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

Exact Characterization of the Convex Hulls of Reachable Sets

Thomas Lew[1], Riccardo Bonalli[2], Marco Pavone[1]

***Abstract*— We study the convex hulls of reachable sets of nonlinear systems with bounded disturbances. Reachable sets play a critical role in control, but remain notoriously challenging to compute, and existing over-approximation tools tend to be conservative or computationally expensive. In this work, we exactly characterize the convex hulls of reachable sets as the convex hulls of solutions of an ordinary differential equation from all possible initial values of the disturbances. This finite-dimensional characterization unlocks a fast sampling-based method to accurately over-approximate reachable sets. We give applications to neural feedback loop analysis and robust model predictive control.**

I. Introduction

Forward reachability analysis plays a critical role in control theory and robust controller design. Generally, it entails characterizing all states that a dynamical system can reach at any time in the future. As such, reachability analysis allows certifying the performance of feedback loops under disturbances and designing controllers with robustness properties. In robust model predictive control (MPC) for instance, it is used to construct tubes around nominal state trajectories to ensure that constraints are satisfied in the presence of external disturbances.

In this work, we study the following forward reachability analysis problem. Let $n \in \mathbb{N}$ be the state dimension, $f : \mathbb{R}^n \to \mathbb{R}^n$ be a function defining the dynamics, and $\mathcal{W} \subset \mathbb{R}^n$ be a compact set of disturbances. Given a time $T > 0$ and an initial state $x^0 \in \mathbb{R}^n$, we consider systems characterized by the ordinary differential equation (ODE)

$$\dot{x}(t) = f(x(t)) + w(t), \quad t \in [0, T], \quad x(0) = x^0, \quad (1)$$

where the disturbances $w : [0, T] \to \mathcal{W}$ are assumed to be integrable ($w \in L^\infty([0, T], \mathcal{W})$). Under standard smoothness assumptions (see Assumptions 1-2), the ODE (1) has a unique solution, denoted by $x_w(\cdot)$. For any time $t \in [0, T]$, we define the reachable set

$$\mathcal{X}_t = \left\{ \begin{array}{r} x_w(t) = x^0 + \int_0^t \big(f(x_w(s)) + w(s)\big)\mathrm{d}s \\ : w \in L^\infty([0, T], \mathcal{W}) \end{array} \right\}, \quad (2)$$

which characterizes all states that are reachable at time t for some disturbance $w \in L^\infty([0, T], \mathcal{W})$.

[1]Department of Aeronautics & Astronautics, Stanford University. {thomas.lew,pavone}@stanford.edu

[2]Laboratory of Signals & Systems, Paris-Saclay University, CNRS, CentraleSupélec riccardo.bonalli@l2s.centralesupelec.fr

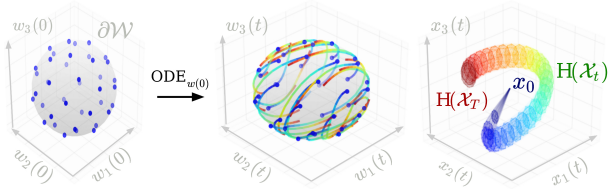

Fig. 1: Theorem 1 states that the convex hulls $\mathrm{H}(\mathcal{X}_t)$ of the reachable sets $\mathcal{X}_t$ can be computed by (a) integrating an augmented ODE (see $\mathbf{ODE}_{w(0)}$) for different initial values of the disturbances $w(0) \in \partial\mathcal{W}$, and (b) taking the convex hulls of the resulting extremal state trajectories $x_{w(0)}$.

Reachability analysis of nonlinear dynamical systems is challenging. Indeed, from (2), computing the reachable sets seemingly requires evaluating an infinite number of state trajectories for all possible disturbances.[1] Due to the complexity of the problem, many existing tools seek convex over-approximations of the reachable sets. However, current methods tend to be conservative or computationally expensive, see Sections II and V.

In this work, we provide a new characterization of the convex hulls of the reachable sets of dynamical systems of the form (1), under smoothness assumptions of f and $\mathcal{W}$ (see Assumptions 1 and 2). Specifically, denoting by $\mathrm{H}(A)$ the convex hull of a set $A \subset \mathbb{R}^n$, we show that

$$\mathrm{H}(\mathcal{X}_t) = \mathrm{H}(F(\partial\mathcal{W}, t)) \quad \text{for all} \quad t \in [0, T], \quad (3)$$

where $F(w(0), t)$ is the solution of an ODE with initial conditions $w(0)$ on the boundary of $\mathcal{W}$, see Theorem 1 and $\mathbf{ODE}_{w(0)}$. Thus, the convex hulls of the reachable sets can now be computed as the convex hull of forward propagations of an ODE for different initial disturbance values. Equation (3) represents a significantly simpler (finite-dimensional) characterization of the convex hulls.

This result suggests a simple approach (Algorithm 1) to estimate the convex hulls $\mathrm{H}(\mathcal{X}_t)$ by integrating an ODE from a sample of initial conditions. This method allows efficiently tackling notoriously challenging problems, such as with neural network controllers in the loop (see Section V). This characterization also informs the design of a robust MPC approach (Algorithm 3) that we demonstrate on a robust spacecraft attitude control task.

Outline: In Section II, we review prior work. In Section III, we state our characterization result of the reachable

[1]Each reachable set $\mathcal{X}_t$ is the image of an infinite-dimensional set. Indeed, by defining the maps $g_t : w \in L^\infty([0, T], \mathcal{W}) \mapsto x_w(t) \in \mathbb{R}^n$, each reachable set is expressed as $\mathcal{X}_t = g_t(L^\infty([0, T], \mathcal{W}))$.

convex hulls (Theorem 1) and propose an estimation algorithm (Algorithm 1). In Section IV, these results are derived using convex geometry and optimal control. We provide results in Section V and conclude in Section VI.

II. Related work

The forward reachable sets of nonlinear systems are generally difficult to characterize. For this reason, many existing approaches seek *convex* over-approximations of the reachable sets, e.g., represented as hyper-rectangles [1], ellipsoidal sets [2], or zonotopes [3], see [4] and [5] for recent surveys. Existing over-approximation methods include techniques based on conservative linearization [3], differential inequalities [6], [7], and Taylor models [8], [9]. In particular, systems with mixed-monotone [10]–[12] or contracting [13] dynamics have been extensively studied, as these properties simplify the fast computation of accurate over-approximations. To tackle smooth systems, a standard and widely applicable approach consists of linearizing the dynamics and bounding the Taylor remainder using smoothness properties of the dynamics [2], [3], [5], [14], [15]. This method has been widely used in robust MPC but is known to be conservative [16], see also results in Section V.

Methods that estimate the reachable sets from a sample of state trajectories have recently found significant interest [16]–[18]. However, the sample complexity of these methods increases with the number of uncertain variables. For systems with disturbances of the form of (1), the number of uncertain variables (and thus the approximation error) increases as the discretization is refined. Thus, naive sampling-based methods are not well-suited for reachability of systems with continuous-time disturbances, see also Section V for comparisons.

The results derived in this work leverage tools from convex geometry and optimal control. This deep connection [19] between geometry, reachability analysis, and optimal control was previously exploited in [20], [21] to characterize the true reachable sets of dynamical systems of dimensions $n \leq 4$ with scalar control inputs (control inputs in [20] take the role of disturbances in (1)). Our derivations also leverage the Pontryagin Maximum Principle (PMP), but we consider a different class of dynamical systems with arbitrary state dimensionality n and the same number of disturbances and states.

Our derivations start with the idea of searching for boundary states that are the furthest in different directions (see $\mathbf{OCP}_d$). This approach is standard in the setting with linear dynamics where the reachable sets are convex [21], [22]. However, in the nonlinear setting, the reachable sets may be non-convex, and finding the extremal disturbance trajectories that generate boundary states requires solving boundary-value problems (BVPs, see $\mathbf{BVP}_d$). This approach was explored in the three-dimensional case [23], but solving BVPs remains computationally challenging. Although sophisticated techniques exist to efficiently solve BVPs [24], [25], additional non-trivial analysis is still required for efficient implementation. Our results show that under the right set of assumptions (see Assumptions 1-2), solving BVPs is not necessary and extremal trajectories take a simple form. The key is the additional idea of sampling initial values of the disturbances. Studying the convex hull of the reachable sets unlocks arguments from convex geometry that allow proving the exactness of the approach.

Notations: We denote by $a^\top b$ the Euclidean inner product of $a, b \in \mathbb{R}^n$, by $\|a\|$ the Euclidean norm of $a \in \mathbb{R}^n$, by $I_{n\times n}$ the identity matrix of size n, by $B(x,r) = \{y \in \mathbb{R}^n : \|y-x\| \leq r\}$ the closed ball of center $x \in \mathbb{R}^n$ and radius $r \geq 0$, and by $\mathcal{S}^{n-1}$ the unit sphere in $\mathbb{R}^n$. For any $v \in \mathcal{S}^{n-1}$, we define the map $\mathsf{Proj}_v : \mathbb{R}^n \to \mathbb{R}^n$,

$$\mathsf{Proj}_v(u) = \left(I_{n\times n} - vv^\top\right) u = u - (v^\top u)v, \quad (4)$$

which denotes the orthogonal projection onto the tangent space of $\mathcal{S}^{n-1}$. Given $A, B \subset \mathbb{R}^n$, we denote by $\mathrm{Int}(A)$, $\overline{A}$, $\partial A = \overline{A} \setminus \mathrm{Int}(A)$, and $A^{\mathsf{c}} = \mathbb{R}^n \setminus A$ the interior, closure, boundary, and complement of A, by $d_A(x) = \inf_{a\in A} \|x - a\|$ the distance from $x \in \mathbb{R}^n$ to A, and by $d_{\mathrm{H}}(A, B) = \max(\sup_{x\in A} d_B(x), \sup_{y\in B} d_A(y))$ the Hausdorff distance between compact sets $A, B \subset \mathbb{R}^n$.

Convex geometry. We denote by $\mathrm{H}(A)$ the convex hull of $A \subset \mathbb{R}^n$ and by $\mathrm{Ext}(A)$ the set of extreme points of a compact set $A \subset \mathbb{R}^n$ [26]. The next result is standard.

Lemma 1 (Support hyperplane): Let $C \subset \mathbb{R}^n$ be convex and closed and $x \in \partial C$. Then, there exists a support hyperplane $\{y \in \mathbb{R}^n : d^\top(y-x) = 0\}$ parameterized by some $d \in \mathcal{S}^{n-1}$ such that $d^\top x \geq d^\top y$ for all $y \in C$.

III. The structure of $\mathrm{H}(\mathcal{X}_t)$

Our results rely on the two following assumptions.

Assumption 1: f is continuously differentiable ($f \in C^1$), globally Lipschitz, and its Jacobian ∇f is Lipschitz.

Assumption 2: $\mathcal{W}$ is compact and its boundary $\partial\mathcal{W}$ is an ovaloid. Equivalently, $\partial\mathcal{W}$ is a smooth $(n-1)$-dimensional submanifold of strictly positive curvature.

Assumption 1 is a standard smoothness assumption [27], [28] that guarantees the existence and uniqueness of solutions to the ODEs in (1) and (7). By multiplying f with a smooth cutoff function whose arbitrarily large support contains states of interest, the Lipschitzianity assumptions are always satisfied if $f \in C^2$. Assumption 2 holds if $\mathcal{W}$ is a sphere or an ellipsoid, and implies that $\mathcal{W}$ is strictly convex [29]. Relaxing this assumption (e.g., to accommodate hyper-rectangular disturbance sets) is left for future work. Assuming that $\mathcal{W}$ is convex is standard to prove that the reachable sets are compact.

Lemma 2 ($\mathcal{X}_t$ is compact): Assume that f and $\mathcal{W}$ satisfy Assumptions 1 and 2. Then, for any $t \in [0, T]$, the reachable set $\mathcal{X}_t$ defined in (2) is compact.

Lemma 2 is standard, see e.g. [30] (from Grönwall's inequality, state trajectories are uniformly bounded thanks to Assumptions 1 and 2, so Lemma 2 follows from [30, Theorem 5.2.1] with minor adaptations).

Thanks to Assumption 2, the *Gauss map* (see Figure 2)

$$n^{\partial\mathcal{W}} : \partial\mathcal{W} \to \mathcal{S}^{n-1} \quad (5)$$

is a diffeomorphism, see [29]. Given $w(t) \in \partial\mathcal{W}$, $n^{\partial\mathcal{W}}(w(t))$ is the unit-norm outward-pointing normal of $\partial\mathcal{W}$ at $w(t)$, such that for all $v \in \mathcal{W}$,

$$n^{\partial\mathcal{W}}(w(t))^\top (w(t) - v) \geq 0. \quad (6)$$

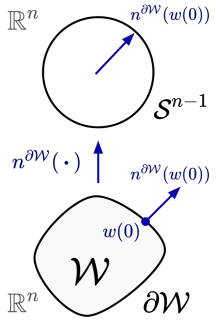

Fig. 2: Gauss map (5).

If $\mathcal{W}$ is a ball of radius r, then the Gauss map is simply given by $n^{\partial\mathcal{W}}(w) = \frac{w}{\|w\|}$.

Given any $w(0) \in \partial\mathcal{W}$, we define the augmented ODE

$$\textbf{ODE}_{w(0)}: \begin{cases} \dot{x}(t) = f(x(t)) + (n^{\partial\mathcal{W}})^{-1}(q(t)), \\ \dot{q}(t) = -\mathsf{Proj}_{q(t)}\left(\nabla f(x(t))^\top q(t)\right), \\ (x(0), q(0)) = (x^0, n^{\partial\mathcal{W}}(w(0))). \end{cases} t \in [0,T], \quad (7)$$

which has a unique solution $(x,q)_{w(0)} \in C([0,T], \mathbb{R}^n \times \mathcal{S}^{n-1})$ thanks to Assumptions 1-2, from standard results on solutions of ODEs [31], [32]. The next result characterizes the convex hull of the reachable sets as the convex hull of solutions to $\textbf{ODE}_{w(0)}$ for all $w(0) \in \partial\mathcal{W}$.

Theorem 1 (Convex hulls of reachable sets $H(\mathcal{X}_t)$): Assume that f and $\mathcal{W}$ satisfy Assumptions 1 and 2. Given $w(0) \in \partial\mathcal{W}$, denote by $(x_{w(0)}, q_{w(0)})$ the unique solution of $\textbf{ODE}_{w(0)}$. Define the map

$$F: \ \partial\mathcal{W} \times [0,T] \to \mathbb{R}^n : (w(0), t) \mapsto x_{w(0)}(t). \quad (8)$$

Then, $\mathrm{H}(\mathcal{X}_t) = \mathrm{H}(F(\partial\mathcal{W}, t))$ for all $t \in [0,T]$ (i.e., (3)).

Theorem 1 states that integrating $\textbf{ODE}_{w(0)}$ for all values of $w(0) \in \partial\mathcal{W}$ (i.e., evaluating $F(w(0),t)$ for different values of $w(0)$) is sufficient to recover the convex hulls of the reachable sets $\mathrm{H}(\mathcal{X}_t)$. This characterization significantly simplifies the reachability analysis problem, which is now finite-dimensional and amounts to integrating an ODE from different initial conditions.

Corollary 1 (Reachable tube): Assume that f and $\mathcal{W}$ satisfy Assumptions 1 and 2 and define the map F by (8) as in Theorem 1. Then, for all $t \in [0,T]$,

$$\mathrm{H}(\mathcal{X}_t) = \mathrm{H}\left(\bigcup_{w(0) \in \partial\mathcal{W}} F(w(0), [0,T])_t \right), \quad (9)$$

where $F(w(0),[0,T])_t = F(w(0),t)$.

Corollary 1 states that to recover the reachable convex hull $\mathrm{H}(\mathcal{X}_t)$ at any time t, it suffices to integrate $\textbf{ODE}_{w(0)}$ over $[0,T]$ only once for each initial disturbance value $w(0) \in \partial\mathcal{W}$. This result implies that all the information required to compute the entire convex reachable tube $\bigcup_{t\in[0,T]} \mathrm{H}(\mathcal{X}_t)$ (e.g., to enforce constraints at all times for robust MPC) is available after evaluating $\mathrm{H}(\mathcal{X}_T)$.

Alg. 1. Estimation of the reachable set convex hulls

Input: M initial disturbances $\{w^i(0)\}_{i=1}^M \subset \partial\mathcal{W}$.
Output: Approximation of the convex hulls $\mathrm{H}(\mathcal{X}_t)$.

1: **for all** $i = 1, \dots, M$ **do**
2: $\quad x^i \leftarrow \text{Integrate}(\textbf{ODE}_{w^i(0)})$
3: **return** $\mathrm{H}\left(\{x^i(t), i = 1, \dots, M\}\right), t \in [0,T]$.

Theorem 1 and Corollary 1 justify using Algorithm 1 to reconstruct the convex hulls $\mathrm{H}(\mathcal{X}_t)$. Error bounds for the approximation follow directly from Theorem 1 using a covering argument, see, e.g., [33, Lemma 4.2].

Corollary 2 (Error bound): Assume that $(f, \mathcal{W})$ satisfy Assumptions 1 and 2. Let $\delta > 0$, $Z_\delta \subset \partial\mathcal{W}$ be a δ-cover of $\partial\mathcal{W}$, and F be defined as in (8). Then,

$$d_{\mathrm{H}}(\mathrm{H}(\mathcal{X}_t), \mathrm{H}(F(Z_\delta, t))) \leq \bar{L}_t \delta \text{ for all } t \in [0,T], \quad (10)$$

where $\bar{L}_t$ denotes the Lipschitz constant of $F(\cdot, t)$.

Corollary 2 implies that padding the set estimates $\mathrm{H}(\{x^i(t)\})$ from Algorithm 1 by $\epsilon_t = \bar{L}_t\delta$ suffices to obtain over-approximations of the reachable sets $\mathcal{X}_t$.

IV. Proof of Theorem 1

We prove Theorem 1 using convex geometry and optimal control. We discuss these results in Section IV-D.

A. Searching for extreme points of $H(\mathcal{X})$

Let $d \in \mathcal{S}^{n-1}$ be a search direction and define the optimal control problem (OCP)

$$\textbf{(OCP}_d\textbf{)} \qquad \begin{aligned} \inf_{w \in L^\infty([0,T], \mathcal{W})} \quad & -d^\top x(T) \\ \text{s.t.} \quad & \dot{x}(t) = f(x(t)) + w(t), \\ & x(0) = x^0, \quad t \in [0,T]. \end{aligned}$$

$\textbf{OCP}_d$ is well-posed under Assumptions 1 and 2, i.e., it admits at least one solution $w_d \in L^\infty([0,T], \mathcal{W})$ (see, e.g., [30, Theorem 6.2.1], and note that $w(\cdot) = 0$ is feasible). Intuitively, solving $\textbf{OCP}_d$ gives a reachable state $x_d(T) \in \mathcal{X}_T$ that is the furthest in the direction d.

B. Reformulating $\textbf{OCP}_d$ using the PMP to reduce the search of solutions from $L^\infty([0,T], \mathbb{R}^n)$ to $\mathbb{R}^n$

The Pontryagin Maximum Principle (PMP) [19], [34], [35] gives necessary conditions of optimality for $\textbf{OCP}_d$. As the Hamiltonian of $\textbf{OCP}_d$ is given by $H(x,w,p) = p^\top(f(x)+w))$, for any locally-optimal solution (x_d, w_d) of $\textbf{OCP}_d$, there exists an absolutely-continuous function $p_d : [0,T] \to \mathbb{R}^n$, called the adjoint vector, such that

$$\dot{p}_d(t) = -\nabla f(x_d(t))^\top p_d(t) \text{ for a.e. } t \in [0,T], \quad (11a)$$

$$p_d(T) = -d^\top \quad (11b)$$

$$w_d(t) = \arg\min_{v \in \mathcal{W}} \ p_d(t)^\top v \quad \text{for a.e. } t \in [0,T], \quad (11c)$$

with $\dot{x}_d(t) = f(x_d(t)) + w_d(t)$ and $x_d(0) = x^0$. A tuple (x_d, p_d, w_d) satisfying the above equations is called (Pontryagin) extremal for $\mathbf{OCP}_d$. These equations indicate that the adjoint vector is non-zero at all times.

Lemma 3 (No singular arcs): Let (x_d, p_d, w_d) be an extremal for $\mathbf{OCP}_d$ with $d \in \mathcal{S}^{n-1}$. Then, $p_d(t) \neq 0$ for every $t \in [0, T]$, i.e., $\mathbf{OCP}_d$ has no singular arcs.

Proof: By contradiction, $p_d(t) = 0$ for some $t \in [0, T]$. Then, 0 is the unique solution to the ODE $\dot{p}(s) \overset{(11a)}{=} -\nabla f(x_d(s))^\top p(s) = 0$ for $s \in [t, T]$ with $p(t) = 0$. Thus, $-d \overset{(11b)}{=} p_d(T) = 0$, a contradiction. ∎

Thanks to Lemma 3 and Assumption 2, the maximality condition (11c) can be simplified. First, since $p_d(t) \neq 0$ for all $t \in [0, T]$ thanks to Lemma 3, (11c) is well-defined. Second, since $\mathcal{W}$ is convex and $v \mapsto p_d(t)^\top v$ is linear, searching for disturbances in $\partial\mathcal{W}$ suffices, i.e., $w_d(t) = \arg\min_{v \in \partial\mathcal{W}} p_d(t)^\top v$. Then,

$$w_d(t) = \arg\min_{v \in \partial\mathcal{W}} \; p_d(t)^\top v = \arg\min_{v \in \partial\mathcal{W}} \; \frac{p_d(t)}{\|p_d(t)\|}^\top v$$
$$= \left(n^{\partial\mathcal{W}}\right)^{-1} \left(-\frac{p_d(t)}{\|p_d(t)\|}\right) \tag{12}$$

where $n^{\partial\mathcal{W}} : \partial\mathcal{W} \to \mathcal{S}^{n-1}$ is the Gauss map in (5), which is a diffeomorphism since $\partial\mathcal{W}$ is an ovaloid [29] by Assumption 2. The last equality in (12) follows from (6) (note that (6) $= 0$ if and only if $v = w(t)$, due to the strict convexity of $\partial\mathcal{W}$). Thus, by combining (11) and (12), we obtain that candidate optimal solutions of $\mathbf{OCP}_d$ must solve the boundary-value problem (BVP)

$$\mathbf{BVP}_d : \begin{cases} \dot{x}_d(t) = f(x_d(t)) + (n^{\partial\mathcal{W}})^{-1}\left(-\frac{p_d(t)}{\|p_d(t)\|}\right), \\ \dot{p}_d(t) = -\nabla f(x_d(t))^\top p_d(t), \quad t \in [0, T], \\ (x_d(0), p_d(T)) = (x^0, -d). \end{cases} \tag{13}$$

With $\mathbf{BVP}_d$, we reduced the search of solutions to $\mathbf{OCP}_d$ from $w \in L^\infty([0, T], \mathbb{R}^n)$ to $p_d(0) \in \mathbb{R}^n$.

C. Reformulating $\mathbf{BVP}_d$ with knowledge of $w_d(0)$

From (12) and $\mathbf{BVP}_d$, we observe that extremal disturbances w_d are continuous. Thus, we expect the (pointwise) knowledge of $w_d(0)$ to be informative. In this section, we flip the problem around and show that sampling initial values of the disturbances $w_d(0)$ suffices to recover the convex hulls of the reachable sets. Specifically, solving BVPs is not necessary: integrating an ODE from different values of $w_d(0) \in \partial\mathcal{W}$ suffices.

$\mathbf{BVP}_d$ indicates that extremal state trajectories follow dynamics that only depend on $\frac{p_d(t)}{\|p_d(t)\|}$. Thus, we define

$$q_d(t) = -\frac{p_d(t)}{\|p_d(t)\|}, \tag{14}$$

which is well-defined thanks to Lemma 3, so that extremal state trajectories follow dynamics $\dot{x}_d = f(x_d) + (n^{\partial\mathcal{W}})^{-1}(q_d)$ that only depend on q_d. The next result states that q_d is the solution of an ODE that is completely characterized by the initial value of the disturbance.

Lemma 4 (Extremals of $\mathbf{OCP}$ are identified by $w(0)$): Let (x_d, p_d, w_d) be an extremal for $\mathbf{OCP}_d$ with $d \in \mathcal{S}^{n-1}$. Then, the map q_d satisfying $q_d(t) = (14) \in \mathcal{S}^{n-1}$ for all $t \in [0, T]$ is well-defined and (x_d, q_d) solves $\mathbf{ODE}_{w_d(0)}$.

Proof: Thanks to Lemma 3, $q_d(t)$ is well-defined for all $t \in [0, T]$. Moreover, $w_d(0) \in \partial\mathcal{W}$ thanks to (12), so $n^{\partial\mathcal{W}}(w_d(0))$ is well-defined. For conciseness, we drop dependencies on d and prove the second claim. First, thanks to (12), $q(0) = n^{\partial\mathcal{W}}(w(0))$. Second,

$$\dot{q}_t = -\frac{\dot{p}_t}{\|p_t\|} + p_t \frac{p_t^\top \dot{p}_t}{\|p_t\|^3} \overset{(11a)}{=} -\mathsf{Proj}_{q_t}\left(\nabla f(x_t)^\top q_t\right),$$

denoting $(x(t), q(t)) = (x_t, q_t)$. This gives $\mathbf{ODE}_{w_d(0)}$. ∎

Lemma 4 yields the following intermediate result.

Lemma 5: Assume that f and $\mathcal{W}$ satisfy Assumptions 1 and 2 and define the map F by (8) as in Theorem 1. Then, for all $t \in [0, T]$, $F(\cdot, t)$ is smooth and

$$(\partial \mathrm{H}(\mathcal{X}_t) \cap \mathcal{X}_t) \subseteq F(\partial\mathcal{W}, t) \subseteq \mathcal{X}_t. \tag{15}$$

Proof: Without loss of generality, we prove the result for $t = T$. Extending the proof to $t \in [0, T)$ (by defining $\mathbf{OCP}_d^t$ maximizing $-d^\top x(t)$, which results in the same expressions for $\mathbf{ODE}$ and F) is straightforward.

First, $F(\cdot, T)$ is smooth since it is the composition of $n^{\partial\mathcal{W}}$ (which is a diffeomorphism thanks to Assumption 2) with the solution of an ODE with smooth coefficients.

Second, $F(\partial\mathcal{W}, T) \subseteq \mathcal{X}_T$ by definition. To show the other inclusion, let $y \in \partial \mathrm{H}(\mathcal{X}_T) \cap \mathcal{X}_T$. Since $y \in \mathcal{X}_T$, there exists some $w \in L^\infty([0, T], \mathcal{W})$ such that $y = x_w(T)$ where x_w solves the ODE in (1). Then, since $x_w(T) = y \in \partial \mathrm{H}(\mathcal{X}_T)$, by the convexity of $\mathrm{H}(\mathcal{X}_T)$ and Lemma 1, $x_w(T)$ is such that $\tilde{w} \mapsto d^\top x_{\tilde{w}}(T)$ is maximized over $\tilde{w} \in L^\infty([0, T], \mathcal{W})$ for some $d \in \mathcal{S}^{n-1}$, i.e., (x_w, w) solves $\mathbf{OCP}_d$. Thus, by Lemma 4, x_w solves $\mathbf{ODE}_{w(0)}$ for some $w(0) \in \partial\mathcal{W}$. We obtain $y = x_w(T) = F(w(0), T) \in F(\partial\mathcal{W}, T)$. ∎

The proof of Theorem 1 almost immediately follows from Lemma 5 and the following geometric result.

Lemma 6: Let $A \subset \mathbb{R}^n$ be a compact set. Then, $\mathrm{H}(A) = \mathrm{H}(\partial A) = \mathrm{H}(\partial \mathrm{H}(A) \cap A)$.

The proof of Lemma 6 uses the Krein-Milman theorem [36] and is given in the Appendix.

Proof of Theorem 1: $\mathcal{X}_t$ is compact (Lemma 2). (3) follows from taking the convex hull on both sides of (15) and as $\mathrm{H}(\partial \mathrm{H}(\mathcal{X}_t) \cap \mathcal{X}_t) = \mathrm{H}(\mathcal{X}_t)$ (Lemma 6). □

Corollary 1 directly follows from Theorem 1. Corollary 2 follows from Theorem 1 and [33, Lemma 4.2].

D. Discussion and insights

In Table I, we summarize the different problems used to derive $\mathbf{ODE}_{w(0)}$ and ultimately prove Theorem 1.

TABLE I: Problems defined to prove Theorem 1.

Problem	unknown variables	number of variables
$\mathbf{OCP}_d$	$w \in L^\infty([0,T],\mathcal{W})$	infinite
$\mathbf{BVP}_d$	$p(0) \in \mathbb{R}^n$	n
$\mathbf{ODE}_{w_d(0)}$	None	0

At first sight, $\mathbf{BVP}_d$ suggests using the procedure outlined in Algorithm 2 to reconstruct the convex hull of the reachable set $\mathrm{H}(\mathcal{X}_T)$. However, this procedure can be computationally expensive and difficult to implement, since solving BVPs is generally challenging. Moreover, the non-convexity of $\mathbf{OCP}_d$ indicates that solutions to $\mathbf{BVP}_d$ could be local minima, so Algorithm 2 would potentially under-estimate the true reachable sets and be unsuitable for applications that require over-approximations of the reachable sets. Carrying on the analysis using samples of $w_d(0)$ as in Section IV-C is key to our characterization.

Alg. 2. Shooting method

for all $d \in \mathcal{S}^{n-1}$ **do**
  $x_d \leftarrow \text{Solve}(\mathbf{BVP}_d)$
return $\mathrm{H}\left(x_d(T), d \in \mathcal{S}^{n-1}\right)$

Lemma 4 implies that extremal trajectories are completely specified by the initial value $w_d(0)$ of the disturbance that is necessarily on the boundary of $\mathcal{W}$ thanks to (12). Thus, given $w_d(0) \in \partial\mathcal{W}$, we can determine the corresponding reachable extremal state $x_d(T)$ by simply integrating $\mathbf{ODE}_{w_d(0)}$, which is independent of the search direction d (d is implicitly encoded in $w_d(0)$). This observation is the key insight behind Algorithm 1, which consists of integrating $\mathbf{ODE}_{w_d(0)}$ for different values of $w_d(0) \in \partial\mathcal{W}$ to recover the convex hulls of the reachable sets. This algorithm is justified by Theorem 1.

V. Results and Applications

In this section, we evaluate Algorithm 1 on a neural feedback reachability problem. We also design a robust MPC method with Algorithm 1, that we evaluate on the problem of robustly stabilizing the attitude of a spacecraft. Computation times are measured on a laptop with an 1.10GHz Intel Core i7-10710U CPU and all code is available at `github.com/StanfordASL/convex_hull_reachability`.

A. Neural feedback loop analysis

We first consider the problem of evaluating the reachable sets of the system $\dot{x}(t) = Ax(t) + B\pi(x(t)) + w(t)$, where π is a neural network and $\mathcal{W} = B(0, \sqrt{2}/20)$. We consider the system and neural network from [37, Sec. VIII.A-C] and [18] (with $n = 2$), but replace the `ReLU` activation functions with smooth `Softplus` activations so that Assumptions 1 and 2 are satisfied.[2] We discretize the dynamics with an Euler scheme at $\Delta t = 0.25$s and predict reachable sets over a horizon $T = 4$s. The initial disturbance values $w^i(0)$ for Algorithm 1 are selected to evenly cover the circle $\partial\mathcal{W}$.

[2]Note that by selecting appropriate hyperparameters, `Softplus` activations can be made arbitrarily close to `ReLU` activation functions.

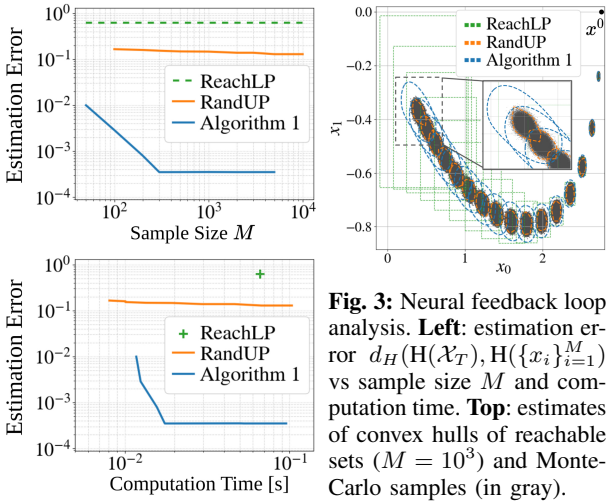

Fig. 3: Neural feedback loop analysis. **Left**: estimation error $d_H(\mathrm{H}(\mathcal{X}_T), \mathrm{H}(\{x_i\}_{i=1}^M)$ vs sample size M and computation time. **Top**: estimates of convex hulls of reachable sets ($M = 10^3$) and Monte-Carlo samples (in gray).

First, we empirically validate Theorem 1. We run Algorithm 1 with $M = 10^3$ samples $w^i(0) \in \partial\mathcal{W}$. Then, we uniformly sample 10^5 disturbances $w(t) \in \mathcal{W}$ at each timestep and evaluate the corresponding trajectories. We verify that all resulting trajectories are within the convex hulls computed by Algorithm 1, which empirically validates Theorem 1. Thus, using Algorithm 1, it suffices to sample initial disturbance values $w^i(0) \in \partial\mathcal{W}$ to reconstruct the convex hulls of the reachable sets.

Second, we compare Algorithm 1 with a baseline that randomly samples disturbances $w(t) \in \mathcal{W}$ at each timestep and returns the convex hulls of closed-loop trajectories (`RandUP` [18]). As ground truth, we use Algorithm 1 with a very large number of samples (10^4), which is justified by Theorem 1. We report results in Figure 3. Algorithm 1 returns estimates that are orders of magnitude more accurate than the baseline's estimates. Given a desired precision, Algorithm 1 is thus orders of magnitude faster than the baseline. This difference is a direct consequence of Theorem 1: only initial values of the disturbances $w(0) \in \partial\mathcal{W}$ need to be evaluated using Algorithm 1. In contrast, the baseline requires sampling a much larger number of variables that increases with the number of prediction timesteps, resulting in significantly worse sample complexity (see [18], [33] for error bounds) and trajectories that are far from the reachable set boundaries. Recall that Algorithm 1 returns under-approximations of the convex hulls of the reachable sets, since any extremal trajectory $x_{w(0)}$ solving $\mathbf{ODE}_{w(0)}$ is contained in the true reachable sets $\mathcal{X}_t$. These simulations show that naive Monte-Carlo estimates poorly approximate the convex hulls of the reachable sets for this problem. In Figure 3, we also report the over-approximations from a formal method (`ReachLP` [37]) that significantly over-estimate the reachable sets. In contrast, Algorithm 1 returns accurate approximations using a small sample size.

B. Robust MPC for the attitude control of a spacecraft

We design an attitude controller for a spacecraft with state $x = (q, \omega) \in \mathbb{R}^7$, control $u \in \mathbb{R}^3$, and dynamics

$$\dot{q}(t) = \Omega(\omega(t))q(t), \tag{16a}$$
$$\dot{\omega}(t) = J^{-1}(u(t) - S(\omega(t))J\omega(t) + w(t)), \tag{16b}$$

where $x(0) = (q^0, \omega^0)$, $w(t) \in \mathcal{W} = B(0, 10^{-2})$, $J = \text{diag}(5, 2, 1)$ is the diagonal inertia matrix, and

$$\Omega(\omega) = \frac{1}{2}\begin{bmatrix} 0 & -\omega_1 & -\omega_2 & -\omega_3 \\ \omega_1 & 0 & \omega_3 & -\omega_2 \\ \omega_2 & -\omega_3 & 0 & \omega_1 \\ \omega_3 & \omega_2 & -\omega_1 & 0 \end{bmatrix}, \; S(\omega) = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}.$$

As in [15], we constrain the angular velocity and input

$$\|\omega(t)\|_\infty \leq 0.1, \quad \|u(t)\|_\infty \leq 0.1, \quad t \in [0, T]. \tag{17}$$

We consider feedback control trajectories parameterized by $u(t) = \bar{u}(t) + K\omega(t)$, where $\bar{u} \in L^\infty([0, T], \mathbb{R}^3)$ is an open-loop control and $K \in \mathbb{R}^{3\times 3}$ is a feedback gain. With this control law, the closed-loop dynamics of the angular velocity ω in (16b) are independent of the spacecraft's orientation q. Intuitively, if opting for a receding horizon implementation, feedback on the orientation q is not necessary since disturbances w only affect the ω-component in (16b). Given $\bar{u} \in L^\infty([0, T], \mathbb{R}^3)$ and $t \in [0, T]$, we define the reachable set

$$\mathcal{R}_t(\bar{u}) = \left\{ \omega_w(\bar{u}, t) : \begin{array}{r} \omega_w(\bar{u}) \text{ solves (16b)} \\ u(t) = \bar{u}(t) + K\omega(t) \\ w \in L^\infty([0, T], \mathcal{W}) \end{array} \right\}. \tag{18}$$

Clearly, the convex hulls of the reachable sets $\text{H}(\mathcal{R}_t(\bar{u}))$ can be estimated using Algorithm 1, where $\textbf{ODE}_{w(0)}$ is defined using (16b) with $u(t) = \bar{u}(t) + K\omega(t)$.[3] Thus, given M samples $w^i(0) \in \partial\mathcal{W}$ and $\epsilon_t > 0$ large-enough, the constraints (17) can be conservatively approximated by

$$-0.1 + \epsilon_t \leq \omega^i(\bar{u}, t)_j \leq 0.1 - \epsilon_t, \tag{19a}$$
$$-0.1 + \epsilon_t \leq (\bar{u}(t) + K\omega^i(\bar{u}, t))_j \leq 0.1 - \epsilon_t, \tag{19b}$$

for all $j = 1, 2, 3$, $i = 1, \ldots, M$, $t \in [0, T]$. The conservatism of (19) follows from the convexity of (17) and Corollary 2, see [33, Cor. 5.4]. Note that the convex hulls of the ω^i are never explicitly computed. Given a reference $x_{\text{ref}} = (1, 0, \ldots, 0)$ and $(Q, R) = (10I_{7\times 7}, I_{3\times 3})$, we define the robust control problem ($\textbf{OCP}(x^0)$):

$$\inf_{\bar{u}} \int_0^T \left((x_0(t) - x_{\text{ref}})^\top Q(x_0(t) - x_{\text{ref}}) + \bar{u}(t)^\top R\bar{u}(t)\right) \text{d}t$$
$$\text{s.t. } x_0 \text{ solves (16) with } w(\cdot) = 0,$$
$$\bar{u} \in L^\infty([0, T], \mathbb{R}^3) \text{ satisfies (19).}$$

By recursively solving $\textbf{OCP}(x^0)$ and applying the computed control inputs, we obtain the receding horizon robust MPC controller in Algorithm 3. We use $M = 100$ samples of $w(0)$ and the error bounds ϵ_t in Corollary 2. Our MPC controller achieves a replanning rate of 10Hz with a Python implementation. We refer to the appendix and the open-source code for further details.

[3] The dynamics (16b) defining $\textbf{ODE}_{w(0)}$ are time-varying due to the control $\bar{u}$. As the equations (11) from the PMP are unchanged for time-varying dynamics $f(t, x)$, the derivations in Section IV still hold, so Algorithm 1 can be used to reconstruct the convex hulls $\text{H}(\mathcal{R}_t(\bar{u}))$.

Alg. 3. Robust model predictive control

Input: M initial disturbances $w^i(0)$, initial state $x(0)$
for all $k = 0, 1, 2 \ldots$ **do**
  $\bar{u}^k \leftarrow \text{Solve}(\textbf{OCP}(x(k\Delta t))$
  Apply $u^k(t) = \bar{u}^k(t) + K\omega(t)$ for $t \in [0, \Delta t]$
  Observe $x((k+1)\Delta t)$

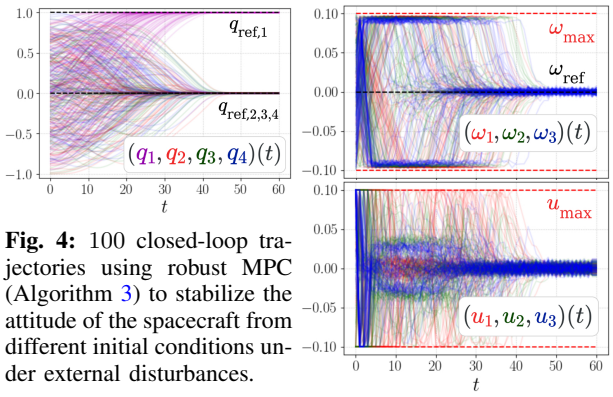

Fig. 4: 100 closed-loop trajectories using robust MPC (Algorithm 3) to stabilize the attitude of the spacecraft from different initial conditions under external disturbances.

MPC results: We evaluate the proposed controller with 100 experiments with uniformly-sampled disturbances and initial states. Results in Figure 4 show that the system stabilizes at the reference x_{ref} and the original constraints (17) are always satisfied. The optimization problem is always feasible in these experiments. The error bounds given in Corollary 2 introduce reasonable conservatism. By increasing the sample size M, this conservatism can be made arbitrarily small.

Comparisons with other reachability methods: After discretizing the dynamics (16) with $\Delta t = 1$s as

$$x((k+1)\Delta t) = \bar{f}(x(k\Delta t), \bar{u}(k\Delta t)) + w(k\Delta t), \tag{20}$$

where $\bar{f}$ is given by a Runge-Kutta (RK4) scheme, we can compare the reachable set convex hull estimates from our method (Algorithm 1) with those from two other standard methods. The first baseline is a sampling-based method (`RandUP` [16]) that estimates the convex hulls $\text{H}(\mathcal{X}_t)$ with the convex hulls of trajectories from (20) with samples of $w(k\Delta t)$. The second standard baseline propagates uncertainty from the disturbances using a linear model of (20) and bounds the approximation error using the Lipschitz constant of the Jacobian $\nabla_x \bar{f}(x, u)$ in (20), see the Appendix for more details.

Given a control trajectory $\bar{u}$ solving $\textbf{OCP}(x^0)$, we present results in Figure 5. Due to space constraints, we only represent the third components of ω and u. First, the Lipschitz-based and the naive sampling-based baselines are the fastest (both have runtime at ~ 0.5 ms),

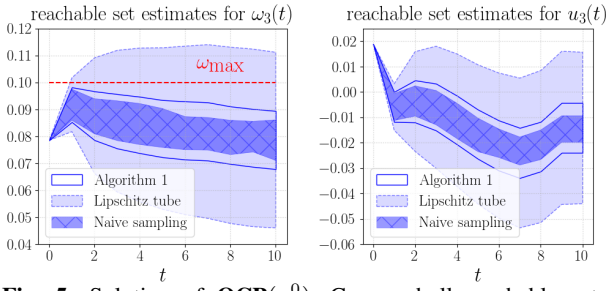

Fig. 5: Solution of $\mathbf{OCP}(x^0)$. Convex hull reachable sets estimates computed with Algorithm 1 and two baselines.

followed by Algorithm 1 (~ 2 ms). However, the over-approximations of the reachable sets from the Lipschitz-based method are significantly more conservative than those from Algorithm 1. A controller using the reachable set estimates from this baseline would deem $\bar{u}$ to potentially violate constraints and would thus be more conservative than the proposed robust MPC approach. Also, the naive sampling-based baseline significantly under-estimates the true convex hulls. One can show that this baseline performs worse as the discretization is refined, see also Section V-A. In contrast, Algorithm 1 is derived in continuous time so its sample complexity is independent of the discretization of the dynamics. Specifically, Algorithm 1 only requires sampling the initial disturbances values $w^i(0)$, so its precision only depends on the accuracy of the discretization of $\mathbf{ODE}_{w(0)}$.

VI. Conclusion and Outlook

Theorem 1 reveals the structure of the convex hull of reachable sets of a general class of nonlinear dynamical systems. The result unlocks an estimation algorithm that consists of integrating an augmented ODE from different initial values of the disturbances. Simulations show that the proposed method is fast and significantly more accurate than baselines. We also demonstrate its applicability to robust model predictive control.

This work opens exciting future directions of research. First, we plan to investigate extensions to reachability problems with uncertain initial states, dynamics of the form $\dot{x} = f(x) + g(x)w$, uncertain parameters, and more general disturbance sets $\mathcal{W}$. Second, studying the boundary structure of the convex hulls could yield tighter error bounds improving upon those in Corollary 2 (perhaps using analysis inspired from [33]). Finally, Theorem 1 and the derivation of $\mathbf{ODE}_{w(0)}$ could inform the design of more efficient reachability tools that exploit additional properties of the dynamics.

Acknowledgements: The authors thank the anonymous reviewers for their helpful feedback. The NASA University Leadership Initiative (grant #80NSSC20M0163) provided funds to assist the authors with their research, but this article solely reflects the opinions and conclusions of its authors and not any NASA entity. Toyota Research Institute provided funds to support this work.

References

[1] P.-J. Meyer, A. Devonport, M. Arcak, *Interval Reachability Analysis*. Springer Cham, 2021.
[2] T. Koller, F. Berkenkamp, M. Turchetta, A. Krause, "Learning-based model predictive control for safe exploration," in *Proc. IEEE Conf. on Decision and Control*, 2018.
[3] M. Althoff, O. Stursberg, M. Buss, "Reachability analysis of nonlinear systems with uncertain parameters using conservative linearization," in *Proc. IEEE Conf. on Decision & Control*, 2008.
[4] B. Houska, M. E. Villanueva, "Robust optimization for MPC," in *Handbook of Model Predictive Control*. Springer International Publishing, 2018, pp. 413–443.
[5] M. Althoff, G. Frehse, A. Girard, "Set propagation techniques for reachability analysis," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, no. 1, pp. 369–395, 2021.
[6] N. Ramdani, N. Meslem, Y. Candau, "A hybrid bounding method for computing an over-approximation for the reachable set of uncertain nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 54, no. 10, pp. 2352–2364, 2009.
[7] J. K. Scott, P. I. Barton, "Bounds on the reachable sets of nonlinear control systems," *Automatica*, vol. 49, no. 1, pp. 93–100, 2013.
[8] M. Berz, K. Makino, "Verified integration of ODEs and flows using differential algebraic methods on high-order Taylor models," *Reliable Computing*, vol. 4, no. 4, pp. 361–369, 1998.
[9] X. Chen, E. Abraham, S. Sankaranarayanan, "Flow*: An analyzer for non-linear hybrid systems," in *Proc. Int. Conf. Computer Aided Verification*, 2013.
[10] P.-J. Meyer, A. Devonport, M. Arcak, "TIRA: Toolbox for interval reachability analysis," in *Hybrid Systems: Computation and Control*, 2019, pp. 224–229.
[11] S. Coogan, M. Arcak, "Efficient finite abstraction of mixed monotone systems," in *Hybrid Systems: Computation and Control*, 2015.
[12] M. Abate, S. Coogan, "Robustly forward invariant sets for mixed-monotone systems," *IEEE Transactions on Automatic Control*, vol. 67, no. 9, pp. 4947–4954, 2022.
[13] S. Singh, A. Majumdar, J.-J. Slotine, M. Pavone, "Robust online motion planning via contraction theory and convex optimization," in *Proc. IEEE Conf. on Robotics and Automation*, 2017.
[14] S. Yu, C. Maier, H. Chen, F. Allgöwer, "Tube MPC scheme based on robust control invariant set with application to lipschitz nonlinear systems," *Systems & Control Letters*, vol. 62, no. 2, pp. 194–200, 2013.
[15] A. P. Leeman, J. Köller, A. Zanelli, S. Bennani, M. N. Zeilinger, "Robust nonlinear optimal control via system level synthesis," 2023, available at https://arxiv.org/abs/2301.04943.
[16] T. Lew, M. Pavone, "Sampling-based reachability analysis: A random set theory approach with adversarial sampling," in *Conf. on Robot Learning*, 2020.
[17] A. J. Thorpe, K. R. Ortiz, O. M. M. K., "Learning approximate forward reachable sets using separating kernels," in *Learning for Dynamics & Control Conference*, 2021.
[18] T. Lew, L. Janson, R. Bonalli, M. Pavone, "A simple and efficient sampling-based algorithm for reachability analysis," in *Learning for Dynamics & Control Conference*, 2022.
[19] A. A. Agrachev, Y. L. Sachkov, *Control Theory from the Geometric Viewpoint*. Springer Berlin Heidelberg, 2004.
[20] A. J. Krener, H. Schättler, "The structure of small-time reachable sets in low dimensions," *SIAM Journal on Control and Optimization*, vol. 27, no. 1, pp. 120–147, 1989.
[21] H. Schättler, U. Ledzewicz, *Geometric Optimal Control*. Springer New York, 2012.
[22] T. Pecsvaradi, K. S. Narendra, "Reachable sets for linear dynamical systems," *Information and Control*, vol. 19, no. 4, pp. 319–344, 1971.
[23] A. Y. Gornov, T. S. Zarodnyuk, E. A. Finkelstein, A. S. Anikin, "The method of uniform monotonous approximation of the

reachable set border for a controllable system," *Journal of Global Optimization*, vol. 66, no. 1, pp. 53–64, 2015.
[24] R. Bonalli, B. Hérissé, E. Trélat, "Solving optimal control problems for delayed control-affine systems with quadratic cost by numerical continuation," in *American Control Conf.*, 2017.
[25] ——, "Optimal control of endoatmospheric launch vehicle systems: Geometric and computational issues," *IEEE Transactions on Automatic Control*, vol. 65, no. 6, pp. 2418–2433, 2020.
[26] R. Schneider, *Convex Bodies: The Brunn-Minkowski Theory*, 2nd ed. Cambridge Univ. Press, 2014.
[27] T. Lorenz, "Boundary regularity of reachable sets of control systems," *System & Control Letters*, vol. 54, no. 9, 2005.
[28] P. Cannarsa, H. Frankowska, "Interior sphere property of attainable sets and time optimal control problems," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 12, no. 2, 2006.
[29] J. Rauch, "An inclusion theorem for ovaloids with comparable second fundamental forms," *Journal of Differential Geometry*, vol. 9, no. 4, 1974.
[30] E. Trélat, *Contrôle optimal: théorie et applications*. Vuibert, Collection "Mathématiques Concrètes", 2008.
[31] A. Bressan, B. Piccoli, *Introduction to the Mathematical Theory of Control*. American Institute of Mathematical Sciences, 2007.
[32] J. M. Lee, *Introduction to Smooth Manifolds*, 2nd ed. Springer New York, 2012.
[33] T. Lew, R. Bonalli, L. Janson, M. Pavone, "Estimating the convex hull of the image of a set with smooth boundary: error bounds and applications," 2023, https://arxiv.org/abs/2302.13970.
[34] L. S. Pontryagin, *Mathematical Theory of Optimal Processes*. Taylor & Francis, 1987.
[35] E. Trélat, "Optimal control and applications to aerospace: Some results and challenges," *Journal of Optimization Theory & Applications*, vol. 154, no. 3, pp. 713–758, 2012.
[36] A. Grothendieck, *Topological Vector Spaces*, 1st ed. New York: Gordon and Breach Science Publishers, 1973, translated by Chaljub, Orlando.
[37] M. Everett, G. Habibi, S. Chuangchuang, J. P. How, "Reachability analysis of neural feedback loops," *IEEE Access*, 2021.
[38] A. González, "Measurement of areas on a sphere using fibonacci and latitude-longitude lattices," *Mathematical Geosciences*, vol. 42, no. 1, pp. 49–64, 2009.
[39] R. Bonalli, T. Lew, M. Pavone, "Analysis of theoretical and numerical properties of sequential convex programming for continuous-time optimal control," *IEEE Transactions on Automatic Control*, 2022.
[40] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, Q. Zhang, "JAX: composable transformations of Python+NumPy programs," 2018.
[41] B. Stellato, G. Banjac, P. Goulart, A. Bemporad, S. Boyd, "OSQP: an operator splitting solver for quadratic programs," *Mathematical Programming Computation*, vol. 12, no. 4, pp. 637–672, 2020.

APPENDIX

A. Proof of Lemma 6 ($H(A)$=$H(\partial A)$=$H(\partial H(A) \cap A)$)

The first equality follows from $\mathrm{H}(\partial A) \subseteq \mathrm{H}(A)$ since $\partial A \subseteq A$ and from $\mathrm{H}(A) = \mathrm{H}(\mathrm{Ext}(A)) \subseteq \mathrm{H}(\partial A)$ by the Krein-Milman theorem [36] and since $\mathrm{Ext}(A) \subseteq \partial A$.

To show the second equality, we first note that $\mathrm{H}(\partial \mathrm{H}(A) \cap A) \subseteq \mathrm{H}(A)$ since $\partial \mathrm{H}(A) \cap A \subseteq A$. Second,

$$\begin{aligned}
\mathrm{H}(A) &= \mathrm{H}(\mathrm{H}(A)) && (\mathrm{H}(A) \text{ is convex}) \\
&= \mathrm{H}(\mathrm{Ext}(\mathrm{H}(A))) && (\text{Krein-Milman theorem [36]}) \\
&= \mathrm{H}\Big([\mathrm{Ext}(\mathrm{H}(A)) \setminus (\mathrm{Ext}(\mathrm{H}(A)) \cap A)] \cup [(\mathrm{Ext}(\mathrm{H}(A)) \cap A)] \Big) \\
&\overset{(1)}{=} \mathrm{H}(\mathrm{Ext}(\mathrm{H}(A)) \cap A) \overset{(2)}{\subseteq} \mathrm{H}(\partial \mathrm{H}(A) \cap A)
\end{aligned}$$

since (1) $\mathrm{Ext}(\mathrm{H}(A)) \setminus (\mathrm{Ext}(\mathrm{H}(A)) \cap A) = \emptyset$ since $\mathrm{Ext}(\mathrm{H}(A)) \subseteq A$ and (2) $\mathrm{Ext}(\mathrm{H}(A)) \subseteq \partial \mathrm{H}(A)$. □

B. Details on the neural feedback loop analysis results

We consider the system and neural network from [37, Sec. VIII.A-C]. We fix the initial condition $x^0 = (11/4, 0)$ and use $\bar{\mathcal{W}} = \{w(t) \in \mathbb{R}^2 : \|w(t)\|_\infty \leq 1/20\}$ for `ReachLP` [37] (which only handles polytopic disturbance sets) and $\mathcal{W} = B(0, \sqrt{2}/20)$ for Algorithm 1 and `RandUP` [18]. Since $\bar{\mathcal{W}} \subset \mathcal{W}$, this choice for $\bar{\mathcal{W}}$ is fair as it reduces the conservatism of `ReachLP`. We refer to the open-source code for further details.

C. Details on the spacecraft attitude control results

Implementation: We use $T = 10$s, discretize the problem with a Runge-Kutta (RK4) scheme with $\Delta t = 1$ s, and enforce the constraints in (19) at each timestep $t_k = k\Delta t$, which is justified by the continuity of the state trajectories. The feedback gain K is given by a standard linear-quadratic regulator (LQR). We select $M = 100$ samples $w^i(0) \in \partial\mathcal{W}$ on a Fibonacci lattice [38], which gives an internal δ-covering of $\partial\mathcal{W}$ for $\delta = 3.5 \cdot 10^{-3}$. We evaluate $\bar{L}_t$ and $\bar{H}_t$ using 10^5 samples[4] and directly use the error bounds ϵ_t predicted in (10). This choice of ϵ_t ensures that the problem with the approximated constraints (19) gives solutions that satisfy the original constraints in (17), see also [33, Corollary 5.3]. We only parameterize the nominal state and control trajectories $(x_0, \bar{u})$ and evaluate (19) and its gradient as a function of $\bar{u}$. We use a standard sequential convex programming scheme [39] to solve the optimization problem. We use a Python implementation with Jax [40] and solve the convexified problems using OSQP [41]. As is common in MPC, we warm-start the optimization with the previously computed solution and perform a single SCP iteration per timestep, which allows solving the problem in 0.1s with a Python implementation. Computation time scales linearly with the sample size.

Lipschitz-based reachability baseline: After discretizing (16) as (20), this standard baseline computes a reachable tube for the angular velocities $\mathcal{T}_t = \{y \in \mathbb{R}^3 : (y - \omega_0(k\Delta t))^\top Q_k^{-1} (y - \omega_0(k\Delta t)) \leq 1\}$ where $x_0 = (q_0, \omega_0)$ follows nominal dynamics $x_0((k+1)\Delta t) = \bar{f}(x_0(k\Delta t), \bar{u}(k\Delta t))$ and the shape matrices $Q_k \in \mathbb{R}^{3\times 3}$ are recursively defined as $Q_0 = 0$, $Q_{k+1} = \frac{c+1}{c} Q_k^{\text{nom}} + (1+c) Q_k^{(\bar{w},\text{Lip})}$ for $k \in \mathbb{N}$, where the first term $Q_k^{\text{nom}} = \bar{A}_k Q_k \bar{A}_k^\top$ with $\bar{A}_k = \nabla_x \bar{f}(x_0(k\Delta t), \bar{u}(k\Delta t))$ propagates uncertainty at time $k\Delta t$ with a linearized model, $Q_k^{(\bar{w},\text{Lip})} = 3(\bar{w} + \frac{\bar{H}}{2} \lambda_{\max}(Q_k)^2)^2 I_3$ with $\bar{w} = 10^{-2}$ and $\bar{H}$ the Lipschitz constant of $\nabla_x \bar{f}$ accounts for the disturbance and the linearization error, and $c^2 = \mathrm{Trace}(Q_k^{\text{nom}})/\mathrm{Trace}(\bar{Q}_k^{(\bar{w},\text{Lip})})$. This standard baseline is described in [2] (see also [16] and [15]) and ensures that the trajectories resulting from the discrete-time dynamics (20) with $w(k\Delta t) \in \mathcal{W}$ are contained in the tube $\mathcal{T}$.

[4] We evaluate the empirical bound $\bar{L}_t = \max_j \|\nabla F(w^j(0), t)\|$ using 10^5 samples of the disturbances $w^j(0)$ and control inputs $\bar{u}^j$.