# Over-Approximating Minimizer Sets of Constrained Convex Programs with Parametric Uncertainty via Reachability Analysis

Brendan Gould[1⋆], Chih-Yuan Chiu[1⋆], Antoine P. Leeman[2],
Kyriakos G. Vamvoudakis[1], Samuel Coogan[1], Glen Chou[1]

***Abstract*— We study the solution set of a strongly convex optimization problem whose cost depends on uncertain, bounded parameters. We compute a certified outer approximation of the corresponding set of optimizers, using convergence properties of the projected gradient descent (PGD) algorithm. By treating the cost parameter as constant but unknown, we interpret PGD as an uncertain dynamical system and analyze its forward reachable sets. Since PGD converges exponentially to the unique optimizer for each fixed parameter, these reachable sets provide outer approximations of the optimizer set, with an exponentially decaying error bound. We apply system-level synthesis (SLS) on the PGD dynamics to optimize the step-size sequence and obtain reachable-set over-approximations. Our method outperforms existing baselines in over-approximating, with low conservativeness, the minimizer sets of convex programs with uncertain costs and high-dimensional decision variables.**

## I. Introduction

In many controls and robotics contexts, optimization supports planning and decision-making, but the data informing these decisions is frequently *uncertain*. For instance, to operate safely in traffic, an autonomous vehicle must anticipate the motion of nearby cars and pedestrians, whose intent can only be partially inferred from observations [1]–[3]. This intent uncertainty naturally induces uncertainty in predictions of agents' motion, which could cause unsafe behavior. Separately, quasistatic [4] and deformable object [5] contact dynamics models in robotic manipulation are convex conic programs, where uncertainty in the mass matrix manifests as uncertainty in the cost function. We present a method for over-approximating the minimizer set of a convex program whose cost is characterized by uncertain parameters. By robustly bounding the solution set for all possible parameter realizations, our work guides safe downstream planning.

The tractability of computing this over-approximation varies significantly depending on the structure of the objective, constraint set, and parameter set. The minimizer set is only generally obtainable in closed form under restrictive assumptions. Sensitivity analysis techniques bound the minimizer's variation as the parameter changes, but such bounds are often highly conservative (Sec. V-B.1), limiting downstream utility. Set-based reachability and robust optimal control study how uncertainty propagates through dynamical systems [6], [7]. Where they seek a single solution minimizing worst-case cost under uncertainty, we over-approximate the full set of minimizers across all parameter realizations.

We use system-level synthesis (SLS) to compute forward reachable-set over-approximations of projected gradient descent (PGD) dynamics with low conservativeness. We then leverage known convergence properties of PGD [8] to convert these reachable sets into certified outer approximations of the corresponding minimizer set. Our contributions are fourfold:

1) We prove that, under standard regularity assumptions, the PGD dynamics' (true) forward reachable sets converge exponentially to the minimizer set. We thus recast our goal as computing forward reachable sets of the PGD dynamics with a fixed but unknown parameter.
2) For *differentiable* PGD dynamics, we over-approximate reachable sets of PGD trajectories with the SLS-based reachability framework [9], scaling tractably to high dimensions with low over-approximation error [9]–[12].
3) To compute reachable sets for *non-differentiable* PGD dynamics, we apply the sampling-based smoothing method in [13] to construct a differentiable approximation of PGD for which the SLS techniques in [9] apply.
4) Our method outperforms sensitivity-analysis baselines in over-approximating parametric minimizer sets of optimization programs with low conservativeness and scales to problems with 64-dimensional variables.

Our work is related to robust optimization, which studies decision-making under worst-case realizations of uncertain data [14]–[16]. However, instead of a single robust decision, we characterize the set of decisions that are optimal for some admissible parameter value. In this sense, prior work on multi-parametric programming and solution-map sensitivity analysis, studying how minimizers change as problem data vary [17]–[20], is closer to our objective. Multi-parametric programming can explicitly characterize optimizers over critical regions for parametric linear and quadratic programs [21], but becomes computationally expensive in high dimensions. In contrast, we compute global over-approximations of the full minimizer set using reachability techniques [22], [23].

**Notation**: For $m, n \in \mathbb{Z}^+$, let $[n] := \{1, \cdots, n\}$ and $[0, n] := \{0\} \cup [n]$, and denote the $n \times n$ identity matrix, $n \times n$ zero matrix, $m \times n$ zero matrix, and the zero vector in $\mathbb{R}^n$ by $I_{n\times n}$, $O_n$, $O_{m\times n}$, and $\mathbf{0}_n$ respectively, with indices

⋆Equal contribution. [1]Georgia Institute of Technology, Atlanta, GA, USA. {bgould6} at gatech dot edu. [2]ETH Zurich, IDSC, Switzerland.

This material is based on work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2545526. 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.

Supported in part by the National Science Foundation under awards #2333488 and SLES-2415479, an Early Career Faculty grant from NASA's Space Technology Research Grants Program, NASA ULI under grant 80NSSC25M7104, and by ARO under grant No. W911NF-24-1-0174.

omitted when clear from context. The block-diagonal matrix constructed from given matrix sub-blocks $M_1, \cdots, M_n$ is blkdiag$\{M_1, \cdots, M_n\}$; when $M_1, \cdots, M_n$ are all scalar, we write diag$\{\cdot\}$. Given $p \in [1, \infty]$ and $M \in \mathbb{R}^{n \times d}$, the unit $p$-norm ball in $\mathbb{R}^d$ is $\mathbb{B}_p^d$, and $M\mathbb{B}_p^d := \{Mx : x \in \mathbb{B}_p^d\}$. Given sets $S, S' \subseteq \mathbb{R}^n$, we define the *diameter* of $S$ by $\text{diam}(S) := \sup_{s,s' \in S} \|s - s'\|_2$ and the *convex hull* of $S$ by $\text{conv}(S)$. The *Minkowski sum* of $S$ and $S'$ is $S \oplus S' := \{s + s' : s \in S, s' \in S'\}$, and the Hausdorff distance between $S$ and $S'$ is $\text{dist}(S, S') := \inf\{\epsilon > 0 : S \subseteq S' \oplus \epsilon I_n \mathbb{B}_2^n, S' \subseteq S \oplus \epsilon I_n \mathbb{B}_2^n\}$.

## II. Problem Statement

Consider an objective $J : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ mapping a variable $\xi \in S_{\mathcal{C}} \subseteq \mathbb{R}^n$ and a parameter $\theta \in \Theta \subset \mathbb{R}^d$ to a real number. We impose the following regularity assumptions.

*Assumption 1:* All partial second derivatives of $J(\xi, \theta)$ exist and are Lipschitz continuous. There exist scalars $m, L > 0$, with $m \leq L$, such that $J(\cdot, \theta)$ is $L$-smooth and $m$-strongly convex, i.e., for each $\theta \in \Theta$ and $\xi, \xi' \in \mathbb{R}^n$:

$$J(\xi', \theta) \geq J(\xi, \theta) + \nabla J(\xi, \theta)^\top (\xi' - \xi) + \tfrac{1}{2} m \|\xi' - \xi\|_2^2,$$
$$J(\xi', \theta) \leq J(\xi, \theta) + \nabla J(\xi, \theta)^\top (\xi' - \xi) + \tfrac{1}{2} L \|\xi' - \xi\|_2^2.$$

*Remark 1:* Assumption 1 is satisfied by strongly convex optimal control problems, including LQR-based trajectory planning and model predictive control.

*Assumption 2:* $S_{\mathcal{C}}$ is convex and closed in $\mathbb{R}^n$.

*Assumption 3:* $\Theta$ is convex and compact in $\mathbb{R}^d$.

Under Assumption 1, for each $\theta \in \Theta$, $J(\cdot, \theta)$ has a unique minimizer over $S_{\mathcal{C}}$, denoted $\xi^\star(\theta)$ below [8, Thm. 2.8].

*Problem 1:* We aim to construct an outer approximation of the minimizer set $S_\xi^\star$, defined by:

$$S_\xi^\star := \textstyle\bigcup_{\theta \in \Theta} \arg\min_\xi J(\xi, \theta) = \big\{\xi^\star(\theta) : \theta \in \Theta\big\}. \quad (1)$$

## III. Preliminaries on PGD and SLS

We introduce fundamental properties of the PGD dynamics (Sec. III-A) and SLS output feedback error controller design.

### A. Iterative Gradient-Based Descent Methods

We review the convergence of PGD on strongly convex problems. Consider $f_{\text{PGD}} : \mathbb{R}^n \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^n$, with steplength sequence $(\alpha_k : k \geq 0)$ and Euclidean projection onto $S_{\mathcal{C}}$ $\text{proj}_{S_{\mathcal{C}}}$ (i.e., $\text{proj}_{S_{\mathcal{C}}}(\xi) := \arg\min_{\xi' \in S_{\mathcal{C}}} \|\xi' - \xi\|_2$ for $\xi \in \mathbb{R}^n$)[1]. Then, for any fixed $\theta \in \Theta$ the PGD updates are:

$$\xi_{k+1} = f_{\text{PGD}}(\xi_k, \theta, \alpha_k) = \text{proj}_{S_{\mathcal{C}}}\left(\xi_k - \alpha_k \nabla_\xi J(\xi_k, \theta)\right). \quad (2)$$

From optimization theory, under Assumptions 1 and 2, for any fixed $\theta \in \Theta$, the iterative PGD dynamics (2) on $J$ converge exponentially to the global minimizer of $J$ over $S_{\mathcal{C}}$ [8, Sec. 7.3.3]. Prop. 2 strengthens this; the convergence rate is agnostic to the specific parameter value $\theta$ in $\Theta$.

*Proposition 1:* Under Assumptions 1–3, $S_\xi^\star$ is bounded.

Prop. 1 is well-known in optimization theory [8, Sec. 2].

*Assumption 4:* There is a known, bounded set $S_\xi^0 \subseteq S_{\mathcal{C}}$ containing the initial iterate $\xi_0$ of the PGD dynamics (2).

[1]As proved in [8, Ch. 7], under Assumption 2, the projection $\text{proj}_{S_{\mathcal{C}}}(\xi)$ of any $\xi \in \mathbb{R}^n$ onto $S_{\mathcal{C}}$ is a unique, well-defined element of $S_{\mathcal{C}}$.

For any $k \geq 0$, given $\xi_0 \in S_\xi^0$, $\theta \in \Theta$, and $\alpha_{0:k-1} := (\alpha_0, \cdots, \alpha_{k-1}) \in \mathbb{R}^k$, let $\phi_k(\xi_0, \theta, \alpha_{0:k-1})$ be the $k$-th PGD iterate. The forward reachable set for (2) at iteration $k$ is:

$$S_{\xi,k}(\xi_0, \alpha_{0:k-1}) := \{\phi_k(\xi_0, \theta, \alpha_{0:k-1}) : \theta \in \Theta\}. \quad (3)$$

*Proposition 2 (**Uniform Convergence of PGD Dynamics**):* Suppose Assumptions 1, 2, and 3 hold. Fix $\xi_0 \in S_\xi^0$, let $c_\alpha, C_\alpha \in (0, \frac{2}{L})$ be given such that $c_\alpha \leq C_\alpha$, and define:

$$\gamma := \max_{\alpha \in [c_\alpha, C_\alpha]} \big\{ \max\{|1 - \alpha m|, |1 - \alpha L|\} \big\} < 1. \quad (4)$$

Fix $\alpha_k \in [c_\alpha, C_\alpha]$, $k \geq 0$. Then, for any $\theta \in \Theta$, $k \in \mathbb{N}$:

$$\|\phi_k(\xi_0, \theta, \alpha_{0:k-1}) - \xi^\star(\theta)\|_2 \leq \gamma^k \cdot \|\xi_0 - \xi^\star(\theta)\|_2. \quad (5)$$

*Proof:* Fix $\xi_0 \in S_\xi^0$, $\theta \in \Theta$, and $\alpha_{0:k-1} \in [c_\alpha, C_\alpha]^k$, and write $\xi_j = \phi_j(\xi_0, \theta, \alpha_{0:j-1})$ for $j \in [0, k]$. From [8, Prop. 7.8], $\forall\ \alpha_k \geq 0$, $\xi^\star(\theta)$ is the unique fixed point of the PGD updates (2). Thus, for any $\theta \in \Theta$ and $j \geq 0$:

$$\begin{aligned}
\|\xi_{j+1} - \xi^\star(\theta)\|_2 &= \big\|\text{proj}_{S_{\mathcal{C}}}\big(\xi_j - \alpha_j \nabla J(\xi_j, \theta)\big) \\
&\quad - \text{proj}_{S_{\mathcal{C}}}\big(\xi^\star(\theta) - \alpha_j \nabla J\big(\xi^\star(\theta), \theta\big)\big)\big\|_2 \\
&\leq \max\big\{|1 - \alpha_j m|, |1 - \alpha_j L|\big\} \cdot \|\xi_j - \xi^\star(\theta)\|_2 \quad (6) \\
&\leq \gamma \cdot \|\xi_j - \xi^\star(\theta)\|_2,
\end{aligned}$$

where (6) follows from [8, Eqn. (7.8)]. Recursively applying the above for $j \in [0, k-1]$, we obtain (5). ∎

Prop. 2 recasts Problem 1 as forward reachability for (2).

*Theorem 1:* (**Over-Approximating $S_\xi^\star$ using the Forward Reachable Sets of PGD Dynamics**) Fix $\xi_0 \in S_\xi^0$ and $\alpha_{0:k-1} \in [c_\alpha, C_\alpha]^k$ arbitrarily, and define $\beta_k(\xi_0) := \gamma^k \cdot \max_{\theta \in \Theta} \|\xi^\star(\theta) - \xi_0\|_2$ for each $k \geq 0$. Then:

$$\text{dist}\big(S_\xi^\star, S_{\xi,k}(\xi_0, \alpha_{0:k-1})\big) \leq \beta_k(\xi_0), \qquad \forall\ k \geq 0. \quad (7)$$

*Proof:* Fix $\xi_0 \in S_\xi^0$ and $\alpha_{0:k-1} \in [c_\alpha, C_\alpha]^k$. By maximizing the right-hand side of (5) over all $\theta \in \Theta$, we obtain that $\|\phi_k(\xi_0, \theta, \alpha_{0:k-1}) - \xi^\star(\theta)\|_2 \leq \beta_k(\xi_0)$ for each $\theta \in \Theta$, yielding (7). ∎

*Remark 2:* $\beta_k(\xi_0)$, though $\xi_0$-dependent, is continuous in $\xi_0$ and thus uniformly bounded over $S_\xi^0$ by Assumption 4.

### B. SLS-based Output Feedback Error Controller Design

To over-approximate $S_\xi^\star$ via Thm. 1, we leverage SLS [9], [10] to compute a forward reachable set for (2), with an uncertain but fixed $\theta \in \Theta$ (8b). Though many methods exist for computing reachable sets [23], [24], we use [9], [10] due to their tractability for high-dimensional dynamics [25].

We consider the following dynamics over a horizon $[0, N]$, with variable-parameter tuples $\zeta := (\xi, \theta)$ as states, PGD steplengths $\alpha$ as control inputs, and variables $\xi$ as outputs:

$$\xi_{k+1} = f_{\text{PGD}}(\xi_k, \theta_k, \alpha_k), \qquad \forall\ k \in [0, N-1], \quad (8a)$$
$$\theta_{k+1} = \theta_k, \qquad \forall\ k \in [0, N-1], \quad (8b)$$
$$y_k = \xi_k, \qquad \forall\ k \in [0, N]. \quad (8c)$$

For notational ease, we define $f : \mathbb{R}^n \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{n+d}$ by $f(\xi, \theta, \alpha) := \big(f_{\text{PGD}}(\xi, \theta, \alpha), \theta\big)$, simplifying (8a)-(8b) to:

$$\zeta_{k+1} = f(\zeta_k, \alpha_k), \qquad \forall\ k \in [0, N-1]. \quad (9)$$

We also write $\boldsymbol{\zeta} := (\zeta_0, \cdots, \zeta_N) \in \mathbb{R}^{(n+d)(N+1)}$, and for any $k_1, k_2 \in [0, N]$ with $k_1 \leq k_2$, we write $\boldsymbol{\zeta}_{k_1:k_2} := (\zeta_{k_1}, \cdots, \zeta_{k_2}) \in \mathbb{R}^{(n+d)(k_2-k_1+1)}$.

To satisfy Prop. 2 and ensure the convergence of PGD dynamics, we enforce that $(\alpha_k : k \in [0, N])$ satisfy:

$$c_\alpha \leq \alpha_k \leq C_\alpha, \quad \forall\ k \in [0, N]. \tag{10}$$

Define $a_1 := (\mathbf{0}_{n+d}, 1), a_2 := (\mathbf{0}_{n+d}, -1) \in \mathbb{R}^{n+d+1}$ and $b_1 := -C_\alpha, b_2 := c_\alpha$. We can then rewrite (10) as $a_p^\top(\zeta_k, \alpha_k) + b_p \leq 0$ for each $p \in [2]$.

To construct reachable sets for PGD (8) over $\theta \in \Theta$ with low conservativeness, we design causal feedback laws $\pi_k : \mathbb{R}^{n(k+1)} \to \mathbb{R}$ for each $k \in [0, N]$, mapping state history $\boldsymbol{\xi}_{0:k}$ to a steplength $\alpha_k \in [c_\alpha, C_\alpha]$. A *fixed* steplength may be used, but doing so may result in looser approximations of $S_\xi^\star$. Since optimizing over all causal feedback laws is intractable, we mimic [9], [10], and reformulate the true nonlinear system (8), encoding initial uncertainty in $\theta_0$ using linear time-varying (LTV) error dynamics (16). We then design (a) nominal state and control trajectories $\hat{\boldsymbol{\zeta}} := (\hat{\xi}_0, \hat{\theta}_0, \cdots, \hat{\xi}_N, \hat{\theta}_N) \in \mathbb{R}^{(n+d)(N+1)}$ and $\hat{\boldsymbol{\alpha}} := (\hat{\alpha}_0, \cdots, \hat{\alpha}_N) \in \mathbb{R}^{N+1}$, satisfying

$$\hat{\zeta}_{k+1} = f(\hat{\zeta}_k, \hat{\alpha}_k), \quad \forall\ k \in [0, N-1], \tag{11}$$

and (b) gains $S_K := \{K_{k,j} \in \mathbb{R}^{1\times n},\ k, j \in [0, N], k \geq j\}$ parameterizing the causal affine output error feedback map:

$$\pi_k(\boldsymbol{y}_{0:k}) = \hat{\alpha}_k + \textstyle\sum_{j=0}^k K_{k,j}(\xi_{k-j} - \hat{\xi}_{k-j}), \tag{12}$$

*Remark 3:* Designing $\hat{\boldsymbol{\zeta}}$ requires choosing an initial nominal state $\hat{\xi}_0$ and parameter $\hat{\theta}_0$. We set $\hat{\xi}_0 = \xi_0$ for simplicity, and $\hat{\theta}_0$ is fixed across iterations per (8b).

## IV. Methods

We outer-approximate $S_\xi^\star$ by computing reachable sets for PGD with SLS (Thm. 1). For differentiable PGD dynamics, we form a LTV system that propagates error between nominal and true trajectories (Sec. IV-A) and bound that error across iterations (Sec. IV-B). We generalize to non-differentiable PGD dynamics via a smoothing technique (Sec. IV-C).

### A. LTV Error System for Smooth PGD Dynamics

We derive the LTV error system for (8), when $S_\mathcal{C}$ is $\mathbb{R}^n$ or an affine subspace of $\mathbb{R}^n$, so that $\text{proj}_{S_\mathcal{C}}(\cdot)$ is affine and $f$ in (8) is differentiable under Assumption 1. The non-differentiable case is treated in Sec. IV-C.

The LTV error system admits, across iterations $k \in [0, N]$, the state errors $\Delta\zeta_k := \zeta_k - \hat{\zeta}_k$ as states and steplength errors $\Delta\alpha_k := \alpha_k - \hat{\alpha}_k$ as controls. Its gain and actuation matrices are the Jacobian linearization of (9) at $(\hat{\zeta}_k, \hat{\alpha}_k)$:

$$A_k(\hat{\zeta}_k, \hat{\alpha}_k) := \tfrac{\partial f}{\partial \zeta}(\hat{\zeta}_k, \hat{\alpha}_k), \tag{13a}$$

$$B_k(\hat{\zeta}_k, \hat{\alpha}_k) := \tfrac{\partial f}{\partial \alpha}(\hat{\zeta}_k, \hat{\alpha}_k), \tag{13b}$$

and the *linearization error* $r_k(\zeta_k, \alpha_k, \hat{\zeta}_k, \hat{\alpha}_k) \in \mathbb{R}^{n+d}$ is:

$$\begin{aligned} r_k(\zeta_k, \alpha_k, \hat{\zeta}_k, \hat{\alpha}_k) &= f(\zeta_k, \alpha_k) - f(\hat{\zeta}_k, \hat{\alpha}_k) \\ &\quad - A_k(\hat{\zeta}_k, \hat{\alpha}_k)\Delta\zeta_k - B_k(\hat{\zeta}_k, \hat{\alpha}_k)\Delta\alpha_k. \end{aligned} \tag{14}$$

We then use (13) and (14) to describe the propagation of parameter errors as the following LTV system while expressing each linearization error as a disturbance $d_k$:

$$d_k = r_k(\zeta_k, \alpha_k, \hat{\zeta}_k, \hat{\alpha}_k), \tag{15}$$

$$\begin{aligned} \Delta\zeta_{k+1} &:= \zeta_{k+1} - \hat{\zeta}_{k+1} = f(\zeta_k, \alpha_k) - f(\hat{\zeta}_k, \hat{\alpha}_k) \\ &= A_k(\hat{\zeta}_k, \hat{\alpha}_k)\Delta\zeta_k + B_k(\hat{\zeta}_k, \hat{\alpha}_k)\Delta\alpha_k + d_k, \end{aligned} \tag{16}$$

Meanwhile, we define the *control error* $\Delta\alpha_k := \alpha_k - \hat{\alpha}_k$ for each $k \in [0, N]$ and the *output matrix* $C := \begin{bmatrix} I_n & O_{n+d} \end{bmatrix} \in \mathbb{R}^{n\times(n+d)}$, and derive from (12) that:

$$\Delta\alpha_k = \textstyle\sum_{j=0}^k K_{k,j} C \Delta\zeta_{k-j}, \ \ \forall\ k \in [0, N]. \tag{17}$$

SLS recasts the design of the output feedback gains $S_K$ as the design of a *system response*, mapping disturbances $\boldsymbol{d}$ directly to state and control errors. We stack the LTV system dynamics (16) and (17) across iterations $k \in [0, N]$. Let $\boldsymbol{d} := (\Delta\zeta_0, d_0, \cdots, d_{N-1})$, and define $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{K}$, and $\mathcal{Z}$ by:

$$\mathcal{A} := \text{blkdiag}\{A_0, \cdots, A_{N-1}, O_{n+d}\}, \tag{18a}$$

$$\mathcal{B} := \text{blkdiag}\{B_0, \cdots, B_{N-1}, \mathbf{0}_{n+d}\}, \tag{18b}$$

$$\mathcal{C} := \text{blkdiag}\{C, \cdots, C, C\}, \tag{18c}$$

$$\mathcal{K} := \begin{bmatrix} K_{0,0} & O & \cdots & O \\ K_{1,0} & K_{1,1} & \ddots & O \\ \vdots & \vdots & \ddots & \vdots \\ K_{N,0} & K_{N,1} & \ddots & K_{N,N} \end{bmatrix}, \tag{18d}$$

$$\mathcal{Z} := \begin{bmatrix} O_{(n+d)\times(n+d)N} & O_{n+d} \\ I_{(n+d)N} & O_{(n+d)N\times(n+d)} \end{bmatrix} \tag{18e}$$

We use (18) to rewrite (16)-(17) across $k \in [0, N]$ as:

$$\Delta\boldsymbol{\zeta} = \mathcal{Z}\mathcal{A}\Delta\boldsymbol{\zeta} + \mathcal{Z}\mathcal{B}\Delta\boldsymbol{\alpha} + \boldsymbol{d}, \tag{19a}$$

$$\Delta\boldsymbol{\alpha} = \mathcal{K}\mathcal{C}\Delta\boldsymbol{\zeta}, \tag{19b}$$

which can be further written as:

$$\begin{bmatrix} \Delta\boldsymbol{\zeta} \\ \Delta\boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} \Phi_{\zeta d} & \Phi_{\zeta v} \\ \Phi_{\alpha d} & \Phi_{\alpha v} \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \\ \mathbf{0}_{n(N+1)} \end{bmatrix} := \Phi \begin{bmatrix} \boldsymbol{d} \\ \mathbf{0}_{n(N+1)} \end{bmatrix}, \tag{20}$$

where the constituent blocks of the *system response* $\Phi \in \mathbb{R}^{(n+d+1)(N+1)\times(2n+d)(N+1)}$ are:

$$\Phi_{\zeta d} := (I_{(n+d)(N+1)} - \mathcal{Z}\mathcal{A} - \mathcal{Z}\mathcal{B}\mathcal{K}\mathcal{C})^{-1} \tag{21a}$$

$$\Phi_{\zeta v} := (I_{(n+d)(N+1)} - \mathcal{Z}\mathcal{A} - \mathcal{Z}\mathcal{B}\mathcal{K}\mathcal{C})^{-1}\mathcal{Z}\mathcal{B}\mathcal{K} \tag{21b}$$

$$\Phi_{\alpha d} := \mathcal{K}\mathcal{C}(I - \mathcal{Z}\mathcal{A} - \mathcal{Z}\mathcal{B}\mathcal{K}\mathcal{C})^{-1} \tag{21c}$$

$$\Phi_{\alpha v} := \mathcal{K}\mathcal{C}(I - \mathcal{Z}\mathcal{A} - \mathcal{Z}\mathcal{B}\mathcal{K}\mathcal{C})^{-1}\mathcal{Z}\mathcal{B}\mathcal{K} + \mathcal{K}. \tag{21d}$$

Conversely, from (21), we also obtain that:

$$\mathcal{K} = \Phi_{\alpha v} - \Phi_{\alpha d}\Phi_{\zeta d}^{-1}\Phi_{\zeta v}. \tag{22}$$

As proved in [26], [27] and described below, the system response $\Phi$ parameterizes all causal output feedback gains $\mathcal{K}$.

*Proposition 3:* [27, Prop. 1] If there exist output feedback gains $\mathcal{K}$ of the form (18d) satisfying (19), then $\Phi_{\zeta d}$, $\Phi_{\zeta v}$, $\Phi_{\alpha d}$, and $\Phi_{\alpha v}$, as computed by (21), satisfy:

$$\begin{bmatrix} I - \mathcal{Z}\mathcal{A} & -\mathcal{Z}\mathcal{B} \end{bmatrix} \Phi = \begin{bmatrix} I & O \end{bmatrix}, \tag{23a}$$

$$\Phi \begin{bmatrix} I - \mathcal{Z}\mathcal{A} \\ -\mathcal{C} \end{bmatrix} = \begin{bmatrix} I \\ O \end{bmatrix}, \tag{23b}$$

$$\Phi_{\zeta d}, \Phi_{\zeta v}, \Phi_{\alpha d}, \Phi_{\alpha v} \text{ are lower block-triangular.} \tag{23c}$$

Conversely, if $\Phi_{\zeta d}$, $\Phi_{\zeta v}$, $\Phi_{\alpha d}$, and $\Phi_{\alpha v}$ satisfy (23), then $\mathcal{K}$, as computed by (22), is lower block triangular.

Below, given $\Phi$ satisfying (23), for each $k \in [0, N]$ and

$j \in [0,k]$, we respectively define $(\Phi_{\zeta d})_{k,j}$ and $(\Phi_{\alpha d})_{k,j}$ to be the $(k,j)$-th block of $\Phi_{\zeta d}$ and $\Phi_{\alpha d}$ in the sense of (18d).

### B. Linearization Error Bounds

Reachable sets for (8) are computed by bounding the error between (13) and $f$ with a simulated disturbance (15). Define

$$S_{\xi,\text{CH}} := \text{conv}\big\{S_{\xi,k}(\xi_0,\alpha_{0:k-1}) : k \geq 0, \xi_0 \in S_\xi^0, \alpha_{0:k-1} \in [c_\alpha, C_\alpha]^k\big\} \tag{24}$$

as the convex hull of all possible rollouts of the PGD dynamics (2) from any $\xi_0 \in S_\xi^0$. Prop. 4 bounds the curvature of $f$ over $S_{\xi,\text{CH}} \times \Theta$ via a Lipschitz argument.

*Proposition 4:* ($f$ **Admits Lipschitz Gradients**) Suppose Assumptions 1, 3, and 4 hold, and $S_\mathcal{C}$ is $\mathbb{R}^n$ or an affine subspace of $\mathbb{R}^n$. For each $i \in [n+d]$, let $f_i : \mathbb{R}^{n+d+1} \to \mathbb{R}$ denote the $i$-th component of the dynamics $f$ in (8). Then, $\forall\ i \in [n+d]$, there exists some $\mu_i > 0$ such that, $\forall\ (\zeta,\alpha),(\zeta',\alpha') \in S_{\xi,\text{CH}} \times \Theta \times [c_\alpha, C_\alpha]$:

$$\|\nabla f_i(\zeta,\alpha) - \nabla f_i(\zeta',\alpha')\|_1 \leq \mu_i \|(\zeta,\alpha) - (\zeta',\alpha')\|_\infty. \tag{25}$$

*Proof:* First note that $S_{\xi,\text{CH}}$ is bounded. By Thm. 1, $\text{dist}(S_{\xi,k}(\xi_0,\alpha_{0:k-1}), S_\xi^\star) \leq \beta_k(\xi_0)$ for all $k \geq 1$. Since $S_\xi^0$ (which contains $S_{\xi,0}(\xi_0,\alpha_{0:k-1}) = \{\xi_0\}$) and $S_\xi^\star$ are bounded (Assumption 4, Prop. 1) and $\beta_k(\xi_0) \searrow 0$ as $k \to \infty$, the sets $S_{\xi,k}(\xi_0,\alpha_{0:k-1})$ are uniformly bounded on $k \geq 0$, $\xi_0 \in S_\xi^0$, and $\alpha_{0:k-1} \in [c_\alpha, C_\alpha]^k$, so $S_{\xi,\text{CH}}$ is bounded.

Next, since $S_\mathcal{C}$ is $\mathbb{R}^n$ or an affine subspace of $\mathbb{R}^n$, the map $\text{proj}_{S_\mathcal{C}}$ is affine. Thus, from (2) and (8), each $\nabla f_i$ is bilinear in $\alpha$ and the first-order and second-order partial derivatives of $J$. It suffices that the second-order partial derivatives of $J$ are Lipschitz continuous and bounded over $S_{\xi,\text{CH}} \times \Theta$, which follows from Assumptions 1 and 3. ∎

Using Prop. 4, we now bound the disturbance $d_k = r(\zeta_k,\alpha_k,\hat{\zeta}_k,\hat{\alpha}_k)$ in (15), similarly to [9, Prop. III.1].

*Proposition 5 (**Linearization Error Bound**):* Suppose Assumptions 1, 3, and 4 hold, and $S_\mathcal{C}$ is $\mathbb{R}^n$ or an affine subspace of $\mathbb{R}^n$. For each $k \in [0,N]$ and $i \in [n+d]$, let $r_{k,i} : \mathbb{R}^{2(n+d+1)} \to \mathbb{R}$ be the $i$-th coordinate of $r_k$. Then for any $\xi_0 \in S_\xi^0$, $\alpha_{0:k}, \hat{\alpha}_{0:k} \in [c_\alpha, C_\alpha]^{k+1}$, any $\zeta_k = (\xi_k,\theta_k) \in S_{\xi,k}(\xi_0,\alpha_{0:k-1}) \times \Theta$ and $\hat{\zeta}_k = (\hat{\xi}_k,\hat{\theta}_k) \in S_{\xi,k}(\xi_0,\hat{\alpha}_{0:k-1}) \times \Theta$, and any $i \in [n+d]$, we have:

$$|r_{k,i}(\zeta_k,\alpha_k,\hat{\zeta}_k,\hat{\alpha}_k)| \leq \mu_i \cdot \|(\Delta\zeta_k,\Delta\alpha_k)\|_\infty^2. \tag{26}$$

*Proof:* By the Mean Value Theorem, for some $\lambda_i \in [0,1]$,

$$\begin{aligned} &f_i(\zeta_k,\alpha_k) - f_i(\hat{\zeta}_k,\hat{\alpha}_k) \\ &= \nabla f_i\big(\lambda_i(\zeta_k,\alpha_k) + (1-\lambda_i)(\hat{\zeta}_k,\hat{\alpha}_k)\big)^\top (\Delta\zeta_k,\Delta\alpha_k). \end{aligned}$$

By construction, $(\zeta_k,\alpha_k),(\hat{\zeta}_k,\hat{\alpha}_k) \in S_{\xi,\text{CH}} \times \Theta \times [c_\alpha, C_\alpha]$, and thus $\lambda_i(\zeta_k,\alpha_k) + (1-\lambda_i)(\hat{\zeta}_k,\hat{\alpha}_k) \in S_{\xi,\text{CH}} \times \Theta \times [c_\alpha, C_\alpha]$ (Note that $S_{\xi,\text{CH}}$ and $[c_\alpha, C_\alpha]$ are convex by construction, and $\Theta$ is convex under Assumption 3). Thus, we have:

$$\begin{aligned} &|r_{k,i}(\zeta_k,\alpha_k,\hat{\zeta}_k,\hat{\alpha}_k)| \\ &= |f_i(\zeta_k,\alpha_k) - f_i(\hat{\zeta}_k,\hat{\alpha}_k) - \nabla f_i(\hat{\zeta}_k,\hat{\alpha}_k)^\top (\Delta\zeta_k,\Delta\alpha_k)| \\ &\leq \|\nabla f_i\big(\lambda_i(\zeta_k,\alpha_k) + (1-\lambda_i)(\hat{\zeta}_k,\hat{\alpha}_k)\big) - \nabla f_i(\hat{\zeta}_k,\hat{\alpha}_k)\|_1 \\ &\qquad \cdot \|(\Delta\zeta_k,\Delta\alpha_k)\|_\infty \\ &\leq \mu_i \cdot \lambda_i \|(\Delta\zeta_k,\Delta\alpha_k)\|_\infty \cdot \|(\Delta\zeta_k,\Delta\alpha_k)\|_\infty, \end{aligned} \tag{27}$$

$$\leq \mu_i \cdot \|(\Delta\zeta_k,\Delta\alpha_k)\|_\infty^2,$$

where (27) follows from Prop. 4. ∎

We define $\boldsymbol{\mu} := \text{diag}\{\mu_1,\cdots,\mu_{n+d}\} \in \mathbb{R}^{(n+d)\times(n+d)}$, which allows us to stack (26) across all $i \in [n+d]$ as:

$$d_k = r_k(\zeta_k,\alpha_k,\hat{\zeta}_k,\hat{\alpha}_k) \in \|(\Delta\zeta_k,\Delta\alpha_k)\|_\infty^2 \boldsymbol{\mu}\mathbb{B}_\infty^{n+d} \tag{28}$$

Below, in Thm. 2, we leverage Prop. 5 to establish robust guarantees for satisfying the steplength constraint (10).

*Theorem 2 (**Robust Steplength Constraint Satisfaction**):* Suppose Assumptions 1, 2, 3, and 4 hold. Given a nominal trajectory $(\hat{\boldsymbol{\zeta}},\hat{\boldsymbol{\alpha}})$ of the dynamics $f$ in (8) and a system response $\Phi$ satisfying (23), if there exist auxiliary error upper bounds $\boldsymbol{\tau} := (\tau_0,\cdots,\tau_N)$ satisfying, $\forall k \in [0,N]$:

$$\begin{aligned} &a_p^\top(\hat{\zeta}_k,\hat{\alpha}_k) + b_p + \left\|a_p^\top \begin{bmatrix} (\Phi_{\zeta d})_{k,0} \\ (\Phi_{\alpha d})_{k,0} \end{bmatrix}\right\|_1 \text{diam}(\Theta) \\ &\quad + \sum_{j=1}^k \tau_{j-1}^2 \left\|a_p^\top \begin{bmatrix} (\Phi_{\zeta d})_{k,j} \\ (\Phi_{\alpha d})_{k,j} \end{bmatrix} \boldsymbol{\mu}\right\|_1 \leq 0, \quad \forall\ p \in [2], \end{aligned} \tag{29}$$

$$\left\|a_p^\top \begin{bmatrix} (\Phi_{\zeta d})_{k,0} \\ (\Phi_{\alpha d})_{k,0} \end{bmatrix}\right\|_\infty \text{diam}(\Theta) + \sum_{j=1}^k \tau_{j-1}^2 \left\|\begin{bmatrix} (\Phi_{\zeta d})_{k,j} \\ (\Phi_{\alpha d})_{k,j} \end{bmatrix} \boldsymbol{\mu}\right\|_\infty \leq \tau_k, \tag{30}$$

then $\|(\Delta\zeta_k,\Delta\alpha_k)\|_\infty \leq \tau_k$ and $\alpha_k \in [c_\alpha, C_\alpha]$ for each $k \in [0,N]$, regardless of the ground truth value of $\theta \in \Theta$.

*Proof:* We proceed by induction. At $k=0$, we prove $\|(\Delta\zeta_0,\Delta\alpha_0)\|_\infty \leq \tau_0$ using (20), as follows:

$$\begin{aligned} \left\|\begin{bmatrix} \Delta\zeta_0 \\ \Delta\alpha_0 \end{bmatrix}\right\|_\infty &= \left\|\begin{bmatrix} (\Phi_{\zeta d})_{0,0} \\ (\Phi_{\alpha d})_{0,0} \end{bmatrix} \Delta\zeta_0\right\|_\infty \leq \left\|\begin{bmatrix} (\Phi_{\zeta d})_{0,0} \\ (\Phi_{\alpha d})_{0,0} \end{bmatrix}\right\|_\infty \|\Delta\zeta_0\|_\infty \\ &\leq \tau_0, \end{aligned} \tag{31}$$

where (31) follows by taking $k=0$ in (30) and noting that $\|\Delta\zeta_0\|_\infty = \|(0,\Delta\theta_0)\|_\infty \leq \text{diam}(\Theta)$. Similarly, we verify robust constraint satisfaction at $k=0$ as follows:

$$\begin{aligned} &a_p^\top(\zeta_0,\alpha_0) + b_p \\ &= a_p^\top(\hat{\zeta}_0,\hat{\alpha}_0) + b_p + a_p^\top(\Delta\zeta_0,\Delta\alpha_0) \\ &= a_p^\top(\hat{\zeta}_0,\hat{\alpha}_0) + b_p + a_p^\top \begin{bmatrix} (\Phi_{\zeta d})_{0,0} \\ (\Phi_{\alpha d})_{0,0} \end{bmatrix} \Delta\zeta_0 \\ &\leq a_p^\top(\hat{\zeta}_0,\hat{\alpha}_0) + b_p + \left\|a_p^\top \begin{bmatrix} (\Phi_{\zeta d})_{0,0} \\ (\Phi_{\alpha d})_{0,0} \end{bmatrix}\right\|_1 \|\Delta\zeta_0\|_\infty \leq\ 0. \end{aligned}$$

From (29), $a_p^\top(\hat{\zeta}_0,\hat{\alpha}_0) + b_p \leq 0$, so $\alpha_0,\hat{\alpha}_0 \in [c_\alpha, C_\alpha]$.

Suppose for some $k \geq 1$, we have $\|(\Delta\zeta_j,\Delta\alpha_j)\|_\infty \leq \tau_j$ for each $j \in [0,k-1]$ and $\alpha_{0:k-1},\hat{\alpha}_{0:k-1} \in [c_\alpha, C_\alpha]^k$. First, we verify that the disturbance bounds given by Prop. 5 are valid for $d_j = r_j(\zeta_j,\alpha_j,\hat{\zeta}_j,\hat{\alpha}_j)$:

$$\zeta_j = (\xi_j,\theta_j) = (\xi_j,\theta_0) \in S_{\xi,j}(\xi_0,\alpha_{0:j-1}) \times \Theta, \tag{32a}$$

$$\hat{\zeta}_j = (\hat{\xi}_j,\hat{\theta}_j) = (\hat{\xi}_j,\hat{\theta}_0) \in S_{\xi,j}(\hat{\xi}_0,\hat{\alpha}_{0:j-1}) \times \Theta, \tag{32b}$$

where $\alpha_{0:k-1},\hat{\alpha}_{0:k-1} \in [c_\alpha, C_\alpha]^k$ by the induction hypothesis, and $\xi_0,\hat{\xi}_0 \in S_\xi^0$. Thus, the conditions of Prop. 5 are met at time $j$, so (26) and (28) imply:

$$d_j \in \|(\Delta\zeta_j,\Delta\alpha_j)\|_\infty^2 \boldsymbol{\mu}\mathbb{B}_\infty^{n+d} \subseteq \tau_j^2 \boldsymbol{\mu}\mathbb{B}_\infty^{n+d}. \tag{33}$$

Then, by (33):

$$\begin{aligned} \left\|\begin{bmatrix} \Delta\zeta_k \\ \Delta\alpha_k \end{bmatrix}\right\|_\infty &= \left\|\begin{bmatrix} (\Phi_{\zeta d})_{k,0} \\ (\Phi_{\alpha d})_{k,0} \end{bmatrix} \Delta\zeta_0 + \sum_{j=1}^k \begin{bmatrix} (\Phi_{\zeta d})_{k,j} \\ (\Phi_{\alpha d})_{k,j} \end{bmatrix} d_{j-1}\right\|_\infty \\ &\leq \left\|\begin{bmatrix} (\Phi_{\zeta d})_{k,0} \\ (\Phi_{\alpha d})_{k,0} \end{bmatrix}\right\|_\infty \|\Delta\zeta_0\|_\infty \end{aligned} \tag{34}$$

$$+\sum_{j=1}^{k}\tau_{j-1}^2\left\|\begin{bmatrix}(\Phi_{\zeta d})_{k,j}\\(\Phi_{\alpha d})_{k,j}\end{bmatrix}\boldsymbol{\mu}\right\|_\infty\le\tau_k.$$

We also prove robust constraint satisfaction, as follows:

$$\begin{aligned}
&a_p^\top(\zeta_k,\alpha_k)+b_p=\ a_p^\top(\hat\zeta_k,\hat\alpha_k)+b_p+a_p^\top(\Delta\zeta_k,\Delta\alpha_k)\\
&=a_p^\top(\hat\zeta_k,\hat\alpha_k)+b_p\\
&\qquad+a_p^\top\begin{bmatrix}(\Phi_{\zeta d})_{k,0}\\(\Phi_{\alpha d})_{k,0}\end{bmatrix}\Delta\zeta_0+\sum_{j=1}^{k}a_p^\top\begin{bmatrix}(\Phi_{\zeta d})_{k,j}\\(\Phi_{\alpha d})_{k,j}\end{bmatrix}d_{j-1}\\
&\le a_p^\top(\hat\zeta_k,\hat\alpha_k)+b_p+\left\|a_p^\top\begin{bmatrix}(\Phi_{\zeta d})_{k,0}\\(\Phi_{\alpha d})_{k,0}\end{bmatrix}\right\|_1\|\Delta\zeta_0\|_\infty\\
&\qquad+\sum_{j=1}^{k}\tau_{j-1}^2\left\|a_p^\top\begin{bmatrix}(\Phi_{\zeta d})_{k,j}\\(\Phi_{\alpha d})_{k,j}\end{bmatrix}\boldsymbol{\mu}\right\|_1\le 0,\qquad(35)
\end{aligned}$$

which also gives $a_p^\top(\hat\zeta_k,\hat\alpha_k)+b_p\le 0$, via (35). ■

We recover the tightest (in $\tau_N$) reachable set for (2) using the nonlinear program (NLP) below, whose constraints ensure robust satisfaction of (10) and hence valid outer approximations of $S_\xi^\star$ via Thm. 1. The convex regularizer $H(\Phi)$ minimizes uncertainty propagation [25]:

$$\min_{\hat{\boldsymbol{\zeta}},\hat{\boldsymbol{\alpha}},\Phi,\boldsymbol{\tau}}\quad \tau_N+H(\Phi)\qquad(36a)$$

$$\text{s.t.}\quad (11),(23),(29),(30).\qquad(36b)$$

We tractably compute a feasible point of (36) using the output feedback SLS solver [10], with $\mathcal{O}(N(n^3+d^3))$ runtime.

### C. Smoothing PGD Dynamics when Non-Differentiable

For non-affine $S_{\mathcal{C}}$, $\text{proj}_{S_{\mathcal{C}}}(\cdot)$ is continuous but in general non-differentiable, in which case $f_{\text{PGD}}$ and $f$ (see (2)) are also non-differentiable. To proceed similarly to Sec. IV-A–IV-B, we *smooth* $f_{\text{PGD}}$ and $f$. Given a *sampling radius* $\delta>0$, for any $\zeta=(\xi,\theta)\in\mathbb{R}^{n+d}$ and $\alpha\in\mathbb{R}$, let $\tilde f_{\text{PGD}}(\cdot;\delta):\mathbb{R}^{n+d+1}\to\mathbb{R}^n$ and $\tilde f(\cdot;\delta):\mathbb{R}^{n+d+1}\to\mathbb{R}^{n+d}$ by:

$$\tilde f_{\text{PGD}}(\zeta,\alpha;\delta):=\mathbb{E}_{q\sim\text{Unif}(\mathbb{B}_2^{n+d+1})}\big[f_{\text{PGD}}\big((\zeta,\alpha)+\delta q\big)\big],\quad(37)$$

$$\tilde f(\zeta,\alpha;\delta):=\big(\tilde f_{\text{PGD}}(\zeta,\alpha;\delta),\theta\big).\qquad(38)$$

We impose a fixed $C_\delta>0$ such that $\delta\in(0,C_\delta]$.

We review properties of smoothed functions [13]. Notably, $\delta$ affects the approximation error and curvature of $\tilde h$ inversely.

*Proposition 6 (**Properties of Smoothed Functions**):* Let $S_h\subseteq\mathbb{R}^m$ and $h:\mathbb{R}^m\to\mathbb{R}^{\overline{m}}$ $\ell_h$-Lipschitz continuous in $\|\cdot\|_\infty$ on $S_h\oplus C_\delta I_m\mathbb{B}_2^m$. Define $\tilde h(\cdot;\delta):\mathbb{R}^m\to\mathbb{R}^{\overline{m}}$ as:

$$\tilde h(\cdot;\delta):=\mathbb{E}_{q\sim\text{Unif}(\mathbb{B}_2^m)}\big[h(z+\delta q)\big],\quad\forall\ z\in\mathbb{R}^m.\qquad(39)$$

Then, for each $i\in[\overline{m}]$ and $z,z'\in S_h$:

1) $\|\tilde h(z;\delta)-h(z)\|_\infty\le\ell_h\delta$;
2) $\nabla\tilde h_i(z;\delta)=\frac{m}{\delta}\mathbb{E}_{q\sim\text{Unif}(\mathbb{B}_2^m)}\big[h_i(z+\delta q)q\big]$;
3) $\|\nabla\tilde h_i(z;\delta)-\nabla\tilde h_i(z';\delta)\|_1\le\frac{m^{3/2}}{\delta}\ell_h\cdot\|z-z'\|_\infty$.

*Proof:* Immediate by [13, Lemma 1], [28, Sec. 4.2]. ■

We apply Prop. 6 to the smoothed dynamics $\tilde f$ in (37).

*Proposition 7:* (**Smoothed PGD Dynamics are Lipschitz and admit Lipschitz Gradients**) There exists $\ell_{\tilde f},\tilde\mu_{i,\delta}$ such that, $\forall\ (\zeta,\alpha),(\zeta',\alpha')\in S_{\xi,\text{CH}}\times\Theta\times[c_\alpha,C_\alpha]$ and $i\in[n+d]$:

$$\|\tilde f(\zeta,\alpha;\delta)-f(\zeta,\alpha)\|_\infty\le\ell_{\tilde f}\delta,\qquad(40)$$

$$\|\nabla\tilde f_i(\zeta,\alpha;\delta)-\nabla\tilde f_i(\zeta',\alpha';\delta)\|_1\le\tilde\mu_{i,\delta}\|(\zeta,\alpha)-(\zeta',\alpha')\|_\infty.\qquad(41)$$

*Proof:* Under Assumption 1 and the boundedness of $S_{\xi,\text{CH}}\times\Theta\times[c_\alpha,C_\alpha]$, the map $f$ is $\ell_{\tilde f}$-Lipschitz continuous for some $\ell_{\tilde f}>0$. Thus, (40)-(41) follow directly from Prop. 6. ■

Equipped with Prop. 7, we adapt the methods in Sec. IV-A-IV-B to settings in which $f$ is non-differentiable. First, we use $\tilde f$, rather than $f$, to re-define the gain and actuation matrices of (13) for the LTV error system as:

$$\tilde A_k(\hat\zeta_k,\hat\alpha_k):=\tfrac{\partial\tilde f}{\partial\zeta}(\hat\zeta_k,\hat\alpha_k),\qquad(42a)$$

$$\tilde B_k(\hat\zeta_k,\hat\alpha_k):=\tfrac{\partial\tilde f}{\partial\alpha}(\hat\zeta_k,\hat\alpha_k).\qquad(42b)$$

Similarly, we modify the definition of the linearization error $\tilde r_k(\zeta_k,\alpha_k,\hat\zeta_k,\hat\alpha_k)\in\mathbb{R}^{n+d}$, as given by (14), into:

$$\begin{aligned}\tilde r_k(\zeta_k,\alpha_k,\hat\zeta_k,\hat\alpha_k)&=\tilde f(\zeta_k,\alpha_k)-\tilde f(\hat\zeta_k,\hat\alpha_k)\\&\quad-\tilde A_k(\hat\zeta_k,\hat\alpha_k)\Delta\zeta_k-\tilde B_k(\hat\zeta_k,\hat\alpha_k)\Delta\alpha_k.\end{aligned}\qquad(43)$$

We also modify the disturbance definition $d_k$ in (15) into $\tilde d_k$, defined below, to additionally account for the difference between the true dynamics $f$ and its smoothed proxy $\tilde f$:

$$\begin{aligned}\tilde d_k&=\tilde r_k(\zeta_k,\alpha_k,\hat\zeta_k,\hat\alpha_k)+\big[f(\zeta_k,\alpha_k)-\tilde f(\zeta_k,\alpha_k)\big]\\&\quad+\big[\tilde f(\hat\zeta_k,\hat\alpha_k)-f(\hat\zeta_k,\hat\alpha_k)\big].\end{aligned}\qquad(44)$$

With (42)-(44), the state error $\Delta\zeta_k:=\zeta_k-\hat\zeta_k$ satisfies an analogue to (16):

$$\Delta\zeta_{k+1}=\tilde A_k(\hat\zeta_k,\hat\alpha_k)\Delta\zeta_k+\tilde B_k(\hat\zeta_k,\hat\alpha_k)\Delta\alpha_k+\tilde d_k.$$

We now adapt Prop. 4, Prop. 5, and Thm. 2 to the setting in which $f$ is non-differentiable. For Prop. 4, we replace (25) with (41). For Prop. 5, we replace $\mu_i$ with $\tilde\mu_{i,\delta}$ in (26), i.e.:

$$|\tilde r_{k,i}(\zeta_k,\alpha_k,\hat\zeta_k,\hat\alpha_k)|\le\tilde\mu_{i,\delta}\cdot\|(\Delta\zeta_k,\Delta\alpha_k)\|_\infty^2,\qquad(45)$$

which can likewise be proved via (41) from Prop. 4.

Next, we augment the disturbance bounds (28) to incorporate the *smoothing error*, i.e., the difference between $f$ and $\tilde f$ as evaluated on the true and nominal trajectories:

$$\begin{aligned}\tilde d_k&=\tilde r_k(\zeta_k,\alpha_k,\hat\zeta_k,\hat\alpha_k)+\big[f(\zeta_k,\alpha_k)-\tilde f(\zeta_k,\alpha_k)\big]\\&\quad+\big[\tilde f(\hat\zeta_k,\hat\alpha_k)-f(\hat\zeta_k,\hat\alpha_k)\big]\\&\in\big(2\ell_{\tilde f}\delta I_{n+d}+\|(\Delta\zeta_k,\Delta\alpha_k)\|_\infty^2\tilde{\boldsymbol{\mu}}_\delta\big)\mathbb{B}_\infty^{n+d},\end{aligned}\qquad(46)$$

where the value $2\ell_{\tilde f}\delta$ in (46) arises by applying (40), and $\tilde{\boldsymbol{\mu}}_\delta:=\text{diag}\{\tilde\mu_{1,\delta},\cdots,\tilde\mu_{n+d,\delta}\}$.

We then modify Thm. 2 by respectively replacing (29) and (30) with (47) and (48), as given below for all $k\in[0,N]$:

$$\begin{aligned}&a_p^\top(\hat\zeta_k,\hat\alpha_k)+b_p+\left\|a_p^\top\begin{bmatrix}(\Phi_{\zeta d})_{k,0}\\(\Phi_{\alpha d})_{k,0}\end{bmatrix}\right\|_1\text{diam}(\Theta)\\&\quad+\sum_{j=1}^{k}\left\|a_p^\top\begin{bmatrix}(\Phi_{\zeta d})_{k,j}\\(\Phi_{\alpha d})_{k,j}\end{bmatrix}\begin{bmatrix}2\ell_{\tilde f}\delta I_{n+d}&\tau_{j-1}^2\tilde{\boldsymbol{\mu}}_\delta\end{bmatrix}\right\|_1\le 0,\forall p\in[2]\end{aligned}\qquad(47)$$

$$\begin{aligned}&\left\|a_p^\top\begin{bmatrix}(\Phi_{\zeta d})_{k,0}\\(\Phi_{\alpha d})_{k,0}\end{bmatrix}\right\|_\infty\text{diam}(\Theta)\\&+\sum_{j=1}^{k}\left\|\begin{bmatrix}(\Phi_{\zeta d})_{k,j}\\(\Phi_{\alpha d})_{k,j}\end{bmatrix}\begin{bmatrix}2\ell_{\tilde f}\delta I_{n+d}&\tau_{j-1}^2\tilde{\boldsymbol{\mu}}_\delta\end{bmatrix}\right\|_\infty\le\tau_k,\ \forall k\in[0,N].\end{aligned}\qquad(48)$$

The validity of (47) and (48) as replacements for (29) and (30) in Thm. 2, for non-differentiable dynamics $f$, can be verified by retracing the proof of Thm. 2, and applying the modified disturbance bound (46) instead of its original form (28) when appropriate. A similar replacement is made in the NLP (36), used to recover the tightest final tube size $\tau_N$.

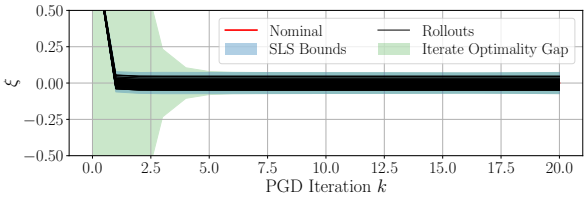



Fig. 1. Robust over-approximation of the parameterized solution set of (49). Individual Monte-Carlo sampled $\theta \in \Theta_{\text{sq}}$ PGD rollouts are in black; the "nominal" $\theta = 0$ is in red. The blue region contains PGD iterates for any $\theta \in \Theta_{\text{sq}}$ (Thm. 2), and the green bounds contain $S_\xi^\star$ at any time (Thm. 1).

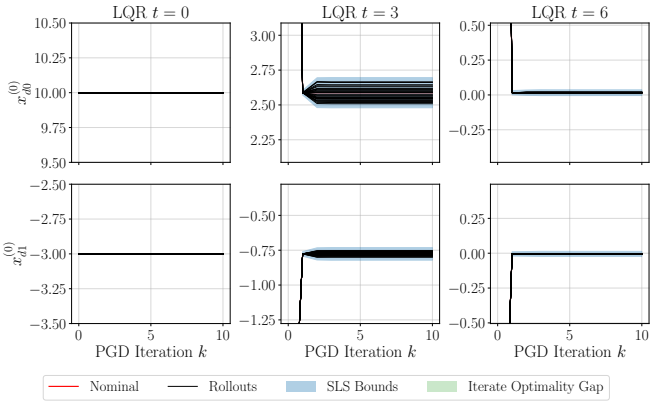



Fig. 2. Robust over-approximation of the planned positions of a parameterized LQR system. All visual elements are as in Fig. 1. For space, we only show the components of $\xi$ corresponding to position at timesteps $t = 0, 3, 6$.

## V. Simulations

Here, we apply Thms. 1-2 to compute over-approximations of the parameterized minimizer sets for specific objectives.

### A. Quadratic Optimization

Consider the unconstrained minimization of a scalar cost:

$$\min_{\xi \in \mathbb{R}} J_{\text{sq}}(\xi, \theta) := \xi^2 + \theta\xi \tag{49}$$

over $\xi \in \mathbb{R}$, with parameter uncertainty over $\theta \in \Theta_{\text{sq}} := [-0.1, 0.1]$. $J_{\text{sq}}$ is $m$-strongly convex with $L$-Lipschitz gradients, where $m = L = 2$. The PGD dynamics are:

$$f_{\text{PGD,sq}}(\xi, \theta, \alpha) := \xi - \alpha(2\xi + \theta). \tag{50}$$

We compute $\nabla^2 f_i$ for $i \in \{1, 2\}$ to find that the curvature of $f$ can be bounded by $\mu_1 = 2$, $\mu_2 = 0$, as in Prop. 4.

Choosing $\xi_0 = 1.0$, $c_\alpha = 0.4$, $C_\alpha = 0.6$ and applying Thm. 2, we bound the $\infty$-norm of the PGD *iterates* over horizon $N = 20$. By (4) and Prop. 2, these iterates converge exponentially to $S_\xi^\star$ with rate $\gamma = 0.2$. SLS ensures that the iterates are bounded between $\overline{\xi} = 10$ and $\underline{\xi} = -\overline{\xi}$, so $\|\xi_0 - \xi^\star(\theta)\| \leq \frac{\|\overline{\xi} - \underline{\xi}\|}{1-\gamma^N}$ for all $\theta \in \Theta$, overapproximating $\beta_k(\xi_0)$ in Thm. 1. These bounds are combined in Fig. 1 to overapproximate the parameterized minimizer set of (49).

The minimizer set for (49) is exactly $S_\xi^\star = \{-\frac{\theta}{2} \mid \theta \in \Theta\}$. While our SLS-computed bounds are conservative, our method provides tighter bounds than baseline methods when no closed-form expression for $S_\xi^\star$ is available (see Sec. V-B).

### B. LQR with Intent Uncertainty

Consider the following parameterized LQR problem, with variable $\xi := (x_0, u_0, \ldots, x_{T-1}, u_{T-1}, x_T) \in \mathbb{R}^{4(T+1)+2T}$, objective $J_{\text{LQR}}(\xi, \theta)$, and parameter $\theta \in \mathbb{R}$:

$$\min_{\xi} \quad \frac{1}{2} x_T^\top Q_T x_T + \tfrac{1}{2} \textstyle\sum_{t=0}^{T-1} \left(x_t^\top Q_t x_t + \theta u_t^\top R_t u_t\right), \tag{51a}$$

$$x_{t+1} = A_t x_t + B_t u_t, \qquad x_0 = \bar{x}_0, \tag{51b}$$

where:

$$A_t = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, B_t = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, Q_t = 0.1I, R_t = 0.1I. \tag{52}$$

$A_t$ and $B_t$ describe a time-invariant double integrator. The combined state $\xi$ for $T = 10$ has dimension 64. Above, $\theta \in \Theta = \Theta_{\text{LQR}} := [0.9, 1.1]$ scales the relative weight of state and control cost. We compactly write (51a) as $J_{\text{LQR}}(\xi, \theta) = \frac{1}{2}\xi^\top H \xi = \frac{1}{2}\xi^\top (H_Q + \theta H_R)\xi$ and dynamics (51b) as $M\xi = b$ for appropriate $H$, $M$, and $b$. Concretely,

$$H_Q = \text{blkdiag}(Q_0, O, \ldots, Q_{T-1}, O, Q_T),$$
$$H_R = \text{blkdiag}(O, R_0, \ldots, O, R_{T-1}, O),$$

The constraint manifold is the affine space given by $\{\xi = Pz + d : z \in \mathbb{R}^n\}$, with $P := I - M^\top (MM^\top)^{-1} M$ and $d = M^\top (MM^\top)^{-1} b$, yielding PGD dynamics

$$f_{\text{PGD,LQR}}(\xi, \theta, \alpha) = P(I - \alpha H)\xi + d. \tag{53}$$

$J_{\text{LQR}}$ is $\min_{\theta \in \Theta_{\text{LQR}}} \lambda_{\min}(H)$=0.09-strongly convex and has Lipschitz gradients with constant $\max_{\theta \in \Theta_{\text{LQR}}} \lambda_{\max}(H) = 0.11$. Due to high-dimensionality, the computation of the curvature bounds $\boldsymbol{\mu}$ is more challenging than for Sec. V-A. Fixing $\overline{\xi} = (20, 20, \ldots, 20)$, $\underline{\xi} = -\overline{\xi}$, $c_\alpha = 9.9$, $C_\alpha = 10.1$ (guaranteed by SLS to contain the true iterates), we compute a local over-approximation in the interval arithmetic toolbox `immrax` [23]. Then, as in Sec. V-A, we bound the iterates $\xi_k$ minimizing (51a) across $N = 10$ iterations, and bloat these bounds using the convergence rate $\gamma = 0.111$. Fig. 2 shows the anytime over-approximations of $S_\xi^\star$ for the position states at timesteps 0, 3, and 6. Our bounds are only slightly larger than those indicated by Monte Carlo-sampled rollouts.

*1) Baseline Comparison:* Compared to our method, classical Implicit Function Theorem methods [8, Section 11.5.2] perform poorly. For dual variables $\lambda$, differentiating the KKT conditions of (51) at an optimal $(\xi^\star, \lambda^\star)$ yields:

$$\begin{bmatrix} H & M^\top \\ M & O \end{bmatrix} \begin{bmatrix} \nabla_\theta \xi^\star(\theta) \\ \nabla_\theta \lambda^\star(\theta) \end{bmatrix} = \begin{bmatrix} -H_R \xi \\ \mathbf{0} \end{bmatrix},$$

with solution $\nabla_\theta \xi^\star = [H^{-1} M^\top (MH^{-1} M^\top)^{-1} MH^{-1} - H^{-1}] H_R \xi := \mathcal{P} H_R \xi$. Over the parameter region $[\underline{\xi}, \overline{\xi}]$ guaranteed to be robustly forward invariant for PGD by SLS, $L_{\xi^\star} := \sup_{\theta \in \Theta_{\text{LQR}}} \|\nabla_\theta \xi^\star\|_2$ is a local Lipschitz constant for $\xi^\star(\theta)$. By sub-multiplicativity of norms and standard linear algebraic techniques, for any $\theta_1, \theta_2 \in \Theta_{\text{LQR}}$

$$\begin{aligned} \|\xi^\star(\theta_1) - \xi^\star(\theta_2)\|_2 &\leq L_\xi^\star \|\theta_1 - \theta_2\|_2 \\ &\leq \|\mathcal{P}\|_2 \|H_R\|_2 \text{diam}(\Theta_{\text{LQR}}) \sup_{\xi \in [\underline{\xi}, \overline{\xi}]} \|\xi\|_2 \\ &\leq \|H^{-1}\|_2 \|H_R\|_2 \text{diam}(\Theta_{\text{LQR}}) \|\overline{\xi}\|_2 \approx 35. \end{aligned}$$

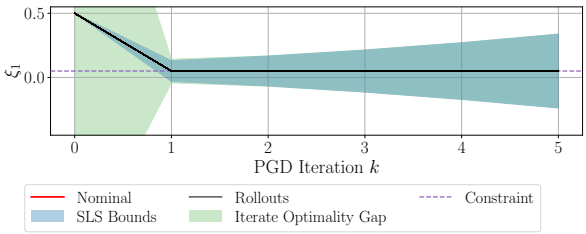



Fig. 3. Robust over-approximation of $S_\xi^\star$ for (54). All visual elements are as in Fig. 1. The boundary of the constraint set $S_\mathcal{C}$ is also shown in purple.

Our method computes the much less conservative bound of 0.2 for each component of $\xi$.

### C. Constrained Quadratic Optimization

Consider the following constrained variant of (49):

$$\min_{\xi\in\mathbb{R}} \quad J_{\text{sqc}}(\xi,\theta) := c_1\xi^2 + c_2\theta\xi, \tag{54a}$$

$$\text{s.t.} \quad \xi \geq c_3. \tag{54b}$$

Since the constraint set above is *not* affine, the PGD dynamics are non-smooth, so we apply the smoothing process in Sec. IV-C to obtain reachable set bounds.

The objective satisfies Assumption 1 with $m = L = 2c_1$. The gradient descent dynamics without projection are similar to (50), and have Lipschitz constant $\ell_{\text{sqc}} = \sup_{(\xi,\theta,\alpha)\in\mathcal{D}}|1 - 2c_1\alpha| + |c_2\alpha| + |2c_1\xi + c_2\theta|$ as in Prop. 6. Since (54) is scalar, projection is non-expansive, and $\ell_{\text{sqc}}$ is a Lipschitz constant for PGD. Applying Prop. 6 with $\delta = 0.1$ bounds the curvature of the smoothed dynamics and their distance from the true PGD update. Then, (46)–(48) over-approximate PGD iterates for any parameterized realization of (54).

Using $\xi_0 = 0.5$, $\Theta = [-0.1, 0.1]$, $c_1 = 0.0521$, $c_2 = 0.0054$, $c_\alpha = 9.54$, and $C_\alpha = 9.56$, we outer-approximate $S_\xi^\star$ for (54) (Fig. 3). Each PGD iteration incurs conservatism due to the smoothing approximation necessary for SLS. Rollouts quickly approach the constraint boundary, requiring Prop. 6 to bound the curvature of the smoothed dynamics, and increasing conservativeness compared to previous examples.

## VI. Conclusion

We present a novel SLS-based method for over-approximating the minimizer set of strongly convex costs with uncertain parameters, by recasting the problem as the forward reachability analysis of PGD dynamics. In numerical experiments, our method outperforms sensitivity analysis baselines in over-approximating the minimizer set with low conservativeness. Future work will quantify the benefits of optimizing step-size policies, investigate alternative smoothing techniques, and fuse minimizer set over-approximation with adaptive techniques that shrink $\Theta$ online.

## References


[1] P. Englert, N. A. Vien, and M. Toussaint, "Inverse KKT: Learning Cost Functions of Manipulation Tasks from Demonstrations," *The International Journal of Robotics Research*, vol. 36, no. 13-14, 2017.

[2] L. Peters, V. Rubies-Royo, C. J. Tomlin, L. Ferranti, J. Alonso-Mora, C. Stachniss, and D. Fridovich-Keil, "Online and Offline Learning of Player Objectives from Partial Observations in Dynamic Games," *The International Journal of Robotics Research*, vol. 42, no. 10, 2023.

[3] M. Menner, P. Worsnop, and M. N. Zeilinger, "Constrained Inverse Optimal Control With Application to a Human Manipulation Task," *IEEE Transactions on Control Systems Technology*, vol. 29, 2021.

[4] W.-C. Li and G. Chou, "Certified Gradient-Based Contact-Rich Manipulation via Smoothing-Error Reachable Tubes," *arXiv preprint arXiv:2602.09368*, 2026.

[5] ——, "A Convex Formulation of Compliant Contact Between Filaments and Rigid Bodies," *arXiv preprint arXiv:2509.13434*, 2025.

[6] M. Althoff, O. Stursberg, and M. Buss, "Reachability Analysis of Nonlinear Systems with Uncertain Parameters using Conservative Linearization," in *IEEE Conference on Decision and Control*, 2008.

[7] L. Schäfer and M. Althoff, "Robust Optimal Control Using Set-based Reachability Analysis," in *European Control Conference*. IEEE, 2025.

[8] S. J. Wright and B. Recht, *Optimization for Data Analysis*. Cambridge University Press, 2022.

[9] A. P. Leeman, J. Köhler, A. Zanelli, S. Bennani, and M. N. Zeilinger, "Robust Nonlinear Optimal Control via System Level Synthesis," *IEEE Transactions on Automatic Control*, vol. 70, no. 7, 2025.

[10] A. P. Leeman, S. Zhan, M. N. Zeilinger, and G. Chou, "Vision-sls: Safe perception-based control from learned visual representations via system level synthesis," *arXiv preprint arXiv:2604.24894*, 2026.

[11] L. Furieri, Y. Zheng, A. Papachristodoulou, and M. Kamgarpour, "An Input-Output Parametrization of Stabilizing Controllers: amidst Youla and System Level Synthesis," *IEEE Control Systems Letters*, vol. 3, no. 4, pp. 1014–1019, 2019.

[12] J. Fang and G. Chou, "Safe Large-Scale Robust Nonlinear MPC in Milliseconds via Reachability-Constrained System Level Synthesis on the GPU," *arXiv preprint arXiv:2604.07644*, 2026.

[13] A. D. Flaxman, A. T. Kalai, and H. B. McMahan, "Online Convex Optimization in the Bandit Setting: Gradient Descent without Gradient," in *ACM-SIAM Symposium on Discrete Algorithms*, USA, 2005.

[14] D. Bertsimas and M. Sim, "The Price of Robustness," *Operations Research*, vol. 52, no. 1, pp. 35–53, 2004.

[15] A. Ben-Tal and A. Nemirovski, "Robust Convex Optimization," *Mathematics of Operations Research*, vol. 23, no. 4, pp. 769–805, 1998.

[16] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and Applications of Robust Optimization," *SIAM Review*, vol. 53, no. 3, 2011.

[17] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*. Springer Science & Business Media, 2013.

[18] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings*. Springer, 2009, vol. 543.

[19] A. Shapiro, "Sensitivity Analysis of Parameterized Variational Inequalities," *Mathematics of Operations Research*, vol. 30, no. 1, 2005.

[20] C. N. Jones, E. C. Kerrigan, and J. M. Maciejowski, "Lexicographic Perturbation for Multiparametric Linear Programming with Applications to Control," *Automatica*, vol. 43, no. 10, pp. 1808–1816, 2007.

[21] P. Tøndel, T. A. Johansen, and A. Bemporad, "An Algorithm for Multi-Parametric Quadratic Programming and Explicit MPC Solutions," *Automatica*, vol. 39, no. 3, pp. 489–497, 2003.

[22] J. M. Bravo, D. Limón, T. Alamo, and E. F. Camacho, "On the Computation of Invariant Sets for Constrained Nonlinear Systems: An Interval Arithmetic Approach," *Automatica*, vol. 41, no. 9, 2005.

[23] A. Harapanahalli, S. Jafarpour, and S. Coogan, "immrax: A Parallelizable and Differentiable Toolbox for Interval Analysis and Mixed Monotone Reachability in JAX," *IFAC-PapersOnLine*, vol. 58, 2024.

[24] S. Bansal, M. Chen, S. Herbert, and C. J. Tomlin, "Hamilton-Jacobi Reachability: A Brief Overview and Recent Advances," in *CDC*, 2017.

[25] A. P. Leeman, J. Kohler, F. Messerer, A. Lahr, M. Diehl, and M. N. Zeilinger, "Fast System Level Synthesis: Robust Model Predictive Control using Riccati Recursions," *IFAC Conference on Nonlinear Model Predictive Control*, vol. 58, no. 18, pp. 173–180, 2024.

[26] J. Anderson, J. C. Doyle, S. H. Low, and N. Matni, "System Level Synthesis," *Annual Reviews in Control*, vol. 47, 2019.

[27] H. Zhou and V. Tzoumas, "Safe Control of Partially-Observed Linear Time-Varying Systems with Minimal Worst-Case Dynamic Regret," in *IEEE Conference on Decision and Control*, 2023.

[28] M. Bravo, D. Leslie, and P. Mertikopoulos, "Bandit Learning in Concave N-Person Games," in *NeurIPS*, 2018.


## Acknowledgement


The authors thank Shuyu Zhan for his assistance with the simulation experiments.