global optimal (see Remark 1 in Reference 40). Assumption 2a implies that driving the cost gradient to zero $\nabla J \rightarrow 0$ results in converging to the optimum $r_t \rightarrow r^*$, which returns the optimal cost by Assumption 2b. Assumption 2a,b will be used to prove the stability of the optimal equilibrium. If the cost J is convex (i.e., $0 \leq \underline{H} \leq \nabla^2 J$) then Assumption 2c implies that Assumption 2a,b. However, these assumptions can hold for non-convex costs J .⁴⁰

The curvature bounds imposed by Assumption 2c ensure that the gradient ∇J is Lipschitz continuous that is, $\underline{H} = -hI$ and $\bar{H} = hI$ implies $\|\nabla J(y_1) - \nabla J(y_2)\| \leq h\|y_1 - y_2\|$. Without Lipschitz continuity it is impossible to bound the estimation error. Our ESC algorithm can exploit more nuanced curvature bounds $\underline{H}, \bar{H}$ than Lipschitz constants h , if available, to produce tighter bounds on the estimated gradient. The bounds $\underline{H}, \bar{H}$ do not need to be tight for the system to converge to optimum. Instead, there is a trade-off between the tightness of these bounds and the amount of data needed to confidently estimate the gradient. Note that the bounds $\underline{H}$ and $\bar{H}$ are not required to be positive definite matrices. Thus, we are *not* assuming that the cost J is convex. Assumption 2c is common in ESC literature.^{20,41,42}

In summary, our ESC solves the following problem.

Problem 1 (ESC). Compute a sequence of inputs u_t that drives the output $y_t \rightarrow \bar{r}^*$ of an unmodeled system (1) to the optimum $\bar{r}^* = \arg \min_y J(y)$ of an unknown cost $J(y_t)$ using real-time measured data $\{y_k, J(y_k)\}_{k=t-T}^t$.

3 | SET-BASED ESC ALGORITHM

We make use of the tracking properties of the unmodeled system from Assumption 1 to reformulate ESC Problem 1. We modify the objective of ESC to driving $r_t \rightarrow \bar{r}^*$ the reference r_t as the input $u_t = r_t$, to the optimum $\bar{r}^*$ that corresponds to the equilibrium

$$(\bar{x}^*, \bar{r}^*) = \arg \min_{\bar{x}, \bar{r}} J(h(\bar{x}, \bar{r})), \quad (2a)$$

$$\text{s.t. } \bar{x} = f(\bar{x}, \bar{r}). \quad (2b)$$

This is sufficient for the cost $J(y_t)$ to converge to its minimum $J(\bar{r}^*)$ as $r_t \rightarrow \bar{r}^*$, since Assumption 1b dictates the output y_t of the system (1) to converge $y_t \rightarrow \bar{r}^*$ to the steady state $\bar{r}^* = h(\bar{x}^*, \bar{r}^*)$ when the input $u_t = r_t = \bar{r}^*$ is constant.

Our ESC is described by Algorithm 1, which updates the reference r_t by using the collected data $\{y_k, J(y_k)\}_{k=t-T}^t$, a real-time history of measured outputs y_k and costs J_k of a horizon T . Algorithm 1 continually switches between two operational modes; exploration and exploitation. In the exploration mode (4), Algorithm 1 perturbs the system using a PE dither signal d_t to improve the estimate $\hat{\theta}$ of the gradient $\theta = \nabla J$ where the reference r_t is the state of the ESC controller. In the exploitation mode (5), Algorithm 1 descends the estimated gradient $\hat{\theta} \approx \nabla J$ toward the optimal equilibrium (2). The operational mode of Algorithm 1 is determined by line 1 which quantifies the informativeness of the collected data $\{y_k, J(y_k)\}_{k=t-T}^t$. Algorithm 1 avoids redundant computation by verifying whether the data is PE, using the information matrix

$$\Lambda = \frac{1}{T} \sum_{k=t-1}^{t-T} \frac{\Delta y_k \Delta y_k^T}{\|\Delta y_k\|^2}. \quad (3)$$

Algorithm 1. Proposed ESC algorithm

- 1: **if** $\underline{\lambda}(\Lambda) < \varepsilon^2$, where $0 < \varepsilon < 1$, and Λ is given by (3),
or $\alpha \|\hat{\theta}^*\|_{\tilde{H}^{-1}} < \frac{1}{2} \|\Delta y_k\|_{\tilde{H}} + \|e_t\|_{\tilde{H}}$ for any $k \in [t - T, t - 1]$, where $\hat{\theta}^*$ is given by (6),
then
- 2: Explore:

$$r_{t+1} = r_t, \quad (4a)$$

$$u_t = r_t + d_t, \quad (4b)$$

where d_t is PE, and $\|d_t\| \leq \delta$,

- 3: **else**
- 4: Exploit:

$$r_{t+1} = r_t - K\hat{\theta}_t^*, \quad (5a)$$

$$u_t = r_t, \quad (5b)$$

where $\hat{\theta}_t^*$ is given by (6) and K satisfies (9).

- 5: **end if.**

The estimated gradient $\hat{\theta}_t^*$ used in the exploitation mode (5) is provided by the following novel gradient estimator

$$\hat{\theta}_t^* = \arg \min_{\hat{\theta} \in \hat{\Theta}_t} \|\hat{\theta}\|_K^2, \quad (6)$$

where the set of consistent gradient estimates $\hat{\Theta}_t$ is the polyhedron

$$\hat{\Theta}_t = \bigcap_{k=t-T}^{t-1} \left\{ \theta \in \mathbb{R}^m : \underline{\omega}_k \leq \Delta \mathcal{J}_k - \Delta y_k^\top \theta \leq \overline{\omega}_k \right\} \quad (7)$$

for some estimation horizon T , where $\Delta y_k = y_t - y_k$ and $\Delta \mathcal{J}_k = \mathcal{J}(y_t) - \mathcal{J}(y_k)$ are the change in the measured output y_k and measured cost $\mathcal{J}(y_k)$, respectively. The uncertainty bounds are

$$\overline{\omega}_k = \Delta y_k^\top \hat{H} e_t + \frac{1}{2} \Delta y_k^\top \overline{H} \Delta y_k + \frac{1}{2} \|\Delta y_k\|_{\tilde{H}} \|e_t\|_{\tilde{H}}, \quad (8a)$$

$$\underline{\omega}_k = \Delta y_k^\top \hat{H} e_t + \frac{1}{2} \Delta y_k^\top \underline{H} \Delta y_k - \frac{1}{2} \|\Delta y_k\|_{\tilde{H}} \|e_t\|_{\tilde{H}}, \quad (8b)$$

where $e_t = r_t - y_t$ is the tracking error, $\hat{H} = \frac{1}{2}(\overline{H} + \underline{H})$ is the median curvature and $\tilde{H} = \overline{H} - \underline{H} > 0$ is the over-estimated range of curvature $\nabla^2 \mathcal{J}$ for the cost $\mathcal{J}$. The source of uncertainty when estimating the gradient $\nabla \mathcal{J}$ is due to the uncertain (but bounded) curvature $\nabla^2 \mathcal{J}$. The set (5) is a polyhedron since it is the intersection of pairs of parallel half-spaces. Thus, the gradient estimator (6) is a QP. We will show that Algorithm 1 only enters the exploitation mode when the direction and amplitude of the estimated gradient is comparable to the actual gradient $\nabla \mathcal{J}$.

The estimation horizon T is a tuning parameter for the gradient estimator (6). To ensure that the gradient set (7) remains bounded, the estimation horizon $T \geq n$ should be greater than or equal to the state dimension n , and the data should be persistently exciting. Increasing T reduces the volume of the set (7) providing more accurate gradient estimates and improving convergence. However, the trade-off is that the number of constraints defining the polytopic set (7) increases, increasing the complexity of the QP (6) used to estimate the gradient.

Remark 1. In Reference 3, a time-varying adaptive step-size was proposed to ensure that a BLS estimated gradient is contained inside an ellipsoidal gradient set. In Corollary 1, we will show that the previous ellipsoidal set is an outer-approximation of the polytopic gradient set (7). Furthermore, our numerical examples will

demonstrate that the optimization-based gradient estimator (6) with the more accurate polytopic gradient set (7) improves performance.

The positive definite controller gain $K = K^\top > 0$ is designed for the controller (5) to provide nominal ISS stability for the closed-loop system (1) and (5) when $\hat{\theta} = \nabla J$. A sufficient condition is the inequality

$$K - K(\bar{H} + \gamma I)K \geq 0, \quad (9)$$

where $\gamma > 0$ is a tuning parameter. The gain condition (9) will be derived from our stability analysis in Proposition 1 in Section 3.2. Additionally, Section 3.2 will discuss tuning γ to promote robustness.

The main **contribution** of this article is the novel gradient estimator (6), which replaces for example, the least-squares (LS) estimators used in other ESC. In Reference 3 and 23, it was shown that LS estimators have bounded estimation errors $\nabla J - \hat{\theta}$ given by an ellipsoidal set $\hat{\Theta}_t^{\text{LS}}$. Thus, $\nabla J \in \hat{\Theta}_t^{\text{LS}} = \hat{\theta} + \hat{\Theta}_t^{\text{LS}}$.

In contrast, we will show that (7) characterizes *all* gradient estimates $\hat{\theta} \in \hat{\Theta}$ consistent with the collected data $\{y_k, J(y_k)\}_{k=t-T}^t$ and Assumption 2 that is, $\nabla J \in \hat{\Theta}_t$. We will show our estimator (6) is strictly less conservative than LS since the circumscribed ellipsoids in the LS are necessarily conservative outer-approximations of polyhedral that is, $\hat{\Theta}_t \subset \hat{\Theta}_t^{\text{LS}}$ is a tighter bound on the true gradient $\nabla J \in \hat{\Theta} \subset \hat{\Theta}^{\text{LS}}$, where our estimator uses the same information about the system as in the LS estimator.

Another advantage of the novel estimator (6) is that it selects the best worst-case gradient estimate $\hat{\theta} \in \hat{\Theta}$. For a LS, the estimated gradient $\hat{\theta}^{\text{LS}}$ is the center of the ellipsoid $\hat{\Theta}^{\text{LS}}$. This is a good choice in the sense that it minimizes the largest possible estimation error that is, it is the solution to the game

$$\hat{\theta}^{\text{LS}} = \min_{\hat{\theta} \in \hat{\Theta}^{\text{LS}}} \max_{\nabla J \in \hat{\Theta}^{\text{LS}}} \|\nabla J - \hat{\theta}\|^2. \quad (10)$$

In other words, choosing a gradient estimate $\hat{\theta}^{\text{LS}}$ other than the center of $\hat{\Theta}^{\text{LS}}$ would increase the worst-case estimation error $\|\nabla J - \hat{\theta}\|^2$. For a polytope, the analogous choice would be the bary-center $\hat{\theta}^{\text{BC}} = \frac{1}{N} \sum_{i=1}^N \hat{\theta}_i$ where $\hat{\theta}_i$ are the vertices of $\hat{\Theta}$. However, our objective is *not* to minimize the worst-case estimation error (10). Instead, our objective is to minimize the *effects* of the gradient estimation errors on the closed-loop performance. These objectives are not synonymous since not all estimation errors $\tilde{\theta} = \hat{\theta} - \nabla J$ have the same influence on the closed-loop system. We will show that our gradient estimator (6) minimizes the destabilizing influence of gradient estimation errors on the closed-loop system (1) and Algorithm 1 (see Theorem 2).

3.1 | Polytopic gradient estimation set $\hat{\Theta}$

The following theorem shows that (7) characterizes *all* gradient estimates $\hat{\theta}$ consistent with the collected data $\{y_k, J(y_k)\}_{k=0}^t$ and Assumption 2.

Theorem 1. *Let Assumption 2 hold. Then, the cost gradient $\theta_t = \nabla J(r_t)$ at $r_t = h(\pi(r_t), r_t)$ is contained in the set (7).*

Proof. According to Taylor's theorem, the cost J satisfies

$$J(y_k) = J(y_t) + \Delta y_k^\top \nabla J(y_t) + \frac{1}{2} \Delta y_k^\top \nabla^2 J(z_k) \Delta y_k, \quad (11)$$

where $\Delta y_k = y_k - y_t$. Note that (11) holds by Taylor's theorem where the second-order Lagrange remainder $\frac{1}{2} \Delta y_k^\top \nabla^2 J(z_k) \Delta y_k$ is evaluated at an unknown point $z_k = \mu_k y_t + (1 - \mu_k) y_k$ for some $\mu_k \in [0, 1]$. The desired gradient $\nabla J(r_t)$ evaluated at the current reference r_t is related to the gradient $\nabla J(y_t)$ evaluated at $y_t = h(x_t, r_t)$ by the mean-value theorem (see Theorem 12.9 in Reference 43) applied to the scalar function $f(z) = d^\top \nabla J(z)$

$$d^\top \nabla J(r_t) = d^\top \nabla J(y_t) + d^\top \nabla^2 J(z_0)(r_t - y_t) \quad (12)$$

for any direction $d \in \mathbb{R}^m$, where the unknown curvature $\nabla^2 \mathcal{J}(z_0)$ is again evaluated at an unknown point $z_0 = \mu_0 y_t + (1 - \mu_0) y_k$ for some $\mu_0 \in [0, 1]$ which depends on the direction d . Combining (11) and (12) in the direction $d = y_k - y_t$ results

$$\Delta \mathcal{J}_k = \Delta y_k^\top \theta + \underbrace{\frac{1}{2} \Delta y_k^\top H_1 \Delta y_k}_{\omega_1} - \underbrace{e_t^\top H_0 \Delta y_k}_{\omega_2}, \quad (13)$$

where $e_t = r_t - y_t$ is the tracking error. Although ω_1 and ω_2 are unknown, they are bounded since the curvature $H = \nabla^2 \mathcal{J}$ of the cost $\mathcal{J}$ is bounded $\underline{H} \leq H \leq \bar{H}$. From the definition of positive definite matrices, the quadratic-form $\omega_1 = \frac{1}{2} \Delta y_k^\top H_1 \Delta y_k$ is contained in the line interval

$$\Omega_1 = \frac{1}{2} \left[\Delta y_k^\top \underline{H} \Delta y_k, \quad \Delta y_k^\top \bar{H} \Delta y_k \right],$$

where $H_1 - \underline{H} \geq 0$ and $\bar{H} - H_1 \geq 0$. Here Ω_1 is the tightest possible bound on the uncertainty ω_1 .

Deriving the bounds Ω_2 on ω_2 is more complex since ω_2 is not a quadratic-form. Since the linear function $f(H) = e_t^\top H \Delta y_k$ is continuous, the image $\Omega_2 = f(\mathcal{H})$ of the connected set $\mathcal{H} = \{H : \underline{H} \leq H \leq \bar{H}\}$ is connected. Furthermore, since $\omega_2 \in \mathbb{R}$ is a scalar, this set Ω_2 is a line interval,

$$\Omega_2 = \left[\min_{\underline{H} \leq H \leq \bar{H}} e_t^\top H \Delta y_k, \quad \max_{\underline{H} \leq H \leq \bar{H}} e_t^\top H \Delta y_k \right], \quad (14)$$

where the semi-definite programs ensure that Ω_2 is the tightest bound on ω_2 . According to Theorem 2.2 from Reference 44, the semi-definite programs (14) have closed-form solution

$$\min_{\underline{H} \leq H \leq \bar{H}} \frac{1}{2} \text{Tr}(CH) = \frac{1}{2} \text{Tr}^- \left(\tilde{H}^{1/2} C \tilde{H}^{1/2} \right) + \frac{1}{2} \text{Tr} (C \underline{H}),$$

where $C = \Delta y_k e_t^\top + e_t \Delta y_k^\top$ is the symmetric cost matrix, and Tr^- is the trace of the projection of a matrix into the negative semidefinite cone, or more simply, the sum of the negative eigenvalues of a matrix. The matrix $\tilde{H}^{1/2} C \tilde{H}^{1/2}$ is rank-2, and it has exactly one negative eigenvalue $\lambda_- = e_t^\top \tilde{H} \Delta y_k - \|\Delta y_k\|_{\tilde{H}} \|e_t\|_{\tilde{H}}$. Thus,

$$\begin{aligned} \min_{\underline{H} \leq H \leq \bar{H}} \frac{1}{2} \text{Tr}(CH) &= \frac{1}{2} \Delta y_k^\top \tilde{H} e_t - \frac{1}{2} \|\Delta y_k\|_{\tilde{H}} \|e_t\|_{\tilde{H}} + e_t^\top \underline{H} \Delta y_k \\ &= \Delta y_k^\top \hat{H} e_t - \frac{1}{2} \|\Delta y_k\|_{\tilde{H}} \|e_t\|_{\tilde{H}}. \end{aligned}$$

where $\bar{H} - \underline{H} > 0$. Similarly, the upper-bound of Ω_2 in (14) can be derived as $\omega_2 \leq \Delta y_k^\top \hat{H} e_t + \|\Delta y_k\|_{\tilde{H}} \|e_t\|_{\tilde{H}}$ that results

$$\Omega_2 = \Delta y_k^\top \hat{H} e_t + \frac{1}{2} \left[-\|\Delta y_k\|_{\tilde{H}} \|e_t\|_{\tilde{H}}, \|\Delta y_k\|_{\tilde{H}} \|e_t\|_{\tilde{H}} \right].$$

The bounds (8) on the total uncertainty $\omega = \omega_1 + \omega_2$ are given by $\Omega_1 \oplus \Omega_2$ where $\oplus$ is the Minkowski sum. For line-intervals, the Minkowski sum is also a line-interval with bounds (8). Thus, the gradient set (7) describes set of gradient estimates $\hat{\theta}$ that satisfy (13) for some $\omega = \omega_1 + \omega_2 \in \Omega_1 \oplus \Omega_2$. ■

Theorem 1 shows that (7) characterizes all consistent gradients under Assumption 2. If the curvature $\nabla^2 \mathcal{J}$ of the cost $\mathcal{J}$ were known and constant, then each data-point $(\Delta y_k, \Delta \mathcal{J}_k)$ would restrict the gradient to an $m - 1$ dimensional affine subspace. Since the cost $\mathcal{J}$ and its curvature are unknown, these affine subspaces become strips in $\mathbb{R}^m$ that is, the region between two hyperplanes. The gradient set (7) is the intersection of these strips. This set will be bounded if the data collection points $\{y_k\}_{k=t-T}^t$ contain an affinely independent subset for example, if Δy_k is PE. Since the bounds (8) depend on the squared norm $\|\Delta y_k\|^2$, smaller PE signals will produce a smaller gradient set (7). In addition, the tracking error

$e_t = y_r - r_t$ affects the size of the set (7) since we are estimating the gradient $\nabla J(r_t)$ at r_t . These results are consistent with the ellipsoidal outer-approximation $\hat{\Theta}^{\text{LS}}$ derived in Reference 3. Next, we show $\hat{\Theta} \subset \hat{\Theta}^{\text{LS}}$.

Corollary 1. *The estimated polyhedral gradient set $\hat{\Theta}$ from (7) is strictly more accurate than the estimated ellipsoidal set $\hat{\Theta}^{\text{LS}}$ derived in Lemma 1 from Reference 3.*

Proof. From Lemma 1 from Reference 3, the cost gradient ∇J belongs to the ellipsoid

$$\hat{\Theta}^{\text{LS}} = \{\theta : \|\Lambda_{LS}(\theta - \hat{\theta}^{\text{LS}})\|^2 \leq 1\}, \quad (15)$$

where $\Lambda_{LS} = \frac{1}{T} \sum_{k=t-1}^{t-T} w_k \Delta y_k \Delta y_k^\top$, with the weights $w_k = \left(\frac{1}{2} \|\Delta y_k\| \|\Delta y_k\|_{\hat{H}} \left(\|e_t\|_{\hat{H}} + \frac{1}{2} \|\Delta y_k\|_{\hat{H}} \right) \right)^{-1}$, and LS estimator

$$\hat{\theta}^{\text{LS}} = \frac{\Lambda_{LS}^{-1}}{T} \sum_{k=t-1}^{t-T} w_k \Delta y_k \underbrace{\left(\Delta J_k + \Delta y_k^\top \hat{H} \left(e_t - \frac{1}{2} \Delta y_k \right) \right)}_{\text{LS estimator}}, \quad (16)$$

which is the center of the ellipsoid $\hat{\Theta}^{\text{LS}}$ (15).

For $\hat{\theta} \in \hat{\Theta}$ in (7), it can be shown that $\hat{\Theta}$ can be represented by the following form ((23) from Reference 3)

$$\underbrace{\Delta J + \Delta y_k^\top \hat{H} \left(e_t - \frac{1}{2} \Delta y_k \right)}_{\text{LS estimator}} = \Delta y_k^\top \hat{\theta} + \frac{1}{w_k \|\Delta y_k\|} v_k,$$

where $v_k \in [-1, 1]$ is the normalized noise. Substituting this in $\hat{\theta}^{\text{LS}}$ (16) yields

$$\hat{\theta}^{\text{LS}} = \underbrace{\Lambda_{LS}^{-1} \left(\frac{1}{T} \sum_{k=t-1}^{t-T} w_k \Delta y_k \Delta y_k^\top \right)}_{\Lambda_{LS}} \hat{\theta} + \frac{\Lambda_{LS}^{-1}}{T} \sum_{k=t-1}^{t-T} \frac{\Delta y_k}{\|\Delta y_k\|} v_k.$$

Thus, $\hat{\theta} \in \hat{\Theta}^{\text{LS}}$ since

$$\|\Lambda_{LS}(\theta - \hat{\theta}^{\text{LS}})\|^2 = \left\| \frac{1}{T} \sum_{k=t-1}^{t-T} \frac{\Delta y_k}{\|\Delta y_k\|} v_k \right\|^2 \leq 1.$$

Therefore, we showed $\hat{\Theta} \subseteq \hat{\Theta}^{\text{LS}}$ that is, the ellipsoid $\hat{\Theta}^{\text{LS}}$ circumscribes $\hat{\Theta}$. Finally, we show the strict inclusion $\hat{\Theta} \subset \hat{\Theta}^{\text{LS}}$. Consider two adjacent vertices $\hat{\theta}_1, \hat{\theta}_2 \in \hat{\Theta}$ of $\hat{\Theta}$. The mid-point $0.5\hat{\theta}_1 + 0.5\hat{\theta}_2$ will lie on the boundary of $\hat{\Theta}$ but strictly inside the ellipsoid $\hat{\Theta}^{\text{LS}}$. Thus, $\hat{\Theta}$ is a strictly less conservative than the least-squares estimation set $\hat{\Theta} \subset \hat{\Theta}^{\text{LS}}$. ■

3.2 | Convergence under the exploitation mode

In this section, we show that the exploitation mode (5) of Algorithm 1 stabilizes the optimal equilibrium (2) despite the uncertainty of estimated gradient $\hat{\theta} \neq \nabla J \in \hat{\Theta}$. First, we show that the exploitation controller (5) stabilizes the equilibrium (2) under the idealistic condition where the gradient is known.

Proposition 1. *Let Assumptions 1 and 2 hold. Let $\nabla J(r_t)$ be known. Let the controller gain K satisfy (9) for some $\gamma > 0$. Then, the optimal equilibrium (2) is asymptotically stable for the plant (1) in closed-loop with the exploitation controller (5) using $\hat{\theta}_t^* = \nabla J(r_t)$.*

Proof. The full proof can be found in Reference 3. Here, we provide a sketch of the proof since subsequent results will follow similar arguments. We define reference error $\tilde{r} = r - \bar{r}^*$, output error $\tilde{x} = x - \pi(r)$, and a Lyapunov function of the form

$$V(\tilde{x}, \tilde{r}) = \frac{\gamma}{\rho} V_x(\tilde{x}) + V_r(\tilde{r}), \quad (17)$$

by some $\gamma > 0$, and $\rho > 0$, where V_x and V_r are Lyapunov functions for the plant (1), and dynamic controller (5), respectively. Then, by the converse Lyapunov function theorem, we show

$$\begin{aligned} V_x(\tilde{x}^+) - V_x(\tilde{x}) &\leq -q_x^0(\|\tilde{x}\|) + \frac{c}{2} \hat{\theta}^\top K G^\top (\Gamma + P) G K \hat{\theta} \\ &\leq -q_x^0(\|\tilde{x}\|) + \frac{\rho}{2} \hat{\theta}^\top K^2 \hat{\theta}, \end{aligned} \quad (18)$$

where $q_x^0 = q_x - \frac{1}{2} \nabla V_x^\top \Gamma^{-1} \nabla V_x \in \mathcal{K}$ by $q_x \in \mathcal{K}$ and proper adjustment of the positive definite $\Gamma > 0$, the matrix G is the Lipschitz constant matrix as $\|\Delta \pi(r)\| \leq \|G \Delta r\|$, the positive definite $P \geq \nabla^2 V_x$ is bound on the curvature of V_x , scalar $c = \|\Gamma + P\| \|(\Gamma + P)^{-1}\|$ is the condition number of $(\Gamma + P)$, and scalar $\rho = c \bar{\lambda}(G^\top (\Gamma + P) G)$. For V_r , we use the unknown cost J of the form

$$V_r(\tilde{r}) = J(\tilde{r} + \bar{r}^*) - J(\bar{r}^*), \quad (19)$$

for which we use Taylor's theorem to get

$$V_r(\tilde{r}^+) - V_r(\tilde{r}) \leq -\nabla J(r)^\top K \hat{\theta} + \frac{1}{2} \hat{\theta}^\top K \bar{H} K \hat{\theta}, \quad (20)$$

where $r^+ - r = -K \hat{\theta}$ by the controller dynamics (5). By combining (18) and (20), and denoting $\theta_t = \nabla J(r_t)$, the combined Lyapunov function (17) satisfies

$$\begin{aligned} \Delta V &\leq -q_x^1(\|\tilde{x}\|) - \theta^\top K \hat{\theta} + \frac{1}{2} \hat{\theta}^\top K (\bar{H} + \gamma I) K \hat{\theta} \\ &\leq -q_x^1(\|\tilde{x}\|) - \theta^\top K \hat{\theta} + \frac{1}{2} \hat{\theta}^\top K \hat{\theta}, \end{aligned} \quad (21)$$

where $\Delta V = V(\tilde{x}^+, \tilde{r}^+) - V(\tilde{x}, \tilde{r})$, $q_x^1 = \frac{\gamma}{\rho} q_x^0 \in \mathcal{K}$, and gain K follows (9). Since $\hat{\theta}_t^* = \hat{\theta} = \theta = \nabla J(r_t)$ is known here, (21) follows

$$\Delta V \leq -q_x^1(\|\tilde{x}\|) - q_r^0(\|\tilde{r}\|), \quad (22)$$

where $q_r^0(\|\tilde{r}\|) = \frac{1}{2} \underline{\lambda}(K) \kappa_3(\|\tilde{r}\|) \in \mathcal{K}$. Thus $(\tilde{x}, \tilde{r}) \rightarrow 0$ as $t \rightarrow \infty$ by Lyapunov's direct method. ■

The proof of Proposition 1 provides insight on choosing the controller gain (9). The scalar γ trades-off convergence of the plant and controller Lyapunov functions. A small γ can cause overshoot as the plant lags tracking the estimated optimal set-point r_t .

The estimated gradient (6) will not necessarily match $\hat{\theta} \neq \nabla J$ the actual gradient ∇J . In principle, this could destabilize the closed-loop (1) and Algorithm 1. However, we will show that our ESC is robust to gradient estimation errors. To analyze the robustness, we will adopt a game-theoretic approach. Consider the following two-player zero-sum game

$$\hat{\theta} = \arg \min_{\hat{\theta} \in \hat{\Theta}} \left(\max_{\theta \in \hat{\Theta}} \frac{1}{2} \|\hat{\theta}\|_K^2 - \theta^\top K \hat{\theta} \right), \quad (23)$$

where the compact polytope $\hat{\Theta}$ was defined in (7). The inner-optimization in (23) is our *adversary* which attempts to destabilize the system. The adversary selects the “actual” gradient $\theta = \nabla J \in \hat{\Theta}$ (consistent with the data) that maximizes the Lyapunov function (21). Here, the adversary has the advantageous position of selecting the worst-case gradient $\nabla J =$

$\kappa(\hat{\theta})$ as a function of our estimated gradient $\hat{\theta} \in \hat{\Theta}$. In the outer-optimization in (23) we select the estimate $\hat{\theta} \in \hat{\Theta}$ that provides the *best worst-case* change in the Lyapunov function (21). This adversarial perspective ensures that (23) provides the best worst-case convergence, ensuring that the ESC is robust to estimation errors. The game (23) is distinct from the game (10), which is tacitly solved in LS estimators. Instead of selecting our gradient estimate $\hat{\theta}$ to minimize the magnitude $\|\theta - \hat{\theta}\|$ estimation errors, we select it to minimize the *effect* (21) of the estimation errors.

We will show that the gradient $\hat{\theta}$ estimated by our novel estimator (6) is also the solution of the game (23). First, we must prove the following lemma.

Lemma 1. *The inequality $\theta^\top K \hat{\theta}^* \geq \|\hat{\theta}^*\|_K^2$ holds $\forall \theta \in \hat{\Theta}$.*

Proof. The Lagrangian of the gradient estimation problem (6) is $L(\hat{\theta}, \lambda) = \frac{1}{2} \hat{\theta}^\top K \hat{\theta} + \lambda^\top (H\hat{\theta} - K)$ where $\hat{\Theta} = \{\hat{\theta} : H\hat{\theta} \leq K\}$ is the half-space representation of the polytopic set (7). The stationarity and dual-feasibility optimality conditions state

$$\nabla L(\hat{\theta}^*, \lambda^*) = K_A \hat{\theta}^* + H_A^\top \lambda_A^* = 0, \quad \lambda_A^* \geq 0,$$

where K_A and H_A describe the subset of constraints active at the optimal. By Farkas' lemma, this means that there does *not* exist $\hat{\theta}^* - \theta$ such that $H_A(\hat{\theta}^* - \theta) = K_A - H_A\theta \geq 0$ and $-(\hat{\theta}^* - \theta)^\top K \hat{\theta}^* < 0$. In other words, $\theta^\top K \hat{\theta}^* \geq \|\hat{\theta}^*\|_K^2$ for all $\theta \in \{\theta : H_A\theta \leq K_A\} \supseteq \Theta$. ■

Lemma 1 shows that the estimator (6) inherently balances direction and amplitude of the estimated gradient that is, a bigger $\hat{\theta}^*$ will have a smaller angle $\propto \theta^\top K \hat{\theta}^*$ with the actual gradient $\theta = \nabla J$. Therefore, big changes to the reference r will only be made when the gradient direction can be confidently estimated. As a result, we can show the cost J is strictly decrescent. By Taylor's theorem

$$J(r^+) - J(r) \leq -\theta^\top K \hat{\theta}^* + \frac{1}{2} \|\hat{\theta}^*\|_{KHK}^2 \leq -\frac{1}{2} \|\hat{\theta}^*\|_K^2 - \frac{\gamma}{2} \|\hat{\theta}^*\|_{K^2}^2,$$

where $\theta^\top K \hat{\theta}^* \geq \|\hat{\theta}^*\|_K^2$ by Lemma 1, and gain K follows (9). The following theorem shows that the solution of the QP (6) is also the solution of the robust optimization (23).

Theorem 2. *The gradient $\hat{\theta}$ estimated from (6) is equivalent to the optimal worst-case gradient estimated from (23).*

Proof. We will show that the optimal solution $\hat{\theta}^*$ of (6) is a feasible lower-bound for (23) and therefore its optimizer. In other words, we will prove the inequality

$$f(\hat{\theta}^*) \leq \min_{\hat{\theta} \in \hat{\Theta}} f(\hat{\theta}) \quad (24)$$

where $f(\hat{\theta}) = \max_{\theta \in \Theta} \frac{1}{2} \|\hat{\theta}\|_K^2 - \kappa(\hat{\theta})^\top K \hat{\theta}$ is the cost of (23) with the optimal adversarial strategy $\theta = \kappa(\hat{\theta})$. From Lemma 1, the following inequality holds

$$f(\hat{\theta}^*) = \max_{\theta \in \Theta} \frac{1}{2} \|\hat{\theta}^*\|_K^2 - \theta^\top K \hat{\theta}^* \leq -\frac{1}{2} \|\hat{\theta}^*\|_K^2 = -\min_{\hat{\theta} \in \hat{\Theta}} \frac{1}{2} \|\hat{\theta}\|_K^2, \quad (25)$$

since $\max_{\theta \in \Theta} -\theta^\top K \hat{\theta}^* \leq -\|\hat{\theta}^*\|_K^2$. By the minimax inequality, we have

$$\max_{\hat{\theta} \in \hat{\Theta}} \min_{\hat{\theta}} \frac{1}{2} \|\hat{\theta}\|_K^2 - \theta^\top K \hat{\theta} \leq \min_{\hat{\theta}} \max_{\theta \in \Theta} \frac{1}{2} \|\hat{\theta}\|_K^2 - \theta^\top K \hat{\theta}.$$

The inner-optimization problem on the left-side selects the best gradient estimate $\hat{\theta}$ as a function of the actual gradient θ . Clearly, the optimal solution is $\hat{\theta} = \theta$, which yields the inequality

$$\max_{\hat{\theta} \in \hat{\Theta}} -\frac{1}{2} \|\hat{\theta}\|_K^2 \leq \min_{\hat{\theta}} \max_{\theta \in \Theta} \frac{1}{2} \|\hat{\theta}\|_K^2 - \theta^\top K \hat{\theta} \leq \min_{\hat{\theta} \in \hat{\Theta}} f(\hat{\theta}) \quad (26)$$

by combining (25) and (26) we obtain the optimizer indicating inequality (24) where $\max_{\theta \in \hat{\Theta}} -\frac{1}{2} \|\hat{\theta}\|_K^2 = -\min_{\hat{\theta} \in \hat{\Theta}} \frac{1}{2} \|\hat{\theta}\|_K^2$. ■

Theorem 2 means that we can compute the best worst-case gradient estimate $\hat{\theta}$ by solving a convex optimization problem (6) instead of the min–max game (23). Since the estimated gradient set (7) is a polytope, the gradient estimator (6) is a QP.

Even though QPs have polynomial complexity, solving (6) each time a new data point $(y_t, \mathcal{J}_t)$ is collected will be computationally expensive. Algorithm 1 checks whether the data is insufficient to confidently estimate the gradient. If not, Algorithm 1 enters exploration mode where it keeps $r_t = \bar{r}$ constant so that the system can converge $y \rightarrow r$ to gather local data, while the system is perturbed with a dither signal d_t to tighten the set $\hat{\Theta}$ around $\nabla \mathcal{J} \neq 0$. Intuitively, the estimation is confident if the uncertainty $\hat{\Theta}$ is sufficiently tight compared to the magnitude of $\theta = \nabla \mathcal{J} \in \hat{\Theta}$. This is proven in the next lemma.

Lemma 2. *Let Assumption 2 hold. Let Algorithm 1 be in the exploitation mode. Then, there exists $\alpha_0 > 0$ such that $\|\hat{\theta}_t^*\|_K \geq \alpha_0 \|\theta_t\|_K$.*

Proof. For every $\theta, \hat{\theta} \in \hat{\Theta}$, where $\hat{\Theta}$ is defined in (7), the distance $|\Delta y_k^\top \theta| - |\Delta y_k^\top \hat{\theta}|$ between θ and $\hat{\theta}$ in the Δy_k direction is less than the width $\bar{\omega}_k - \underline{\omega}_k$ of the interval $[\underline{\omega}_k, \bar{\omega}_k] - \Delta \mathcal{J}_k$, or equivalently

$$|\Delta y_k^\top \theta| - |\Delta y_k^\top \hat{\theta}| \leq \bar{\omega}_k - \underline{\omega}_k.$$

We can divide both sides by $\|\Delta y_k\|_{\tilde{H}}$ to get

$$\varepsilon_k \|\theta\|_{\tilde{H}^{-1}} - \|\hat{\theta}\|_{\tilde{H}^{-1}} \leq \left(\frac{1}{2} \|\Delta y_k\|_{\tilde{H}} + \|e_t\|_{\tilde{H}} \right), \quad (27)$$

where we used Cauchy–Schwarz inequality $|\Delta y_k^\top \hat{\theta}| \leq \|\Delta y_k\|_{\tilde{H}} \|\hat{\theta}\|_{\tilde{H}^{-1}}$, and we defined $\varepsilon_k \in [0, 1]$ where $|\Delta y_k^\top \theta| = \varepsilon_k \|\Delta y_k\|_{\tilde{H}} \|\theta\|_{\tilde{H}^{-1}}$. For some instance k , where $\varepsilon_k = \max_{k' \in [t-T, t-1]} \varepsilon_{k'}$ over the interval $k' \in [t-T, t-1]$, we can show $\sqrt{C_{\tilde{H}}} \varepsilon_k \geq \varepsilon$, where $C_{\tilde{H}}$ is the condition number of $\tilde{H}$. Since $\underline{\lambda}(\Lambda) \geq \varepsilon^2$ holds in the exploitation mode by line 1 of Algorithm 1, we have

$$\underline{\lambda}(\Lambda) \leq \frac{1}{T} \sum_{k'=t-1}^{t-T} \frac{\theta^\top \Delta y_{k'} \Delta y_{k'}^\top \theta}{\|\theta\|^2 \|\Delta y_{k'}\|^2} \leq \frac{C_{\tilde{H}}}{T} \sum_{k'=t-1}^{t-T} \varepsilon_{k'}^2 = C_{\tilde{H}} \varepsilon_k^2.$$

Furthermore, since $\frac{1}{2} \|\Delta y_k\|_{\tilde{H}} + \|e_t\|_{\tilde{H}} \leq \alpha \|\hat{\theta}^*\|_{\tilde{H}^{-1}}$ holds according to line 1 of Algorithm 1, (27) yields

$$\frac{\varepsilon}{\sqrt{C_{\tilde{H}}}} \|\theta\|_{\tilde{H}^{-1}} \leq (1 + \alpha) \|\hat{\theta}^*\|_{\tilde{H}^{-1}},$$

where $\hat{\theta} = \hat{\theta}^*$. Therefore, $\|\hat{\theta}^*\|_K \geq \alpha_0 \|\theta\|_K$ holds at time t by equivalence of norms, where $\alpha_0 \geq \varepsilon / ((1 + \alpha) C_{\tilde{H}} \sqrt{C_K}) > 0$, where C_K is the condition numbers of K . ■

Lemma 2 means that if we can force the estimate gradient to converge $\hat{\theta} \rightarrow 0$ (i.e., $\|\hat{\theta}\|_K \rightarrow 0$) then we can force the actual gradient to zero that is, $\|\theta\|_K \leq \frac{1}{\alpha_0} \|\hat{\theta}\|_K \rightarrow 0$. The following corollary shows that this promotes convergence to the optimal equilibrium (2).

Corollary 2. *Let Assumptions 1 and 2 hold. Let the controller gain K satisfy (9). Then, the optimal equilibrium (2) is asymptotically stable for the plant (1) in closed-loop with the Algorithm 1 in the exploitation mode.*

Proof. We prove ISS stability using the Lyapunov function (17) from Proposition 1. The proof is identical to Proposition 1 until the bound (21) on the decrease of the Lyapunov function which assumed the gradient was known $\hat{\theta} = \nabla \mathcal{J}$. Under estimation error $\hat{\theta} \neq \nabla \mathcal{J}$, the best worst-case decrease can be bounded by the game (23)

$$\Delta V \leq -q_x^1(\|\tilde{x}\|) + \max_{\theta \in \hat{\Theta}} \left(\frac{1}{2} \|\hat{\theta}^*\|_K^2 - \theta^\top K \hat{\theta}^* \right). \quad (28)$$

Using Lemma 1, we obtain the bound

$$\begin{aligned} \Delta V &\leq -q_x^1(\|\tilde{x}\|) + \frac{1}{2} \|\hat{\theta}^*\|_K^2 - \min_{\theta \in \hat{\Theta}} \theta^\top K \hat{\theta}^* \\ &\leq -q_x^1(\|\tilde{x}\|) - \frac{1}{2} \|\hat{\theta}^*\|_K^2 \leq -q_x^1(\|\tilde{x}\|) - q_r^1(\|\tilde{r}\|), \end{aligned}$$

where $q_r^1(\|\tilde{r}\|) = \frac{1}{2} \alpha_0 \underline{\lambda}(K) \kappa_3(\|\tilde{r}\|) \leq \frac{1}{2} \alpha_0 \|\theta\|_K^2 \leq \frac{1}{2} \|\hat{\theta}^*\|_K^2$ is class- $\mathcal{K}$ function $q_r^1 \in \mathcal{K}$ by Lemma 2 and Assumption 2a. Therefore, the states $(\tilde{x}, \tilde{r})$ of the closed-loop system converge to the optimal equilibrium (2). ■

Corollary 2 shows that the optimal equilibrium (2) is asymptotically stabilized when Algorithm 1 is in the exploitation mode (5). Since $\hat{\theta}^*$ from (6) also optimizes the game (23), the decrease of the Lyapunov function (28) is maximized, providing the best performance under uncertainty. This convergence relies on the restrictive conditions (line 1 of Algorithm 1) for entering the exploitation mode. In the following section, we will show that it is possible to satisfy these restrictive conditions.

3.3 | Convergence under the ESC algorithm

In this section, we show that Algorithm 1 drives the plant (1) to the optimal equilibrium (2). By Corollary 2, this will occur if Algorithm 1 always returns to the exploitation mode after finite-time when $\nabla J(r_t) \neq 0$. Or equivalently, if Algorithm 1 always leaves exploration mode in finite-time as shown in the following lemma.

Lemma 3. *Let Assumption 1 hold. Let $\theta = \nabla J(r_t) \neq 0$, the dither d_t be pe, and sufficiently small that is, $\|d_t\| \leq \delta$. Then, there is a finite time $\tau < \infty$ such that the system leaves the exploration mode.*

Proof. According to line 1, Algorithm 1 leaves the exploration, if the output Δy_k is PE $\underline{\lambda}(\Lambda) \geq \varepsilon^2$ and the data is sufficiently local that is, $\alpha \|\hat{\theta}^*\|_{\tilde{H}^{-1}} \geq \frac{1}{2} \|\Delta y_k\|_{\tilde{H}} + \|e_t\|_{\tilde{H}}$.

Since the system is controllable, the PE dither d_t will produce a PE state x_t and likewise, the PE state will produce a PE output y_t since the system is observable. Thus, for a estimation horizon T larger than the sum of the controllability and observability indices, Δy_k is PE and $\underline{\lambda}(\Lambda) \geq \varepsilon^2$ will be satisfied.³

For the second condition, $\sqrt{C_{\tilde{H}}} \varepsilon_k \geq \varepsilon$ for some k when Δy is PE according to Lemma 2 (see (27)). Thus, for $\hat{\theta} = \hat{\theta}^*$ (27) yields

$$\|\hat{\theta}^*\| \geq \frac{\varepsilon}{\sqrt{C_{\tilde{H}}}} \|\theta\|_{\tilde{H}^{-1}} - \left(\frac{1}{2} \|\Delta y_k\|_{\tilde{H}} + \|e_t\|_{\tilde{H}} \right).$$

To enter the exploitation mode, we need to show that the right hand side becomes greater than $\frac{1}{\alpha} \left(\frac{1}{2} \|\Delta y_k\|_{\tilde{H}} + \|e_t\|_{\tilde{H}} \right)$, or equivalently

$$\frac{1}{2} \|\Delta y_k\|_{\tilde{H}} + \|e_t\|_{\tilde{H}} \leq \frac{\alpha}{1 + \alpha} \frac{\varepsilon}{\sqrt{C_{\tilde{H}}}} \|\theta\|_{\tilde{H}^{-1}}$$

for $t > \tau - T$ and $k \in [t - T, t - 1]$, so that both conditions of Algorithm 1 line 1 will be violated after finite-time $\tau > T$. To show this, we bound the signals e_t and Δy_k . Since the plant output map (1b) is Lipschitz continuous, we have

$$\|e_t\| = \|y_t - r\| = \|h(x_t) - h(\bar{x})\| \leq \ell_h \|x_t - \bar{x}\|,$$

where ℓ_h is the Lipschitz constant for h and $\bar{x}$ is the equilibrium state corresponding to the reference r . Likewise,

$$\begin{aligned}\|y_k - y_t\| &\leq \|h(x_t) - h(\bar{x}) - (h(x_k) - h(\bar{x}))\| \\ &\leq \ell_h \|x_t - \bar{x}\| + \ell_h \|x_k - \bar{x}\|.\end{aligned}$$

Therefore,

$$\frac{1}{2}\|\Delta y_k\|_{\tilde{H}} + \|e_t\|_{\tilde{H}} \leq \bar{\lambda}(\tilde{H})\left(\frac{1}{2}\ell_h\|x_k - \bar{x}\| + \frac{3}{2}\ell_h\|x_t - \bar{x}\|\right).$$

Since the plant is ISS, there exists a finite time τ and bound δ on the dither amplitude $\|d_t\| \leq \delta$ such that for $t > \tau_0 = \tau - T$,

$$\|x_t - \bar{x}\| \leq \beta(\|x_0 - \bar{x}\|, \tau_0) + \zeta(\delta) < \frac{\varepsilon\alpha\|\theta\|_{\tilde{H}^{-1}}}{2\bar{\lambda}(\tilde{H})\ell_h(1+\alpha)\sqrt{C_{\tilde{H}}}},$$

where $\beta \in \mathcal{KL}$ and $\zeta \in \mathcal{K}$. Therefore, there exists dither amplitude δ such that Algorithm 1 leaves the exploration mode. ■

Algorithm 1 leaves the exploration mode if the tracking error e_t and output “velocity” Δy_k are sufficiently informative and small compare to the magnitude of $\theta = \nabla J$. We note that the switching condition in line 1 of Algorithm 1 does not depend on either T or δ . Thus, we only need to show the existence of these bound to show that Algorithm 1 will not become trapped in the exploration mode.

Finally, we prove that Algorithm 1 drives the plant (1) to the optimal equilibrium (2).

Theorem 3. *Let Assumptions 1 and 2 hold. Let the controller gain K satisfy (9). Let the dither d_t be bounded $\|d_t\| \leq \delta$ and pe. Then the plant (1), controlled by Algorithm 1, is ISS at the equilibrium (2).*

Proof. In the exploitation mode, the combined Lyapunov function (17) follows

$$\Delta V \leq -q_x^1(\|\tilde{x}\|) - q_r^1(\|\tilde{r}\|),$$

according to Corollary 2. In the exploration mode, since the system is ISS,

$$\Delta V \leq -q_x^1(\|\tilde{x}\|) + \sigma^0(\delta),$$

where $\sigma^0(\delta) = \frac{1}{2}\gamma\|(\Gamma + P)\| \|G\|^2 \delta^2 \in \mathcal{K}$.

We use theoretical principles from the switched system literature⁴⁵ for the proof. Let $i \in \mathbb{N}$ be the mode-switching indices, where t_{2i} denote the time-indices that the system enters exploration mode. Then, the combined Lyapunov function follows

$$V(\tilde{x}_{2i+2}, \tilde{r}_{2i+2}) - V(\tilde{x}_{2i}, \tilde{r}_{2i}) \leq -q(\|\tilde{x}_{2i}, \tilde{r}_{2i}\|) + \sigma(\delta), \quad (29)$$

where $\tilde{x}_i = \tilde{x}_{t_i}$, $\tilde{r}_i = \tilde{r}_{t_i}$, and

$$q(\|\tilde{x}_{2i}, \tilde{r}_{2i}\|) = \sum_{t=t_{2i}}^{t_{2i+2}} (q_x^1(\|\tilde{x}_t\|) + \sum_{t=t_{2i+1}}^{t_{2i+2}} q_r^1(\|\tilde{r}_t\|)) \in \mathcal{K},$$

and $\sigma(\delta) = (t_{2i+2} - t_{2i})\sigma^0(\delta) \in \mathcal{K}$. Therefore, by (29) and Proposition 2.3 in Reference 46, there exists $\beta^1 \in \mathcal{KL}$ and $\zeta^1 \in \mathcal{K}$ such that

$$\|\tilde{x}_{2i}, \tilde{r}_{2i}\| \leq \beta^1(\|\tilde{x}_0, \tilde{r}_0\|, 2k) + \zeta^1(\delta).$$

Therefore, the state (x, r) of the closed-loop system of plant (1) controlled by Algorithm 1 converges to a neighborhood of the optimal equilibrium (2) as the exploitation time goes to infinity. By Lemma 2 the system leaves

the exploration mode in finite time. Thus, the total time $\sum_{i=1}^{\infty} T_{2i+1}$ that the system is in the exploitation mode goes to infinity $\sum_{i=1}^{\infty} T_{2i+1} \rightarrow \infty$ as $t \rightarrow \infty$. ■

Theorem 3 proves that the closed-loop systems (1) and Algorithm 1 is ISS with respect to the dither signal d_t . Thus, the closed-loop system converges to a neighborhood of the optimal equilibrium that depends on the dither amplitude. As previously mentioned, asymptotic stability of the optimal equilibrium can be achieved using a decaying dither.^{28,47}

4 | NUMERICAL EXAMPLES

In this section, we illustrate our ESC Algorithm 1 through three benchmark and one practical examples. The numerical results were simulated in MATLAB R2022b using Windows 11 on a laptop with a 2.30 GHz Intel Core i7-12700H processor, 16 GB RAM, and 500 GB SSD.

4.1 | First-order benchmark example

The following example is taken from Reference 18. It features linear first-order dynamics with a non-convex cost. The plant dynamics are

$$\dot{x} = -x + u, \quad (30a)$$

$$y = x, \quad (30b)$$

which satisfy Assumption 1. The unknown cost function is

$$\mathcal{J}(y) = 3 - \frac{1}{\sqrt{1 + (y - 2)^2}}, \quad (31)$$

which is non-convex and only locally satisfies Assumption 2, indicating that assumptions on the cost function are only sufficient and not necessary. We set a sample rate of 10 Hz, an estimation horizon of $T = 5$, bounds $\underline{H} = -2$ and $\bar{H} = 2$ on the curvature $\nabla^2 \mathcal{J}$ of the cost, an initial condition of $x_0 = 100$ and $r_0 = 100$, a dither of $d_t = 0.001 \sin t$, design parameters $\varepsilon = 10^{-4}$, $\alpha = 10^4$, and a gain $K = 0.5I$.

Figure 1 shows our algorithm converges to optimum faster than the previous algorithm from Reference 3. This is a result of using the less conservative polytopic set (7) over the ellipsoidal sets used by LS estimated, as discussed in Corollary 1. Additionally, our algorithm converged in roughly half of the time as the least-squares ESC from Reference 18.

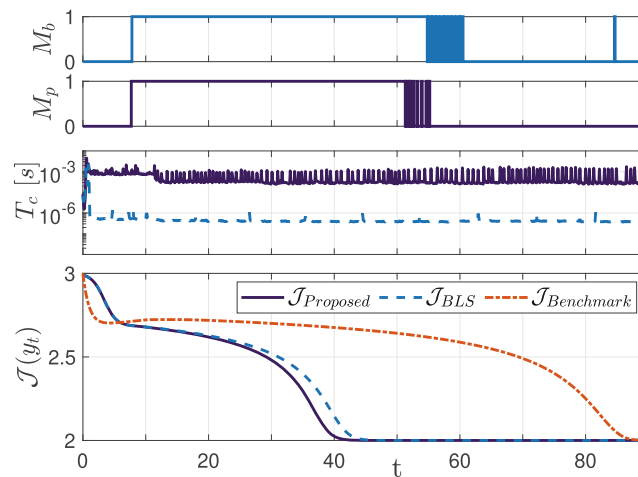


FIGURE 1 Results for the first-order problem; time t is in seconds. $\mathcal{J}$ shows the cost (31) optimized by one-dimensional ESCs, including the presented method, the BLS of Reference 3, and the least squares of Reference 18, where y_t follows (30). M_b and M_p show the operation modes for the BLS and proposed methods, respectively. Value 0 indicates exploration and 1 exploitation. T_c is the computation time.

4.2 | Second-order benchmark example

The following second-order ESC problem is taken from Reference 29. The plant dynamics are

$$\dot{x} = R(x)u + w, \quad (32a)$$

$$y = x, \quad (32b)$$

where $R(x) \in \mathbb{R}^{2 \times 2}$ is a known planar rotation matrix with angle $x_1 + x_2$ and $w(t) = [\sin(2t), \cos(t)]^\top$ is a periodic disturbance. We pre-stabilize the system using the feedback-linearizing input

$$u = -R(x)^\top (F(x - r) - w),$$

where the matrix $F = -10I$ ensures tracking $y_t \rightarrow \bar{r}$, so that the plant (32) satisfies Assumption 1. To preserve nonlinearity and challenge our method, the plant (32) is simulated in continuous-time using MATLAB's `ode45` solver, while the controller is updated in discrete-time. For $t \in [t_k, t_{k+1})$, zero-order hold is applied for the control input $u(t) = u(x_k)$, which is computed using sampled state $x(t_k)$ and disturbance $w(t_k)$. The unknown cost function is considered

$$\mathcal{J}(y) = \|y - 1\|^2 + 2018, \quad (33)$$

which globally satisfies Assumption 2.

We set a sample rate of 20 Hz, an estimation horizon of $T = 10$, an initial condition of $x_0 = 0$ and $r_0 = 0$, and a gain of $K = 0.5I$. We perform two experiments here. First we use fairly over-estimated bounds $\underline{H} = 0I$ and $\bar{H} = 10I$ on the curvature $\nabla^2 \mathcal{J} = 2I$ of the cost, and design parameters $\varepsilon = 10^{-4}$ and $\alpha = 10^4$. For the second experiment, we use loose bounds $\underline{H} = -1000I$ and $\bar{H} = 1000I$ on the curvature $\nabla^2 \mathcal{J}$ of the cost, the same $\varepsilon = 10^{-4}$, and reduced $\alpha = 1$, which increases the confidence of gradient estimation.

Figure 2 shows our polyhedral-based algorithm converges faster than the recent ellipsoidal-based BLS from Reference 3 when the curvature bounds are tight. Again this a consequence of Corollary 1. In addition, the presented ESC outperforms that ESC with adaptive dither from Reference 29. Figure 3 shows our algorithm still converges to the optimum in less than 3 s when the curvature bounds are loosened. In contrast, the BLS from Reference 3 does not converge within the allotted 5 s.

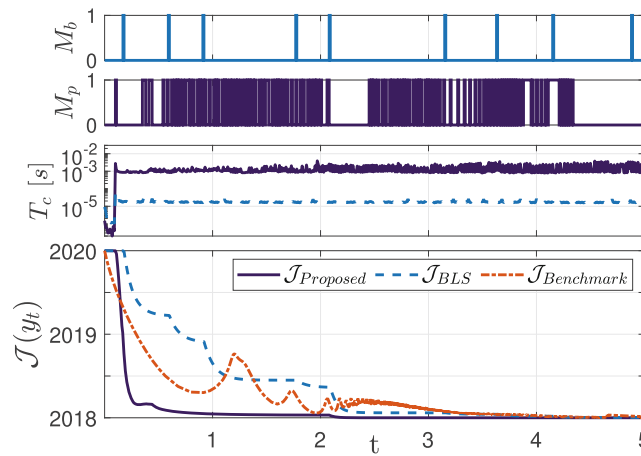


FIGURE 2 Results for the second-order problem; time t is in seconds. $\mathcal{J}$ shows the cost (33) optimized by two-dimensional ESCs, including the presented method, the BLS of Reference 3, and the adaptive dither method of Reference 29, where y_t follows (32). M_b and M_p show the operation modes for the BLS and proposed methods, respectively. Value 0 indicates exploration and 1 exploitation. T_c is the computation time.

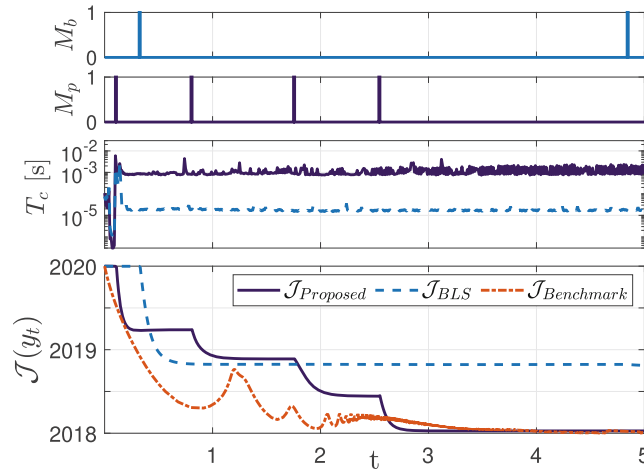


FIGURE 3 Results for the second experiment on the second-order problem. Configurations are similar to those in Figure 2.

4.3 | Third-order benchmark example

The following third-order ESC problem is taken from Reference 28. The plant dynamics are

$$\dot{x}_1 = x_1^2 + u_2^2, \quad (34a)$$

$$\dot{x}_2 = -x_2 + u_1, \quad (34b)$$

$$\dot{x}_3 = -x_3 + u_2 x_2. \quad (34c)$$

Indeed the plant (34) does not satisfy the asymptotic tracking assumption of Assumption 1b. This can be rectified, if the steady-state map of the plant is known, by using the transformation

$$\begin{aligned} \dot{u}_1 &= r_1 / (1 + \sqrt{r_2}), \\ \dot{u}_2 &= -\sqrt{r_2}, \end{aligned}$$

where $r_2 \geq 0$ is enforced by setting $r_2 = 0$ if $r_2 < 0$. However, (34) is only locally Lipschitz continuous, indicating that Assumption 1 about the system is only sufficient but not necessary. Similar to the previous example, the plant (34) is simulated in continuous-time using MATLAB's `ode45` solver, while the controller is updated in discrete-time that is, using the measurements at t_k where zero-order hold is applied for the control inputs for $t \in [t_k, t_{k+1})$. The unknown cost is

$$\mathcal{J}(y) = y_1^2 + 2y_2, \quad (35)$$

where $y_1 = x_2 + x_3$ and $y_2 = x_1 + x_2 - u_1$ are the measured outputs. The cost (35) globally satisfies Assumption 2.

We set a sample rate of 4 Hz, an estimation horizon of $T = 5$, over-estimated bounds $\underline{H} = 0I$ and $\bar{H} = 10I$ on the curvature $\nabla^2 \mathcal{J}$ of the cost, an initial condition of $x_0 = [1, 5, 5]^\top$ and $r_0 = [10, 1]^\top$, design parameters $\varepsilon = 10^{-4}$, $\alpha = 10^4$, a gain of $K = I$, and a dither of $d_t = 0.001[\sin(t), \cos(2t)]^\top$.

Figure 4 shows our algorithm converges faster than the recent BLS from Reference 3 due to the less conservative gradient estimator. Furthermore, the convergence is roughly 1% of the convergence time of the perturbation-based ESC from Reference 28.

4.4 | Practical example: Drone leak inspection

In this section, we apply our ESC algorithm to the problem of autonomously locating the source of gas emissions. This application is useful for dangerous tasks such as sampling material from volcanic craters where line-of-sight limits both

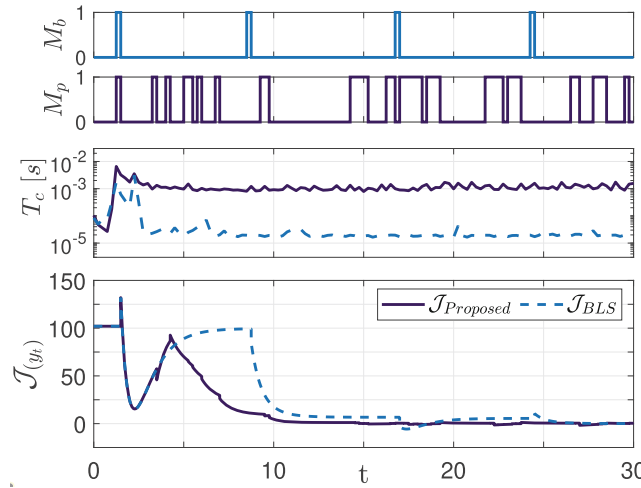


FIGURE 4 Results for the third-order problem; time t is in seconds. J shows the cost (35) optimized by the presented method, and the BLS method of Reference 3, where y_t follows the third-order dynamical plant (34). M_b and M_p show the operation modes for the BLS and proposed methods, respectively. Value 0 indicates exploration and 1 exploitation. T_c is the computation time.

visibility and communications, requiring autonomous operation. Furthermore, this application is useful in industrial setting to locate toxic gas leaks without endangering a human operator.

Typically, ESC controllers consider a model-free approach for both the plant dynamics and objective function. However, for this application, the dynamics of the drone will be well characterized. Furthermore, the drone will be generally equipped with GPS and a tracking control system allowing it to move to prescribed locations. The premise of this article is that this knowledge can be exploited to improve the performance of the ESC algorithm. In contrast, the objective function (gas concentration) is unknown. However gas concentration is a function of the drone position (plant output) rather than the motor throttles (plant input), as typically considered in the ESC literature. Thus, this problem is an ideal application of the ESC algorithm proposed in this article.

We use the standard drone dynamics,⁴⁸

$$m\ddot{y}_1 = (u_1 + u_2 + u_3 + u_4) \sin \theta_2, \quad (36a)$$

$$m\ddot{y}_2 = (u_1 + u_2 + u_3 + u_4) \sin \theta_1, \quad (36b)$$

$$J\ddot{\theta}_1 = (u_1 - u_2 - u_3 + u_4)\ell, \quad (36c)$$

$$J\ddot{\theta}_2 = (u_1 + u_2 - u_3 - u_4)\ell, \quad (36d)$$

where $m = 1$ kg is the mass, $\ell = 3$ cm is the half length, $J = 0.09$ kg m^2 is the moment of inertia for each axis, $y = (y_1, y_2)$ is the planar drone position, $\theta = (\theta_1, \theta_2)$ is the pitch and roll, and u_i $i = 1, \dots, 4$ are the propellers' forces. The yaw and vertical positions of the drone are ignored for simplicity of our presentation. We designed a linear-quadratic-integral controller with parameters $Q = I_2$ and $R = 0.01I_2$ using the linearization of the drone dynamics (36a) to satisfy Assumption 1. The cost to be minimized is the negative response of the gas sensor installed on the drone. We assume the concentration of gas is a function of location only, specifically the Gaussian plume model

$$J(y) = -\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}(y - y^*)^\top \Sigma^{-1}(y - y^*)\right) \quad (37)$$

is considered as the negative concentration of gas in the plane, where $y^* = [200, 100]^\top$ is the location of the source, and

$$\Sigma(y) = \begin{cases} \sigma^2(I - dd^\top) & \text{if } d^\top y \geq 0 \\ \sigma_0^2 I & \text{otherwise,} \end{cases}$$

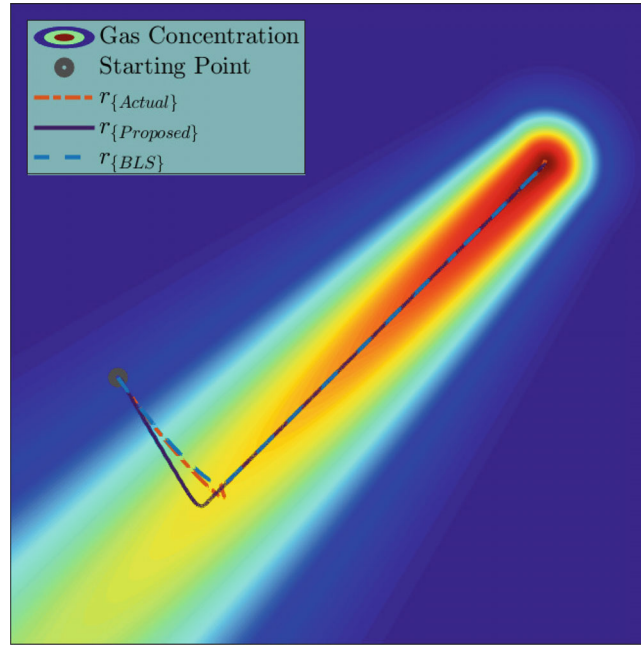


FIGURE 5 Gas is emitted from the source and blown in 45° southwest direction. Blue indicates low concentration while red indicates high concentration. The presented controller generates the connected purple path, and the controller from Reference 3 generates the dashed blue path. The red dot-dashed is generated by using the actual cost gradient for ESC.

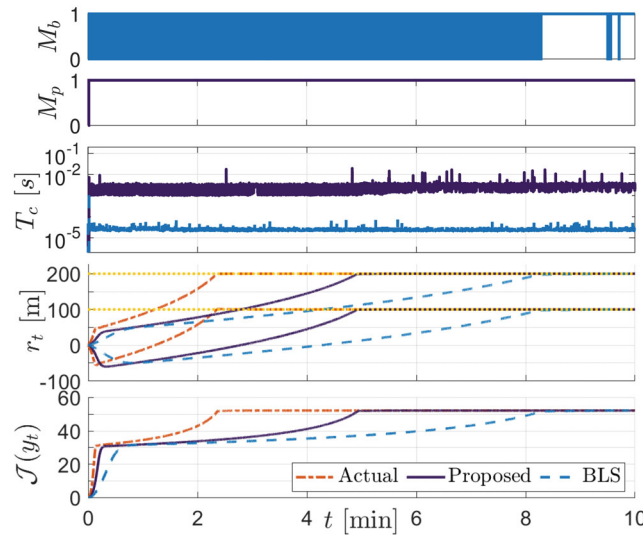


FIGURE 6 $\mathcal{J}$ shows the cost (31) optimized by the presented ESC (purple connected), the BLS ESC from Reference 3 (blue dashed), and ESC using the actual gradient (red dot-dashed). r_t is the reference command generated by the ESC controllers. The yellow dotted line indicates the location of the leak source. M_b and M_p show the operation modes for the BLS and proposed methods, respectively. Value 0 indicates exploration and 1 exploitation. T_c is the computation time.

where $d = [\cos(-\pi/4), \sin(-\pi/4)]^\top$ is the wind direction, $\sigma_0 = 15$, and $\sigma = \sigma_0 + d^\top(y - y^*)/\nu$ grows in the wind direction, where the wind speed is $\nu = 10$ meters/second. The non-convex cost (37) locally satisfies Assumption 2. We set a sample rate of 20 Hz, an estimation horizon of $T = 20$, design parameters $\epsilon = 10^{-4}$, $\alpha = 10^4$, a gain of $K = 100I$, and over-estimated bounds $\underline{H} = -3 \times 10^{-4}I$ and $\bar{H} = 3 \times 10^{-4}I$. The PE dither is $d_t = [\sin(2\pi t), \cos(2\pi t)]^\top$.

Figure 5 shows the gas concentration and the path generated by the presented method and the BLS-ESC from Reference 3 which used an ellipsoidal gradient set. Since the gradient direction is initially perpendicular to the wind direction, the drone first enters the gas plume, then turns towards the source of gas. The path is similar to the path produced by the previous algorithm. As Figure 6 shows, the reference command r_t , generated by the proposed method converges

to the location of the source $r_t \rightarrow \bar{r}^*$ by the rate of 70% faster than the one generated by the BLS-ESC from Reference 3. Moreover, the proposed controller is always in exploitation mode whereas the BLS-ESC from Reference 3 frequently switches modes.

To further validate the presented ESC algorithm, we compare its performance with gradient descent (5) that uses perfect knowledge of the gradient that is, $\hat{\theta} = \nabla J$. With perfect knowledge of the gradient, we can decrease the convergence time to 48% as shown in Figure 6. This simulation shows that about half of the convergence time was due to the ill-conditioning cost (37) that is, the Hessian of the cost varies significantly over the domain of the cost. This ill-conditioning fundamentally limits the convergence rate of gradient descent (5).⁴⁹ Unfortunately, past research has shown that optimization techniques used to address ill-conditioning do not translate well to ESC.⁵⁰ This simulation also shows that about half of the convergence time was due to the conservativeness of the gradient estimator (6) due to the uncertainty of the cost. However, if we tighten the bounds $\underline{H}, \bar{H}$ on the Hessian then the gradient estimator becomes less conservative reducing the convergence time. For instance, when we re-simulated with 3× tighter bounds $\underline{H} = -10^{-4}I$ and $\bar{H} = 10^{-4}I$, the convergence time for the proposed ESC was only 32% slower than the perfect ESC algorithm.

5 | CONCLUSIONS

We presented an ESC algorithm with a novel gradient estimator. We showed that ESC stabilizes the optimal equilibrium of the closed-loop system despite estimation errors. Furthermore, since we select the estimated gradient that optimizes the worst-case convergence of the joint Lyapunov function, our ESC provides fast and robust convergence. This was demonstrated through three benchmark examples with state-of-the-art ESC algorithms. Finally, we demonstrated the practical utility of our ESC algorithm for autonomous leak inspection. In future works, a Hessian estimator can be integrated with the presented method to estimate bounds on the curvature required for our gradient estimator. Moreover, we will investigate a stochastic set-based frameworks for the estimated gradient set when the data is noisy.

AUTHOR CONTRIBUTIONS

Ali Kashani: Conceptualization (lead); Data Curation; investigation; visualization (lead); validation (lead); resources (lead); formal analysis (lead); writing - original draft (lead); Software (lead); writing - review and editing (equal); Methodology (equal). **Shirin Panahi:** writing - review and editing (equal). **Ankush Chakrabarty:** Methodology (equal); writing - review and editing (equal). **Claus Danielson:** Funding Acquisition; Project Administration; supervision; Methodology (equal); conceptualization (supporting); resources (supporting); writing - original draft (supporting); writing - review and editing (equal); validation (supporting); visualization (supporting).

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under NSF Grant Number CMMI-2105631. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Ali Kashani  <https://orcid.org/0000-0002-7383-5792>

REFERENCES

1. LeBlanc M. *Sur l'électrification des Chemins de fer au Moyen de Courants alternatifs de Frequence Elevee*. Vol 2. Revue Générale de l'électricité; 1922:275-277.
2. Danielson C, Lacy S, Hindman B, Collier P, Hunt J, Moser R. Extremum seeking control for simultaneous beam steering and wavefront correction. *2006 American Control Conference*. IEEE; 2006:6.
3. Danielson C, Bortoff SA, Chakrabarty A. Extremum seeking control with an adaptive gain based on gradient estimation error. *IEEE Trans Syst Man Cybern Syst*. 2022;53:152-164.
4. Grushkovskaya V, Michalowsky S, Zuyev A, May M, Ebenbauer C. A family of extremum seeking laws for a unicycle model with a moving target: theoretical and experimental studies. *2018 European Control Conference (ECC)*. IEEE; 2018:1-6.

5. Cai W, Alhashim AN, N'Doye I, Laleg-Kirati TM. An extremum seeking control based approach for alignment problem of mobile optical communication systems. *IFAC Pap OnLine*. 2020;53(2):5368-5374.
6. Zhang M, Wang H, Wu J. On UAV source seeking with complex dynamic characteristics and multiple constraints: a cooperative standoff monitoring mode. *Aerosp Sci Technol*. 2022;121:107315.
7. Suttner R, Dashkovskiy S. Robustness and averaging properties of a large-amplitude, high-frequency extremum seeking control scheme. *Automatica*. 2022;136:110020.
8. Scheinker A. 100 years of extremum seeking: a survey. *Automatica*. 2024;161:111481.
9. Dong L, Li Y, Mu B, Xiao Y. Self-optimizing control of air-source heat pump with multivariable extremum seeking. *Appl Therm Eng*. 2015;84:180-195.
10. Zhang C, Ordonez R. Numerical optimization-based extremum seeking control with application to abs design. *IEEE Trans Automat Contr*. 2007;52(3):454-467.
11. Creaby J, Li Y, Seem JE. Maximizing wind turbine energy capture using multivariable extremum seeking control. *Wind Eng*. 2009;33(4):361-387.
12. Lei P, Li Y, Chen Q, Seem JE. Extremum seeking control based integration of MPPT and degradation detection for photovoltaic arrays. *Proceedings of the 2010 American Control Conference*. IEEE; 2010:3536-3541.
13. Dubbioso S, di Grazia LE, De Tommasi G, Mattei M, Mele A, Pironti A. Vertical stabilization of tokamak plasmas via extremum seeking. *IFAC J Syst Control*. 2022;21:100203.
14. Trollberg O, Jacobsen EW. Greedy extremum seeking control with applications to biochemical processes. *IFAC Pap OnLine*. 2016;49(7):109-114.
15. Scheinker A, Krstić M. Minimum-seeking for CLFs: universal semiglobally stabilizing feedback under unknown control directions. *IEEE Trans Automat Contr*. 2012;58(5):1107-1122.
16. Ghaffari A, Krstić M, Seshagiri S. Power optimization for photovoltaic microconverters using multivariable newton-based extremum seeking. *IEEE Trans Control Syst Technol*. 2014;22(6):2141-2149.
17. Poveda JI, Teel AR. A robust event-triggered approach for fast sampled-data extremization and learning. *IEEE Trans Automat Contr*. 2017;62(10):4949-4964.
18. Hunnekens B, Haring MA, van de Wouw N, Nijmeijer H. A dither-free extremum-seeking control approach using 1st-order least-squares fits for gradient estimation. *53rd IEEE Conference on Decision and Control*. IEEE; 2014:2679-2684.
19. Yuan S, Fabiani F, Baldi S. Numerical optimization-based extremum seeking control of a class of constrained nonlinear systems via finite-time state transition. *Int J Robust Nonlinear Control*. 2022;32(11):6379-6394.
20. Xu P, Roosta F, Mahoney MW. Newton-type methods for non-convex optimization under inexact hessian information. *Math Program*. 2020;184(1):35-70.
21. Guay M, Benosman M. Finite-time extremum seeking control for a class of unknown static maps. *Int J Adapt Control Signal Process*. 2021;35(7):1188-1201.
22. Haring M, Johansen TA. On the accuracy of gradient estimation in extremum-seeking control using small perturbations. *Automatica*. 2018;95:23-32.
23. Guay M, Burns DJ. A proportional integral extremum-seeking control approach for discrete-time nonlinear systems. *Int J Control*. 2017;90(8):1543-1554.
24. Ariyur KB, Krstic M. *Real-Time Optimization by Extremum-Seeking Control*. John Wiley & Sons; 2003.
25. Benosman M. *Learning-Based Adaptive Control: An Extremum Seeking Approach—Theory and Applications*. Butterworth-Heinemann; 2016.
26. Wang L, Chen S, Ma K. On stability and application of extremum seeking control without steady-state oscillation. *Automatica*. 2016;68:18-26.
27. He Z, Chen S, Sun Z, Wang L, Ma K. Twisting sliding mode extremum seeking control without steady-state oscillation. *Int J Control*. 2021;94(2):411-421.
28. Haring M, Johansen TA. Asymptotic stability of perturbation-based extremum-seeking control for nonlinear plants. *IEEE Trans Automat Contr*. 2016;62(5):2302-2317.
29. Suttner R. Extremum seeking control with an adaptive dither signal. *Automatica*. 2019;101:214-222.
30. Mele A, De Tommasi G, Pironti A. Finite-time stabilization of linear systems with unknown control direction via extremum seeking. *IEEE Trans Automat Contr*. 2021;67(10):5594-5601.
31. Guay M, Dochain D. A time-varying extremum-seeking control approach. *Automatica*. 2015;51:356-363.
32. Atta KT, Guay M. Adaptive amplitude fast proportional integral phasor extremum seeking control for a class of nonlinear system. *J Process Control*. 2019;83:147-154.
33. Chakrabarty A, Danielson C, Di Cairano S, Raghunathan A. Active learning for estimating reachable sets for systems with unknown dynamics. *IEEE Trans Cybern*. 2020;52(4):2531-2542.
34. Sun W, Tan Y. Model guided extremum seeking control of electromagnetic micromirrors. *Sci Rep*. 2021;11(1):17613.
35. Dewasme L, Vande Wouwer A. Model-free extremum seeking control of bioprocesses: a review with a worked example. *Processes*. 2020;8(10):1209.
36. Lee T-H, Narendra KS. Robust adaptive control of discrete-time systems using persistent excitation. *Automatica*. 1988;24(6):781-788.
37. Liu S, Niu B, Zong G, Zhao X, Xu N. Data-driven-based event-triggered optimal control of unknown nonlinear systems with input constraints. *Nonlinear Dyn*. 2022;109(2):891-909.

38. Kashani A, Kalhor A, Araabi BN, Danielson C. Dead-beat identification for model reference adaptive control. *2022 IEEE 61st Conference on Decision and Control (CDC)*. IEEE; 2022:2827-2832.
39. Tamizi MG, Kashani AAA, Azad FA, Kalhor A, Masouleh MT. Experimental study on a novel simultaneous control and identification of a 3-DoF delta robot using model reference adaptive control. *Eur J Control*. 2022;67:100715.
40. Liao C-K, Manzie C, Chapman A, Alpcan T. Constrained extremum seeking of a MIMO dynamic system. *Automatica*. 2019;108:108496.
41. Marchetti AG, François G, Faulwasser T, Bonvin D. Modifier adaptation for real-time optimization—methods and applications. *Processes*. 2016;4(4):55.
42. Bunin GA. On the equivalence between the modifier-adaptation and trust-region frameworks. *Comput Chem Eng*. 2014;71:154-157.
43. Apostol TM. *Mathematical Analysis*. 2nd ed. Addison-Wesley; 1974.
44. Xu Y, Sun W, Qi L. A feasible direction method for the semidefinite program with box constraints. *Appl Math Lett*. 2011;24(11):1874-1881.
45. Branicky MS. Multiple Lyapunov functions and other analysis tools for switched and hybrid systems. *IEEE Trans Automat Contr*. 1998;43(4):475-482.
46. Jiang Z-P, Wang Y. Input-to-state stability for discrete-time nonlinear systems. *Automatica*. 2001;37(6):857-869.
47. Bhattacharjee D, Subbarao K. Extremum seeking control with attenuated steady-state oscillations. *Automatica*. 2021;125:109432.
48. Bresciani T. *Modelling, Identification and Control of a Quadrotor Helicopter*. MSc Theses. Lund University; 2008.
49. Bhaya A, Kaszkurewicz E. *Control Perspectives on Numerical Algorithms and Matrix Problems*. SIAM; 2006.
50. Labar C, Garone E, Kinnaert M, Ebenbauer C. Newton-based extremum seeking: a second-order lie bracket approximation approach. *Automatica*. 2019;105:356-367.

How to cite this article: Kashani A, Panahi S, Chakrabarty A, Danielson C. Robust data-driven dynamic optimization using a set-based gradient estimator. *Optim Control Appl Meth*. 2024;1-21. doi: 10.1002/oca.3157