

Generalized Multi-Constraint Extremum Seeking

Alan Williams¹ Jorge Cortés² Alexander Scheinker³

Abstract—We generalize the Safe Extremum Seeking algorithm to address the minimization of an unknown objective function subject to multiple unknown inequality and equality constraints, relying on recent results of gradient flow systems. These constraints may represent safety or other critical conditions. The proposed ES algorithm functions as a general nonlinear programming tool, offering practical maintenance of all constraints and semiglobal practical asymptotic stability, utilizing a Lyapunov argument on the penalty function and the set-valued Lie derivative. The efficacy of the algorithm is demonstrated on a 2D problem.

I. INTRODUCTION

Optimization under constraints is a central problem in control and engineering, appearing in applications ranging from robotics to autonomous systems and real-time decision-making. Various approaches exist for solving constrained optimization problems, but the choice of method depends significantly on the information available about the objective and constraint functions. In this paper, we leverage two recently proposed methods for addressing different problems using constrained optimization: one employing control barrier functions (CBFs) and quadratic programming (QP) for general nonlinear programming, under multiple constraints with known gradients, and another using extremum seeking (ES) to optimize an objective under a single CBF constraint, when gradients of the problem are unknown. These approaches tackle different problems, but their combination presents an opportunity to develop a more general ES tool merging the ideas from both algorithms.

The first work [1] introduces a general constrained optimization framework based on the safe gradient flow. The method treats a general nonlinear programming problem (under multiple inequality and equality constraints) as a control problem, where constraints are enforced through a CBF-inspired QP solved at every time instant. The result is a continuous-time optimization algorithm that ensures that the feasible set is forward invariant, while driving trajectories towards solutions to the nonlinear program. The approach has been shown [16] to be closely related to projected gradients methods, and may be understood as a continuous, smoothed-version of projected gradient descent methods. The work

[1] assumes the gradients of the objective and constraints are known which may not apply to scenarios where only functions values are measured.

In contrast, the work [31] presents a “Safe” Extremum Seeking (Safe ES) algorithm, designed for optimization problems where the objective and CBF gradients are unknown but can be estimated through real-time measurements. ES methods introduce perturbations to the optimization variables to estimate the gradient of an unknown function in real-time. To enforce safety constraints, Safe ES also incorporates a QP-based safety filter, which prevents constraint violations while seeking the extremum of the objective function. This approach is particularly useful in settings where analytical expressions of the constraint and objective are unavailable, and it ensures “practical safety”, meaning that the constraint violation can be made arbitrarily small via more conservative design constants. However, Safe ES, in its current form, is limited to a single inequality constraint as its design is based on the classical safety filter design in [4].

Contributions: We propose a Generalized Multi-Constraint Extremum Seeking (GMC-ES) algorithm that merges ES-based gradient estimation with a QP framework to handle multiple inequality and equality constraints using only function evaluations. The method solves a QP in real time with estimated gradients, and its aggressiveness in enforcing constraints is tunable through a design parameter. Using a penalty-function as a Lyapunov function, we establish semiglobal practical asymptotic stability for a class of nonconvex problems. More so, the practical maintenance of the feasible set is also achieved. This extends Safe ES to Generalized Multi-Constraint ES (GMC-ES), going beyond single-constraint settings, providing a real-time, constraint-enforcing optimizer for nonlinear programs. Along the way, we also provide conditions for global convergence of the safe gradient flow by means of the set-valued Lie derivative.

Literature (ES, CBFs, safe gradient flow): ES is a powerful control method for dynamic systems with uncertainties, with its first formal stability proof established in the seminal work of [22] paper, sparking renewed interest in this method. Constrained ES specifically has been explored in cases where constraints are unknown, often assuming convexity, incorporating techniques such as switching ES algorithms using projection maps [9] and convex problem-solving via saddle points of modified barrier functions [23]. Studies in [17], [18] provide Lie bracket-based analysis of saddle-point dynamics, while [5] considers projection-based gradient methods. We recommend [28] for a comprehensive survey on extremum seeking.

The literature on safe control has grown significantly,

This work was supported by the U.S. Department of Energy (DOE), Office of Science, Office of High Energy Physics contract number 89233218CNA000001 and the Los Alamos National Laboratory LDRD Program Directed Research (DR) project 20220074DR

¹Alan Williams and ³Alexander Scheinker are with the Accelerator Operations and Technology - Instrumentation and Control (AOT-IC) group at Los Alamos National Lab, Los Alamos, NM 87545, USA (e-mail: {awilliams, ascheink}@lanl.gov).

²Jorge Cortés is with the Department of Mechanical and Aerospace Engineering, University of California, San Diego, CA 92093-0411, USA (e-mail: cortes@ucsd.edu).

driven by foundational work in [2], [30], particularly in the development of Control Barrier Functions (CBFs) and safety filters to enforce constraint satisfaction in real-time control applications [3].

The safe gradient flow algorithm [1] has found applications in power systems [13], reinforcement learning [19], [24], and more such as for feedback control design for dynamical systems [10]. Here, we extend the range of applications to extremum seeking control. The connections of the safe gradient flow with other dynamical system-inspired approaches to solve constrained optimization problems, in particular those based on projected dynamical systems [16], [26], are discussed in [1].

II. PRELIMINARIES ON NONSMOOTH ANALYSIS

Suppose $V : \mathbb{R}^n \rightarrow \mathbb{R}$ is locally Lipschitz, and Ω_V is a set of measure zero containing the points where V is not differentiable. Denoting the convex hull of a set $\mathcal{S}$ as $\text{co}(\mathcal{S})$, the *generalized gradient* is defined as

$$\partial V(x) := \text{co} \left\{ \lim_{i \rightarrow \infty} \nabla V(x_i) : x_i \rightarrow x, x_i \notin \Omega_V \right\}.$$

Intuitively, the generalized gradient $\partial V(x)$, at some point x , describes a convex combination of all nearby gradients where $V(x)$ is strictly differentiable. See [11] and [14] for more information on generalized gradients.

Regularity of a function proves useful in calculating generalized gradients. The definition of regularity can be seen in [11], but the following class of regular functions is sufficient background for this paper, rewritten from [11, Proposition 2.3.12], which provides a useful expression for the generalized gradient.

Proposition 1: Let $g_k : \mathbb{R}^n \rightarrow \mathbb{R}$ be differentiable functions for $k = 1, 2, \dots, m$, and let $V : \mathbb{R}^n \rightarrow \mathbb{R}$ be defined as the maximum of the collection of functions

$$V(x) = \max\{g_k(x) : k = 1, 2, \dots, m\}. \quad (1)$$

Let $I_V(x)$ denote the set of indices k for which $V(x) = g_k(x)$ — the “active” functions. Then:

- 1) V is regular and locally Lipschitz.
- 2) The generalized gradient can be expressed as

$$\partial V(x) = \text{co} \{ \nabla g_i(x) : i \in I_V(x) \}.$$

See [11, Chapter 2] for more information on generalized gradients and useful calculus rules. Lipschitz functions we encounter in this work can often be written in the “pointwise maximum” form (1) — note these two basic identities $|g(x)| = \max\{g(x), -g(x)\}$ and $g(x) + [h(x)]_+ = \max\{g(x), g(x) + h(x)\}$ where, we denote $[h]_+ = \max\{0, h\}$ for $h \in \mathbb{R}$.

We now define a set-valued form of the Lie derivative, utilizing the generalized gradient, which is generally defined for a set-valued map but we give a simpler definition for a Lipschitz function, sufficient for this work [14]. Consider the system

$$\dot{x} = f(x). \quad (2)$$

Definition 1: Let $V : \mathbb{R}^n \rightarrow \mathbb{R}$ and $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be locally Lipschitz. The *set-valued Lie derivative* of V with respect to f is

$$\mathcal{L}_f V(x) = \{a \in \mathbb{R} : f(x)^\top p = a \text{ for all } p \in \partial V(x)\}.$$

If V is differentiable, then one recovers the standard Lie derivative $\mathcal{L}_f V(x) = \{f(x)^\top \nabla V(x)\}$. If V is differentiable everywhere then with some abuse of notation we may ignore the set notation, writing $\mathcal{L}_f V(x) = f(x)^\top \nabla V(x)$.

The following result [6], [14] is a generalization of the invariance principle, and it is the primary tool used to study stability in this work. In the case where the set-valued quantity $\mathcal{L}_f V(x)$ is the empty set, we use the convention that $\max\{\emptyset\} = -\infty$.

Theorem 1 (Invariance Principle): Let $V : \mathbb{R}^n \rightarrow \mathbb{R}$ be locally Lipschitz and regular. Let $\mathcal{S} \subset \mathbb{R}^n$ be compact and positively invariant for (2). Suppose $\max \mathcal{L}_f V(x) \leq 0$ for all $x \in \mathcal{S}$. Then all solutions of (2), with $x(t_0) \in \mathcal{S}$, converge to the largest invariant set $\mathcal{M}$ contained in

$$\mathcal{S} \cap \overline{\{x \in \mathbb{R}^n : 0 \in \mathcal{L}_f V(x)\}}.$$

The next result is Nagumo’s theorem, giving a useful condition which may be used to show that a set is invariant — note that Theorem 1 *requires* an invariant, compact set $\mathcal{S}$. The original work is given in [25], but more simply reformulated from [8, Theorem 4.7].

Theorem 2 (Nagumo’s Theorem): Suppose $\mathcal{S} \subseteq \mathbb{R}^n$ is a closed set. Then $\mathcal{S}$ is invariant with respect to (2) with f locally Lipschitz if and only if

$$f(x) \in \mathcal{T}_{\mathcal{S}}(x) \text{ for all } x \in \text{bd}(\mathcal{S}),$$

where $\mathcal{T}_{\mathcal{S}}(x)$ is the Bouligand tangent cone [8, Definition 4.6].

The following result characterizes Bouligand’s tangent cone for sublevel sets defined by the maximum of differentiable functions, showing that for a sublevel set $\mathcal{S} = \{V(x) \leq \bar{V}\}$, with V given in (1), the set $\mathcal{T}_{\mathcal{S}}(x)$ can be equivalently expressed using the set-valued Lie derivative $\mathcal{L}_f V(x)$

Proposition 2: Let $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be locally Lipschitz and $V : \mathbb{R}^n \rightarrow \mathbb{R}$ regular (which includes functions of the form as in (1)). Consider $\mathcal{S} = \{x \in \mathbb{R}^n : V(x) \leq \bar{V}\}$ for $\bar{V} \in \mathbb{R}$. Then for $x \in \text{bd}(\mathcal{S})$ such that $0 \notin \partial V(x)$,

$$\max \mathcal{L}_f V(x) \leq 0 \iff f(x) \in \mathcal{T}_{\mathcal{S}}(x).$$

Proof: First, note that “Bouligand’s tangent cone” in the terminology of [11] is referred to as the “contingent cone” — see [12, Chapter 7]. Since V is regular by Proposition 1 and $0 \notin \partial V(x)$ for $x \in \text{bd}(\mathcal{S})$, then by [11, Theorem 2.4.7] $\mathcal{S}$ is regular and the “tangent cone” (in Clarke’s terminology) is equivalent to Bouligand’s tangent cone (or the contingent cone in Clarke’s terminology) with $\mathcal{T}_{\mathcal{S}}(x) = \{v \in \mathbb{R}^n : V^\circ(x; v) \leq 0\}$, where $V^\circ(x; v)$ is the “generalized directional derivative” of V at the point x in the direction v . By

[11, Proposition 2.1.2], $V^\circ(x; v) = \max\{\xi^\top v : \xi \in \partial V(x)\}$, and so

$$\mathcal{T}_S(x) = \{v \in \mathbb{R}^n : \max\{\xi^\top v : \xi \in \partial V(x)\} \leq 0\}.$$

Forward implication: take some $x \in \text{bd}(S)$ and suppose $\max \mathcal{L}_f V(x) \leq 0$. Then for all $a \in \mathcal{L}_f V(x)$, where $a = f(x)^\top p$ for all $p \in \partial V(x)$, we have $a = f(x)^\top p \leq 0$, which implies $f(x) \in \mathcal{T}_S$.

Backward implication: take some $x \in \text{bd}(S)$ and suppose $f(x) \in \mathcal{T}_S$, implying $f(x)^\top p \leq 0$ for all $p \in \partial V(x)$. For any $a \in \mathcal{L}_f V(x)$ we have $a = f(x)^\top p$, for all $p \in \partial V(x)$. Since $a = f(x)^\top p \leq 0$ then $\max \mathcal{L}_f V(x) \leq 0$. ■

Theorem 1 directly implies that if the system $\dot{x} = f(x)$ has an equilibrium x_e on some compact invariant set $\mathcal{S}$, and if $\max \mathcal{L}_f V(x) < 0$ for $x \in \mathcal{S} \setminus \{x_e\}$, then solutions converge to x_e . But in the context of this work, we ultimately are interested in the convergence to a *neighborhood* around a point.

Consider instead the system

$$\dot{x} = \tilde{F}(x) = F(x) + O(a), \quad (3)$$

where $F(x_e) = 0$ and $O(a)$ is Lipschitz in x but can be made arbitrarily small by choosing the parameter $a > 0$ sufficiently small (on a compact set). In this scenario, we would like to show that for any $\Delta > 0$, trajectories of (3) from the initial condition $x(0) \in \bar{B}_\Delta(x_e)$ converge to $\bar{B}_\nu(x_e)$, for any $\nu > 0$, no matter how small, for some $0 < a < a^*$. This is known as “SPA stability” or semiglobal practical asymptotic stability, see [27, Assumption 2] for the definition of this property. Theorem 1 is the tool we use to demonstrate this SPA property in our work, and Theorem 2 is the tool used to find an invariant set $\mathcal{S}$, a requirement in Theorem 1.

Organization: In Section III we study the safe gradient flow algorithm and derive conditions for which global convergence is guaranteed. Rather than imposing conditions on the nonlinear program which is being solved, as studied in [1], we extend the work and derive conditions on the QP defining the dynamics, and characterize results in terms of the set-valued Lie derivative. Then in Section IV, we give an extremum seeking design, GMC-ES, which mimics the safe gradient flow dynamics, extending the work in [31] to handle multiple constraints. We also give a high level overview of the analysis and present a theorem showing practical convergence and practical anytime maintenance of all constraints. Finally in Section V we implement GMC-ES in a simple 2D example.

III. STABILITY OF THE SAFE GRADIENT FLOW

A. Lipschitzness, Equilibrium, and Invariance

In this section we state the form of the safe gradient flow from [1], the basis for the extremum seeking design of this work, and characterize some key results in terms of the set-valued Lie derivative. The results in this section differ slightly in nature to the safe gradient properties described in [1]. Here, we impose a *global constraint qualification on the QP* (7) defining the safe gradient flow, whereas the work

of [1] imposes *constraint qualifications on the nonlinear program* (4). We take the most direct route of imposing global conditions on the QP (7), since our goal is to study global convergence of (6) under the dynamics (7). Therefore, many of the results in this section, *which hold globally*, have analogs in the work of [1], *which hold locally* or in a region near the feasible set $\mathcal{C}$. For example, Proposition 5 has an analog [1, Proposition 5.1] which holds given a constraint qualification on nonlinear program at an optimal point θ^* . Proposition 6 also has an analog [1, Theorem 5.4] given a constraint qualification on the nonlinear program for $\theta \in \mathcal{C}$.

Consider the nonlinear program:

$$\begin{aligned} & \underset{\theta \in \mathbb{R}^n}{\text{minimize}} && f(\theta) \\ & \text{subject to} && g(\theta) \leq 0 \\ & && h(\theta) = 0, \end{aligned} \quad (4)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^l$ are differentiable with locally Lipschitz derivatives. Let

$$\mathcal{C} = \{\theta \in \mathbb{R}^n \mid g(\theta) \leq 0, h(\theta) = 0\}, \quad (5)$$

denote the feasible set.

Consider the safe gradient flow dynamics:

$$\dot{\theta} = \mathcal{G}_\alpha(\theta) \quad (6)$$

where $\mathcal{G}_\alpha$ satisfies the quadratic program with $\alpha > 0$:

$$\begin{aligned} \mathcal{G}_\alpha(\theta) = \arg \min_{\xi \in \mathbb{R}^n} & \quad \left\{ \frac{1}{2} \|\xi + \nabla f(\theta)\|^2 \right\} \\ \text{subject to} & \quad \frac{\partial g(\theta)}{\partial \theta} \xi \leq -\alpha g(\theta), \\ & \quad \frac{\partial h(\theta)}{\partial \theta} \xi = -\alpha h(\theta). \end{aligned} \quad (7)$$

The dynamics $\mathcal{G}_\alpha$ can be thought of as a gradient descent based method, and under no constraints the minimization under ξ results in $\xi^* = -\nabla f(\theta)$ — standard gradient descent.

The Karush–Kuhn–Tucker (KKT) conditions for the optimal point ξ^* along with associated multipliers $(u, v) \in \mathbb{R}^m \times \mathbb{R}^l$ are:

$$\xi^* + \nabla f(\theta) + \frac{\partial g(\theta)}{\partial \theta}^\top u + \frac{\partial h(\theta)}{\partial \theta}^\top v = 0 \quad (8a)$$

$$\frac{\partial g(\theta)}{\partial \theta} \xi^* + \alpha g(\theta) \leq 0 \quad (8b)$$

$$\frac{\partial h(\theta)}{\partial \theta} \xi^* + \alpha h(\theta) = 0 \quad (8c)$$

$$u \geq 0 \quad (8d)$$

$$u^\top \left(\frac{\partial g(\theta)}{\partial \theta} \xi^* + \alpha g(\theta) \right) = 0 \quad (8e)$$

where $\xi^* = \mathcal{G}_\alpha(\theta)$. Let the Lagrange multiplier set be defined as

$$\Lambda(\theta) = \{(u, v) \in \mathbb{R}_{\geq 0}^m \times \mathbb{R}^l \mid \exists \xi^* \in \mathbb{R}^n \text{ such that } (\xi^*, u, v) \text{ solves (8)}\}. \quad (9)$$

We impose the Mangasarian–Fromovitz constraint qualification (MFCQ) for all $\theta \in \mathbb{R}^n$ on the QP in (7).

Assumption 1 (MFCQ holds for the QP): MFCQ holds for the QP (7) at each $\theta \in \mathbb{R}^n$. Namely, for each $\theta \in \mathbb{R}^n$:

- 1) $\frac{\partial h(\theta)}{\partial \theta}$ has full row rank, and
- 2) $\exists d \in \mathbb{R}^n$ such that $\nabla h_j(\theta)^\top d = 0$ for all $j = 1, \dots, l$ and $\nabla g_i(\theta)^\top d < 0$ for all $i \in \mathcal{A}_{\text{QP}}(\theta)$, where

$$\mathcal{A}_{\text{QP}}(\theta) := \{i \in \{1, \dots, m\} \mid \nabla g_i(\theta)^\top \xi^* + \alpha g_i(\theta) = 0\}.$$

Proposition 3 (MFCQ implies bounded multipliers): Suppose Assumption 1 holds.

- i) Then the Lagrange multiplier set $\Lambda(\theta)$ associated with (8) is nonempty and bounded. In particular, for all $\theta \in \mathbb{R}^n$ there exists an $M(\theta) < \infty$ such that every $(u, v) \in \Lambda(\theta)$ satisfies $u_i \leq M(\theta)$ and $|v_j| \leq M(\theta)$ for all $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, l\}$.
- ii) Slater's condition holds for (7) at the same θ , which means $\exists \xi' \in \mathbb{R}^n$ satisfying $\nabla g_i(\theta)^\top \xi' + \alpha g_i(\theta) < 0$ and $\nabla h_j(\theta)^\top \xi' + \alpha h_j(\theta) = 0$ for all $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, l\}$.

Proof: Statement i) the set $\Lambda(\theta)$ is bounded by [20]. Statement ii) see [7, Proposition 3.3.8 proof]. ■

The following proposition gives an additional fact and $\mathcal{G}_\alpha(\theta)$ is locally Lipschitz.

Proposition 4 (Lipschitzness of safe gradient flow): Under Assumption 1, for all $\theta \in \mathbb{R}^n$, $\mathcal{G}_\alpha(\theta)$ is well-defined and locally Lipschitz.

Proof: Under Assumption 1, MFCQ provides a direction d with $\nabla h_j^\top d = 0$ and $\nabla g_i^\top d < 0$ for all active i , so no inequality can be an equality for all feasible points; hence in [15, (3.5)-(3.6)], B_0 is vacuous, the strict-slack condition for A_0 follows from Slater's condition (implied by MFCQ), and the constant-rank hypothesis [15, (3.8)] reduces to rank $D(\theta) = l$ uniformly by the full row rank of $\partial h / \partial \theta$. Since the the problem data in (7) are locally Lipschitz in θ and the quadratic cost is strongly convex, [15, Theorem 3.10] applies, yielding $\mathcal{G}_\alpha(\theta)$ locally Lipschitz in θ . ■

Under a constraint qualification, a local solution $\theta^* \in \mathcal{C}$ to (4) satisfies the KKT conditions

$$\nabla f(\theta^*) + \frac{\partial g(\theta^*)}{\partial \theta}^\top u^* + \frac{\partial h(\theta^*)}{\partial \theta}^\top v^* = 0, \quad (10a)$$

$$g(\theta^*) \leq 0, \quad (10b)$$

$$h(\theta^*) = 0, \quad (10c)$$

$$u^* \geq 0, \quad (10d)$$

$$(u^*)^\top g(\theta^*) = 0, \quad (10e)$$

with some Lagrange multipliers $(u^*, v^*) \in \mathbb{R}^m \times \mathbb{R}^l$ associated with $\theta^* \in \mathbb{R}^n$ [7]. We make the assumption that the KKT conditions uniquely determines the global solution to the nonlinear problem (4), in anticipation of a global study of convergence.

Assumption 2: There exists a unique solution $\theta^* \in \mathbb{R}^n$ to (4), and is the only point that satisfies (10) with some Lagrange multipliers $(u^*, v^*) \in \mathbb{R}^m \times \mathbb{R}^l$.

If θ^* is a local solution to (4), and a constraint qualification holds at θ^* , then the KKT conditions are satisfied at θ^* . However, the converse is not true: a point θ may satisfy the KKT conditions, without being a (local or global) solution of (4). Consequently, to ensure correctness in the study of the global convergence of (6), we impose Assumption 2. The next result also follows from the work of [1, Proposition 5.1], demonstrating a single equilibrium of (6).

Proposition 5 (Equilibrium at KKT point): Under Assumptions 1 and 2, $\mathcal{G}_\alpha(\theta) = 0$ if and only if $\theta = \theta^*$.

Proof: If $\mathcal{G}_\alpha(\theta) = \xi^* = 0$, then the conditions in (8) with $\alpha > 0$ can be simplified to exactly that of (10), which implies $\theta = \theta^*$ uniquely by Assumption 2. If $\theta = \theta^*$, notice that $\xi^* = 0$ satisfies (8) along with $u = u^*$ and $v = v^*$. Since the QP in (7) is strongly convex and Slater's condition holds, $\xi^* = 0$ is unique. ■

We define

$$\mathcal{C}_{\bar{g}, \bar{h}} = \{\theta \in \mathbb{R}^n \mid g(\theta) \leq \bar{g}, |h(\theta)| \leq \bar{h}\}, \quad (11)$$

where $\bar{g} \in \mathbb{R}_{\geq 0}^m$, $\bar{h} \in \mathbb{R}_{\geq 0}^l$. Note that $\mathcal{C} \subseteq \mathcal{C}_{\bar{g}, \bar{h}}$.

Proposition 6 (Time derivatives and Invariance): Under Assumptions 1 and 2 the following hold:

- i) All sets of the form $\mathcal{C}_{\bar{g}, \bar{h}}$, including the feasible set $\mathcal{C}$, are invariant with respect to (6).
- ii) For any $(u, v) \in \Lambda(\theta)$ satisfying (8),

$$\mathcal{L}_{\mathcal{G}_\alpha} f(\theta) = -\|\mathcal{G}_\alpha(\theta)\|^2 + \alpha u^\top g(\theta) + \alpha v^\top h(\theta)$$

- iii) For any $\theta \in \mathcal{C}$,

$$\mathcal{L}_{\mathcal{G}_\alpha} f(\theta) \leq 0$$

and $\mathcal{L}_{\mathcal{G}_\alpha} f(\theta) = 0$ if and only if $\theta = \theta^*$.

Proof: Statement i): $\xi^* = \mathcal{G}_\alpha(\theta)$ are Lipschitz in θ and satisfy (8b) implying $\frac{dg_i(\theta)}{dt} = \nabla g_i(\theta)^\top \mathcal{G}_\alpha(\theta) \leq -\alpha g_i(\theta)$ for all $i \in \{1, \dots, m\}$, according to the dynamics (6). Similarly, (8c) is satisfied, implying and $\frac{dh_j(\theta)}{dt} = -\alpha h_j(\theta)$ for $j \in \{1, \dots, l\}$. Therefore the sets $S_{g_i} = \{\theta : g_i(\theta) \leq \bar{g}_i\}$ and $S_{h_j} = \{\theta : |h_j(\theta)| \leq \bar{h}_j\}$ are invariant [21, Lemma 3.4], and since $\mathcal{C}_{\bar{g}, \bar{h}}$ is the intersection of all S_{g_i} and S_{h_j} , then $\mathcal{C}_{\bar{g}, \bar{h}}$ is invariant. See [1, Lemma 5.8] for proofs of statements ii) and iii). ■

Proposition 6i allows one to study trajectories on the invariant set $\mathcal{C}_{\bar{g}, \bar{h}}$, rather than all of $\mathbb{R}^n$. If $\mathcal{C}_{\bar{g}, \bar{h}}$ extends unbounded in some direction (it is not compact), then additional assumptions will need to be imposed for global stability, Theorem 1 requires a compact invariant set. Regardless, the invariance of $\mathcal{C}_{\bar{g}, \bar{h}}$ is useful in that it demonstrates that all $h_j(\theta(t))$ remain bounded above and below in time, and all $g_i(\theta(t))$ remain bounded above in time.

B. The Set-Valued Lie Derivative and Stability

Consider the regular function $V_\epsilon : \mathbb{R}^n \rightarrow \mathbb{R}$ with parameter $\epsilon > 0$,

$$V_\epsilon(\theta) = f(\theta) + \frac{1}{\epsilon} \sum_{i=1}^m [g_i(\theta)]_+ + \frac{1}{\epsilon} \sum_{j=1}^l |h_j(\theta)|. \quad (12)$$

Then, using Proposition 1, and since $g(\theta)$ and $h(\theta)$ are differentiable, the generalized gradient is

$$\partial V_\epsilon(\theta) = \left\{ \nabla f(\theta) + \frac{1}{\epsilon} \sum_{i=1}^m q_i \nabla g_i(\theta) + \frac{1}{\epsilon} \sum_{j=1}^l r_j \nabla h_j(\theta) : \right. \\ \left. q \in Q(\theta), r \in R(\theta) \right\} \quad (13)$$

where q (associated with g) and r (associated with h) are

$$Q(\theta) = \left\{ q \in \mathbb{R}^m : q_i \in \begin{cases} \{0\} & \text{if } g_i(\theta) < 0 \\ \{1\} & \text{if } g_i(\theta) > 0 \\ [0, c_1] & \text{if } g_i(\theta) = 0 \end{cases} \right\}, \quad (14)$$

$$R(\theta) = \left\{ r \in \mathbb{R}^l : r_j \in \begin{cases} \{\text{sign}(h_j(\theta))\} & \text{if } h_j(\theta) \neq 0 \\ [-c_2, c_2] & \text{if } h_j(\theta) = 0 \end{cases} \right\}, \quad (15)$$

for some $c_1, c_2 > 0$, which are determined by the number of “active” functions (for $g_i = 0$ and $h_j = 0$) in order to satisfy the convex combination.

Therefore, the set-valued Lie derivative $\mathcal{L}_{\mathcal{G}_\alpha} V_\epsilon(\theta)$ is

$$\mathcal{L}_{\mathcal{G}_\alpha} V_\epsilon(\theta) = \left\{ \nabla f(\theta)^\top \mathcal{G}_\alpha(\theta) + \frac{1}{\epsilon} \sum_{i=1}^m q_i \nabla g_i(\theta)^\top \mathcal{G}_\alpha(\theta) \right. \\ \left. + \frac{1}{\epsilon} \sum_{j=1}^l r_j \nabla h_j(\theta)^\top \mathcal{G}_\alpha(\theta) : q \in Q(\theta), r \in R(\theta) \right\}. \quad (16)$$

We take some element $p \in \mathcal{L}_{\mathcal{G}_\alpha} V_\epsilon(\theta)$, and using Proposition 6, we have $\mathcal{L}_{\mathcal{G}_\alpha} f(\theta) = \nabla f(\theta)^\top \mathcal{G}_\alpha(\theta)$ and

$$p = -\|\mathcal{G}_\alpha(\theta)\|^2 + \alpha u^\top g(\theta) + \alpha v^\top h(\theta) \\ + \frac{1}{\epsilon} \sum_{i=1}^m q_i \nabla g_i(\theta)^\top \mathcal{G}_\alpha(\theta) + \frac{1}{\epsilon} \sum_{j=1}^l r_j \nabla h_j(\theta)^\top \mathcal{G}_\alpha(\theta). \quad (17)$$

From the rows of (8c), we have

$$r_j \nabla h_j(\theta)^\top \mathcal{G}_\alpha(\theta) = \alpha r_j h_j(\theta) = -\alpha |h_j(\theta)|. \quad (18)$$

and from the rows of (8b),

$$q_i \nabla g_i(\theta)^\top \mathcal{G}_\alpha(\theta) \leq -\alpha q_i g_i(\theta) = -\alpha [g_i(\theta)]_+. \quad (19)$$

Therefore

$$p \leq -\|\mathcal{G}_\alpha(\theta)\|^2 + \alpha \sum_{i=1}^m \left(u_i g_i(\theta) - \frac{[g_i(\theta)]_+}{\epsilon} \right) \\ + \alpha \sum_{j=1}^l \left(v_j h_j(\theta) - \frac{|h_j(\theta)|}{\epsilon} \right). \quad (20)$$

Because $u_i g_i(\theta) \leq u_i [g_i(\theta)]_+$ and $v_j h_j(\theta) \leq |v_j| |h_j(\theta)|$ then

$$p \leq -\|\mathcal{G}_\alpha(\theta)\|^2 + \alpha \sum_{i=1}^m [g_i(\theta)]_+ \left(u_i - \frac{1}{\epsilon} \right) \\ + \alpha \sum_{j=1}^l |h_j(\theta)| \left(|v_j| - \frac{1}{\epsilon} \right). \quad (21)$$

From this form of the bound on $p \in \mathcal{L}_{\mathcal{G}_\alpha} V_\epsilon(\theta)$ we can see that if ϵ is chosen small, we can dominate the multiplier terms u_i and $|v_j|$ in order to force $\max \mathcal{L}_{\mathcal{G}_\alpha} V_\epsilon(\theta) < 0$. The constraint qualification, Assumption 1, guarantees that u, v are bounded.

A simple condition which guarantees global asymptotic stability is that one of the inequality constraints $g_i(\theta)$ has compact levels – or with many inequality constraints, the combination of sublevels yields a compact set. For example, consider $\theta \in \mathbb{R}^2$ with no equality constraints (so $\mathcal{C}_{\bar{g}, \bar{h}}$ reduces to $\mathcal{C}_{\bar{g}}$) and two inequality constraints. Suppose $\mathcal{C}_{\bar{g}} = \{\theta \in \mathbb{R}^2 \mid g_1(\theta) \leq \bar{g}, g_2(\theta) \leq \bar{g}\}$ for $\bar{g} > 0$ where $g_1(\theta) = -\theta_2$ and $g_2(\theta) = -1 + \theta_1^2 + \theta_2$. Therefore $g_i(\theta(t)) \leq g_i(\theta(0))e^{-\alpha t}$ for $i = 1, 2$ and $\mathcal{C}_{\bar{g}}$ is compact and invariant for all $\bar{g} \geq 0$.

Assumption 3: For any θ there exists a compact $\mathcal{C}_{\bar{g}, \bar{h}}$ such that $\theta \in \mathcal{C}_{\bar{g}, \bar{h}}$.

Assumption 3 implies any initial condition can be contained in a compact invariant set $\theta \in \mathcal{C}_{\bar{g}, \bar{h}}$, with invariance being guaranteed by Proposition 6. Therefore Theorem 1 can directly be applied, with the Lyapunov function V_ϵ where $\epsilon > 0$ is chosen sufficiently small to dominate all u, v , which are bounded on $\mathcal{C}_{\bar{g}, \bar{h}}$ by Proposition 3.

Theorem 3: Let Assumptions 1, 3, and 2 hold. Then the dynamics (6) are globally asymptotically stable with respect to the equilibrium θ^* .

Proof: Any initial condition $\theta(t_0)$ lies in a particular set $\mathcal{C}_{\bar{g}, \bar{h}}$ for some $\bar{g}, \bar{h} > 0$, and $\mathcal{C}_{\bar{g}, \bar{h}}$ is compact (Assumption 3) and invariant (Proposition 6). Consider (21) for $p \in \mathcal{L}_{\mathcal{G}_\alpha} V_\epsilon(\theta)$. Choose $\epsilon > 0$ such that $\epsilon < 1/\bar{M}$ where $\bar{M} = \max_{\theta \in \mathcal{C}_{\bar{g}, \bar{h}}} M(\theta)$ for $M(\theta)$ given in Proposition 3. Therefore $p < 0$ for all $\theta \in \mathcal{C}_{\bar{g}, \bar{h}} \setminus \{\theta^*\}$ and by Theorem 1 we have the result. ■

Comments on when $\mathcal{C}_{\bar{g}, \bar{h}}$ is unbounded: Proving global stability on a potentially unbounded feasible set $\mathcal{C}$ (which means all sets $\mathcal{C}_{\bar{g}, \bar{h}}$ are unbounded) is more challenging. One will need to find an arbitrarily large compact invariant $\mathcal{M}$ (so that it subsumes any initial condition) without the use of $\mathcal{C}_{\bar{g}, \bar{h}}$. Instead, one could use of sublevel sets of the function V_ϵ (with radial unboundedness), since compactness of $\mathcal{C}_{\bar{g}, \bar{h}}$ cannot be relied upon. One challenge which must be dealt with is, for a fixed $\bar{V} > 0$, making ϵ small (to dominate u, v in (21)) makes the sublevel $\mathcal{M} = \{\theta : V_\epsilon(\theta) \leq \bar{V}\}$ small, due to the form in (12). In other words, the choice of a small ϵ is at odds with the goal of finding a large invariant set $\mathcal{M}$, as the compact invariant set is *dependent* on ϵ .

IV. EXTREMUM SEEKING

A. Design

In the absence of knowledge of ∇f , ∇g_i , and ∇h_j we can consider an approximation of the design in (7), through the use of techniques used in extremum seeking to estimate the

gradients. The GMC-ES dynamics are given as

$$\dot{\hat{\theta}} = k\omega_f \hat{\mathcal{G}}_\alpha \quad (22)$$

$$\dot{G}_f = -\omega_f(G_f - (f(\theta) - \eta_f)M(t)) \quad (23)$$

$$\dot{G}_{g_i} = -\omega_f(G_{g_i} - (g_i(\theta) - \eta_{g_i})M(t)) \quad (24)$$

$$\dot{G}_{h_j} = -\omega_f(G_{h_j} - (h_j(\theta) - \eta_{h_j})M(t)) \quad (25)$$

$$\dot{\eta}_f = -\omega_f(\eta_f - f(\theta)) \quad (26)$$

$$\dot{\eta}_{g_i} = -\omega_f(\eta_{g_i} - g_i(\theta)) \quad (27)$$

$$\dot{\eta}_{h_j} = -\omega_f(\eta_{h_j} - h_j(\theta)) \quad (28)$$

for $i = 1, \dots, m$ and $j = 1, \dots, l$, where $\hat{\mathcal{G}}_\alpha$ represents the approximate solution to (7) and is defined as the solution to the quadratic program with $\alpha > 0$:

$$\begin{aligned} \hat{\mathcal{G}}_\alpha = \arg \min_{\xi \in \mathbb{R}^n} \quad & \left\{ \frac{1}{2} \|\xi + G_f\|^2 \right\} \\ \text{subject to} \quad & J_g \xi \leq -\alpha \eta_g, \\ & J_h \xi = -\alpha \eta_h. \end{aligned} \quad (29)$$

We define the relation

$$\hat{\theta} := \hat{\theta} + S(t) \quad (30)$$

for $i = 1, \dots, m$ and $j = 1, \dots, l$ where $G_f, G_{g_i}, G_{h_j} \in \mathbb{R}^n$ are estimates of $\nabla f(\theta), \nabla g_i(\theta), \nabla h_j(\theta)$, and $\eta_f, \eta_{g_i}, \eta_{h_j} \in \mathbb{R}$ are filtered measurements of $f(\theta), g_i(\theta), h_j(\theta)$ respectively. The Jacobian estimates of g and h are grouped as

$$J_g = \begin{bmatrix} G_{g_1}^\top \\ \vdots \\ G_{g_m}^\top \end{bmatrix}, \quad J_h = \begin{bmatrix} G_{h_1}^\top \\ \vdots \\ G_{h_l}^\top \end{bmatrix}. \quad (31)$$

for a design parameters $k, \omega_f, \alpha, \rho_\alpha, \epsilon_h > 0$. Note that $\eta_g = [\eta_{g_1}, \dots, \eta_{g_m}]^\top$ and $\eta_h = [\eta_{h_1}, \dots, \eta_{h_l}]^\top$. As in classical extremum seeking, the perturbation signal S and demodulation signal M are given by $S_i(t) = a \sin(\omega_i t)$ and $M_i(t) = \frac{2}{a} \sin(\omega_i t)$ and contain additional design parameters $\omega_i, a \in \mathbb{R}_{>0}$. The parameter dynamics are a function of filtered measurements and gradients (shown in (29) without arguments for space) of the nonlinear programming problem: $\hat{\mathcal{G}}_\alpha = \hat{\mathcal{G}}_\alpha(G_f, G_{g_1}, \dots, G_{h_l}, \eta_f, \dots, \eta_{g_1}, \dots, \eta_{h_l}, \dots)$.

Naming Conventions: following [1], we use α for the constraint enforcement parameter, while [31] used c . In [31], h was used for inequality constraints, whereas here, h are the equality constraints, and g are the inequalities (where $g \leq 0$ defines the feasible set, and not $g \geq 0$). In [31], α was used as a parameter in the Lyapunov analysis, but here the equivalent parameter is $\frac{1}{\epsilon}$, using conventions from [1].

B. Analysis

In this section we give a high level preview of the analysis used to claim SPA stability of the system (22)-(28).

We first perform the series of transformations [27, Section 3] and summarize the details here:

- 1) Transform (22)-(28) to a new time variable $\tau = \omega_f t$.
- 2) In the time scale τ perform averaging in the classical sense [21] over the common period of perturbation, deriving an average system.

- 3) Make another time transformation to $s = k\tau$, and taking $k = 0$ derive a singularly perturbed or “reduced” system and a boundary layer system, which is UGAS.

We arrive at an averaged and then reduced system in the new time variable s :

$$\frac{d\theta_r}{ds} = \hat{\mathcal{G}}_{\alpha,r}(\theta_r; a) \quad (32)$$

where

$$\begin{aligned} \hat{\mathcal{G}}_{\alpha,r}(\theta_r; a) = \arg \min_{\xi \in \mathbb{R}^n} \quad & \left\{ \frac{1}{2} \|\xi + \nabla f(\theta_r) + O(a)\|^2 \right\} \\ \text{subj. to} \quad & \left(\frac{\partial g(\theta_r)}{\partial \theta_r} + O(a) \right) \xi \leq -\alpha (g(\theta_r) + O(a)), \\ & \left(\frac{\partial h(\theta_r)}{\partial \theta_r} + O(a) \right) \xi = -\alpha (h(\theta_r) + O(a)). \end{aligned} \quad (33)$$

The terms $O(a)$ represent Lipschitz functions of θ_r that have a norm which can be bounded linearly in $a > 0$ — so they can be made arbitrarily small by choosing a sufficiently small $a > 0$. Since the problem data of the QP (33) are perturbed by Lipschitz quantities, then by [15, Theorem 3.10], for sufficiently small a , we may express the solution $\hat{\mathcal{G}}_{\alpha,r}(\theta_r; a)$ as

$$\frac{d\theta_r}{ds} = \hat{\mathcal{G}}_{\alpha,r}(\theta_r; a) = \mathcal{G}_\alpha(\theta_r) + O(a) \quad (34)$$

where $\mathcal{G}_\alpha(\theta_r)$ is the solution to the unperturbed QP (7), where $a > 0$ is taken sufficiently small. Therefore, the averaged and reduced system dynamics can be written as the exact safe gradient flow system with a small additive disturbance (proportional to the amplitude of the perturbation signals). We give the following without proof due to space limits.

Theorem 4 (Stable and Practically Constrained): Let Assumptions 1, 2, and 3 hold. Then there exists $\beta_\theta \in \mathcal{KL}$ such that: for any $\Delta, \delta, \nu > 0$ there exist $r, \omega_f^*, a^* > 0$, such that for any $\omega_f \in (0, \omega_f^*), a \in (0, a^*)$, there exists $k^*(a) > 0$ such that for any $k \in (0, k^*(a))$ the solutions to (22)-(28) satisfy

$$\|\hat{\theta}(t) - \theta^*\| \leq \beta_\theta \left(\|\hat{\theta}(t_0) - \theta^*\|, k \cdot \omega_f \cdot (t - t_0) \right) + \nu,$$

$$g_i(\theta(t)) \leq g_i(\theta(t_0)) e^{-\alpha k \omega_f (t - t_0)} + O(\delta)$$

$$\|h_j(\theta(t))\| \leq \|h_j(\theta(t_0))\| e^{-\alpha k \omega_f (t - t_0)} + O(\delta)$$

for all $\|\hat{\theta}(t_0) - \theta^*\| \leq \Delta$, $\|G_{g_i}(t_0) - \nabla g_i(\theta(t_0))\| \leq r$, $\|G_{h_j}(t_0) - \nabla h_j(\theta(t_0))\| \leq r$, $\|\eta_{g_i}(t_0) - g_i(\theta(t_0))\| \leq r$, $\|\eta_{h_j}(t_0) - h_j(\theta(t_0))\| \leq r$, for each $i = 1, \dots, m$, $j = 1, \dots, l$ and for all $t \geq t_0 \geq 0$.

Proof idea: under the assumptions discussed in Section III, the dynamics $\mathcal{G}_\alpha(\theta_r)$ are locally Lipschitz, implying that the averaged and reduced dynamics of the ES scheme in (34) are also locally Lipschitz for small $a > 0$, so that the MFCQ assumption still holds for the QP (33). This is crucial as the work of [27] requires this system to be locally Lipschitz, admitting classical solutions (see the conditions in [29, Lemma 1, Lemma 2], which are relied upon in the proof sketch of [27, Theorem 1]). The result [27, Theorem 1]

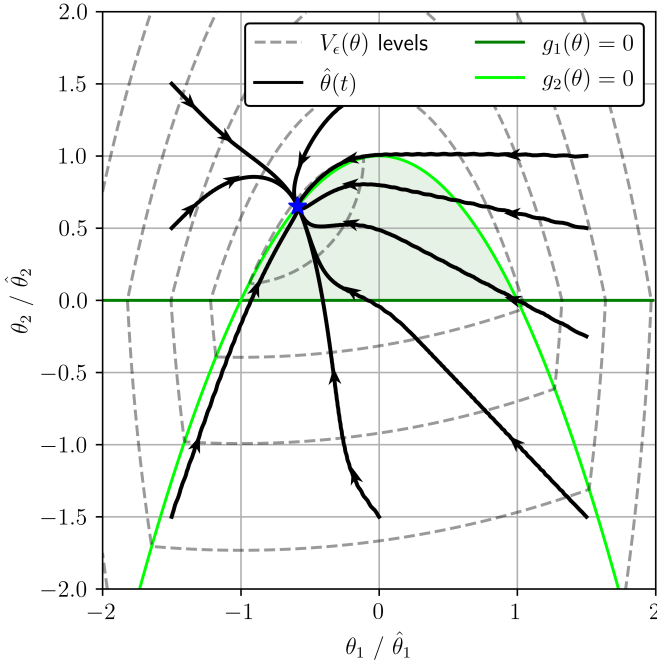


Fig. 1. Trajectories $\hat{\theta}(t)$ plotted from various initial conditions along with the zero level sets of g_1, g_2 and level sets of V_ϵ with $\epsilon = 0.1$. The inequality set $g(\theta) \leq 0$ is shaded in green, and the point θ^* is the blue star.

states that if (34) is SPA stable, then the GMC-ES scheme in (22)-(28) will also be SPA stable. Theorem 4 can be argued for using [27, Theorem 1] and a similar Lyapunov based argument with the same V_ϵ which was used in Section III to study the exact system $\dot{\theta} = \mathcal{G}_\alpha(\theta)$. See [31] as well, which provides a similar analysis albeit under a single inequality constraint (so $m = 1, l = 0$). Finally, since the Lipschitzness of (29) may not hold for the transient if the estimators in (24), (25), (27) and (28) are not close to the true values $\nabla g_i(\theta(t)), \nabla h_j(\theta(t)), g_i(\theta(t))$ and $h_j(\theta(t))$, we restrict the initialization of these estimators to be close to their true values, so that [27, Theorem 1] applies.

V. 2D EXAMPLE

Consider the problem below with unknown functions

$$\begin{aligned} f(\theta) &= (\theta_1 + 1)^2 + (\theta_2 - 1)^2, \\ g_1(\theta) &= -\theta_2, \\ g_2(\theta) &= -1 + \theta_1^2 + \theta_2. \end{aligned}$$

The design constants are chosen as $a = 0.1$, $\alpha = 1$, $k = 0.03$, $\omega_f = 0.5$, $\omega_1 = 10$, and $\omega_2 = 13$. The GMC-ES scheme is integrated with the Euler method and a $\Delta t = 0.048$. Fig. 1 demonstrates the performance of the algorithm in parameter space.

Assumption satisfaction: Assumption 1(ii) is nontrivial (no equalities). Since $\nabla g_1(\theta) = (0, -1)^\top$ and $\nabla g_2(\theta) = (2\theta_1, 1)^\top$, when $\theta_1 \neq 0$ the two gradients are linearly independent, so we can find a single direction d with $\nabla g_i(\theta)^\top d < 0$ for both $i = 1, 2$. When $\theta_1 = 0$ the gradients are opposite, $(0, -1)$ and $(0, 1)$. If both were active at the QP solution ξ^* , the active equalities $-\xi_2^* - \alpha\theta_2 = 0$

and $\xi_2^* + \alpha(-1 + \theta_2) = 0$ implies $-\alpha\theta_2 = \alpha(1 - \theta_2)$, a contradiction; hence $|\mathcal{A}_{QP}(\theta)| \leq 1$. If the active set is $\{i\}$, take $d = -\nabla g_i(\theta)$ so that $\nabla g_i^\top d = -\|\nabla g_i(\theta)\|^2 < 0$; if it is empty, the condition is vacuous. Therefore Assumption 1 holds for all θ . Since f is strongly convex and the feasible set is convex with Slater's condition, (4) has a unique minimizer θ^* , and the KKT conditions are necessary and sufficient, so there exists at least one multiplier pair (u^*, v^*) such that (θ^*, u^*, v^*) satisfies (10), and no other θ does. Hence Assumption 2 holds. One can also check that Assumption 3 also holds.

Finding θ^* : The KKT conditions (10) state there exist $u_1, u_2 \geq 0$ such that

$$\begin{aligned} \text{(stationary)} \quad & 2(\theta_1 + 1) + 2u_2\theta_1 = 0, \\ & 2(\theta_2 - 1) - u_1 + u_2 = 0, \\ \text{(feasibility)} \quad & \theta_2 \geq 0, \quad \theta_1^2 + \theta_2 \leq 1, \\ \text{(complementary)} \quad & u_1\theta_2 = 0, \quad u_2(\theta_1^2 + \theta_2 - 1) = 0. \end{aligned}$$

Suppose g_2 is active at the solution: $\theta_1^2 + \theta_2 = 1$ and $u_2 \geq 0$. If $\theta_2 = 0$ then $\theta_1 = \pm 1$, but neither choice satisfies the stationary condition with $u_1, u_2 \geq 0$, so $\theta_2 > 0$ and thus $u_1 = 0$. With $u_1 = 0$, the stationary condition gives $u_2 = 2(1 - \theta_2)$ and, using $g_2 = 0$ (i.e., $\theta_2 = 1 - \theta_1^2$), we obtain

$$u_2 = 2\theta_1^2, \quad (1 + u_2)\theta_1 + 1 = 0 \Rightarrow 2\theta_1^3 + \theta_1 + 1 = 0.$$

This cubic has a unique real root and so $\theta_1^* = -0.58975$, and $\theta_2^* = 1 - (\theta_1^*)^2 = 0.65219$.

VI. CONCLUSION

We introduced the Generalized Multi-Constraint Extremum Seeking algorithm, extending Safe ES to handle multiple unknown inequality and equality constraints. Using the safe gradient flow with ES-based gradient estimates, we established semiglobal practical asymptotic stability and practical constraint satisfaction. This work broadens the use of ES to general nonlinear programming problems, providing a real-time, constraint-enforcing optimization tool.

REFERENCES

- [1] A. Allibhoy and J. Cortés, "Control barrier function-based design of gradient flows for constrained nonlinear programming," *IEEE Transactions on Automatic Control*, vol. 69, no. 6, pp. 3499–3514, 2024.
- [2] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs for safety critical systems," *IEEE Transactions on Automatic Control*, vol. 62, pp. 3861–3876, 2017.
- [3] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada, "Control barrier functions: Theory and applications," in *2019 18th European control conference (ECC)*. Ieee, 2019, pp. 3420–3431.
- [4] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs for safety critical systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 3861–3876, 2016.
- [5] K. T. Atta, M. Guay, and R. Lucchese, "A geometric phasor extremum seeking control approach with measured constraints," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 1494–1500.
- [6] A. Bacciotti and F. Ceragioli, "Stability and stabilization of discontinuous systems and nonsmooth lyapunov functions," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 4, pp. 361–376, 1999.

- [7] D. P. Bertsekas, "Nonlinear programming," *Journal of the Operational Research Society*, vol. 48, no. 3, pp. 334–334, 1997.
- [8] F. Blanchini and S. Miani, *Set-Theoretic Methods in Control*. Springer, 2008, vol. 78.
- [9] X. Chen, J. I. Poveda, and N. Li, "Continuous-time zeroth-order dynamics with projection maps: Model-free feedback optimization with safety guarantees," *arXiv preprint arXiv:2303.06858*, 2023.
- [10] Y. Chen, L. Cothren, J. Cortés, and E. Dall'Anese, "Online regulation of dynamical systems to solutions of constrained optimization problems," *IEEE Control Systems Letters*, vol. 7, pp. 3789–3794, 2023.
- [11] F. H. Clarke, *Optimization and Nonsmooth Analysis*. SIAM, 1990.
- [12] F. H. Clarke, Y. S. Ledyaev, R. J. Stern, and R. Wolenski, *Nonsmooth analysis and control theory*. Springer, 1998.
- [13] A. Colot, Y. Chen, B. Cornélusse, J. Cortés, and E. Dall'Anese, "Optimal power flow pursuit via feedback-based safe gradient flow," *IEEE Transactions on Control Systems Technology*, vol. 33, pp. 658–670, 2025.
- [14] J. Cortés, "Discontinuous dynamical systems," *IEEE Control Systems Magazine*, vol. 28, no. 3, pp. 36–73, 2008.
- [15] J. W. Daniel, "The continuity of metric projections as functions of the data," *Journal of Approximation Theory*, vol. 12, no. 3, pp. 234–239, 1974.
- [16] G. Delimpaltadakis, J. Cortés, and W. P. M. H. Heemels, "Continuous approximations of projected dynamical systems via Control Barrier Functions," *IEEE Transactions on Automatic Control*, vol. 70, no. 1, pp. 681–688, 2024.
- [17] H.-B. Dürr and C. Ebenbauer, "A smooth vector field for saddle point problems," in *2011 50th IEEE Conference on Decision and Control and European Control Conference*. IEEE, 2011, pp. 4654–4660.
- [18] H.-B. Dürr, C. Zeng, and C. Ebenbauer, "Saddle point seeking for convex optimization problems," *IFAC Proceedings Volumes*, vol. 46, no. 23, pp. 540–545, 2013.
- [19] J. Feng, W. Cui, J. Cortés, and Y. Shi, "Bridging transient and steady-state performance in voltage control: A reinforcement learning approach with safe gradient flow," *IEEE Control Systems Letters*, vol. 7, pp. 2845–2850, 2023.
- [20] J. Gauvin, "A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming," *Mathematical Programming*, vol. 12, no. 1, pp. 136–138, 1977.
- [21] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 2002.
- [22] M. Krstic and H.-H. Wang, "Stability of extremum seeking feedback for general nonlinear dynamic systems," *Automatica*, vol. 36, no. 4, pp. 595–601, 2000.
- [23] C. Labar, E. Garone, M. Kinnaert, and C. Ebenbauer, "Constrained extremum seeking: a modified-barrier function approach," *IFAC-PapersOnLine*, vol. 52, no. 16, pp. 694–699, 2019.
- [24] P. Mestres, A. Marzabal, and J. Cortés, "Anytime safe reinforcement learning," in *7th Annual Learning for Dynamics & Control Conference*. PMLR, 2025, pp. 221–232.
- [25] M. Nagumo, "Über die lage der integralkurven gewöhnlicher differentialgleichungen," *Proceedings of the Physico-Mathematical Society of Japan. 3rd Series*, vol. 24, pp. 551–559, 1942.
- [26] A. Nagurney and D. Zhang, *Projected dynamical systems and variational inequalities with applications*. Springer Science & Business Media, 2012, vol. 2.
- [27] D. Nesić, Y. Tan, W. H. Moase, and C. Manzie, "A unifying approach to extremum seeking: Adaptive schemes based on estimation of derivatives," in *49th IEEE conference on Decision and Control (CDC)*. IEEE, 2010, pp. 4625–4630.
- [28] A. Scheinker, "100 years of extremum seeking: A survey," *Automatica*, vol. 161, p. 111481, 2024.
- [29] Y. Tan, D. Nesić, and I. Mareels, "On non-local stability properties of extremum seeking control," *IFAC Proceedings Volumes*, vol. 38, no. 1, pp. 550–555, 2005.
- [30] P. Wieland and F. Allgöwer, "Constructive safety using control barrier functions," *IFAC Proceedings Volumes*, vol. 40, no. 12, pp. 462 – 467, 2007, 7th IFAC Symposium on Nonlinear Control Systems.
- [31] A. Williams, M. Krstic, and A. Scheinker, "Semiglobal safety-filtered extremum seeking with unknown CBFs," *IEEE Transactions on Automatic Control*, vol. 70, no. 3, pp. 1698–1713, 2025.