Received XX September 2021; revised ; accepted XX October 2021; Date of publication XX November 2021; date of current version XX November 2021. This paper was recommended by Associate Editor —-.

Digital Object Identifier 10.1109/OJCSYS.2021.Doi Number

Cautious optimization via data informativity

Jaap Eising[1], Jorge Cortés[2]

[1]Automatic Control Laboratory, ETH Zürich

[2]Department of Mechanical and Aerospace Engineering, University of California, San Diego

CORRESPONDING AUTHOR: J. Eising (e-mail: jeisign@ethz.ch)

A preliminary version of this work was submitted as [17] to the IEEE Conference on Decision and Control.

ABSTRACT This paper deals with the problem of accurately determining guaranteed suboptimal values of an unknown cost function on the basis of noisy measurements. We consider a set-valued variant to regression where, instead of finding a best estimate of the cost function, we reason over all functions compatible with the measurements and apply robust methods explicitly in terms of the data. Our treatment provides data-based conditions under which closed-forms expressions of upper bounds of the unknown function can be obtained, and regularity properties like convexity and Lipschitzness can be established. These results allow us to provide tests for point- and set-wise verification of suboptimality, and tackle the cautious optimization of the unknown function in both one-shot and online scenarios. We showcase the versatility of the proposed methods in two control-relevant problems: data-driven contraction analysis of unknown nonlinear systems and suboptimal regulation with unknown dynamics and cost. Simulations illustrate our results.

INDEX TERMS Data-based optimization, set-membership identification, data-driven control, contraction analysis

I. Introduction

In many real-world applications involving optimization, the cost or reward function is not fully known and has to be ascertained from measurements. Such situations can arise due to a variety of factors, including system complexity, large-scale structure, or lack of access to the relevant information in unknown or adversarial scenarios. These considerations motivate the need for quantifiable performance guarantees that can be used for control and optimization in safety-critical applications. In this paper, we are therefore interested in cautious optimization of unknown functions, i.e., accurately determining guaranteed suboptimal values on the basis of noisy measurements. We consider both one-shot scenarios, where we are required to make decisions on the basis of a given set of data, and online scenarios, where the suboptimization of the unknown function can be repeatedly refined based on the collection of additional measurements. Our approach, based on data informativity, allows us to derive results which are robust against worst-case situations.

Literature review: Data-based optimization is a widely applicable and thoroughly investigated area. Giving a full overview of all methods and results is not possible, and thus we focus here on those most relevant to our treatment. The problems considered here are robust optimization programs, see e.g. [2, 4] and references therein. Solution approaches to data-based optimization problems usually start by using the data to obtain an estimate of the unknown function which in some sense 'best' explains the measurements. This requires the definition of the class of admissible functions. Popular techniques are based on Gaussian processes, as in [37] or the approximation properties of neural networks [13, 40]. Of particular relevance to our analysis here is the work on set-membership estimation [11, 32, 33], which, instead of characterizing the best function, considers model-free set-valued uncertainties using Lipschitz interpolation.

In contrast to these approaches, we consider the situation where the unknown function can be written as a linear combination of a set of basis functions. In practice, this is not always possible, and thus the choice of basis induces a

model mismatch error, which needs to be taken into account. Common choices that balance expressiveness and the number of parameters are linear, polynomials, Gaussian, and sigmoidal functions. Depending on the problem, one may also choose a basis by taking into account physical considerations or employing spectral methods [9, 21, 25]. In the specific case of nonlinear system identification, an overview of the problem of selecting a suitable parametrization can be found in [34]. Once a basis is chosen, the data is used to perform regression [22, 42] on the parameters. What remains to be defined then is the metric chosen to decide the best estimate. Common methods are least squares (minimal Frobenius norm), ridge regression (minimal L_2 norm), and sparse or Lasso regression (minimal L_1 norm). The nonlinear system-identification literature has also brought forward methods which determine models that are sparse [8], of low rank (via dynamic mode decomposition) [27, 38, 39], or both [24]. Regardless of how one obtains an estimate, the second step in data-based optimization is to employ either a certainty-equivalent or robust method to find optimal values.

The alternative approach we use here is in line with the literature of data-driven control methods on the basis of Willems' fundamental lemma [47]. More specifically, we consider the concept of data informativity [43, 46]. Simply put, these methods take the viewpoint that guaranteed conclusions can be drawn from data only if *all* systems compatible with the measurements have this property. Of specific interest to this paper are works that deal with nonlinear systems and a number of recent works have worked with systems given in terms of bilinear [31] or polynomial basis functions [15, 20], albeit in discrete time. Simultaneously, there have been efforts in generalizing Willems' lemma to continuous time [30, 36] and linking the results in discrete and continuous time by investigating sampling [18]. A related approach is to perform system identification while keeping in mind the uncertainty bounds used in (robust) control, which is known as identification for control [19]. The convergence over time of similar set-based identification schemes is investigated in [28].

Apart from considering optimization, we apply our results to data-based contraction analysis of nonlinear systems [10, 29] and the regulation of an unknown system to a suboptimal point of an unknown cost function. Data-based regulation with known cost functions has been investigated in a number of forms, including linear quadratic regulation, see [16] and references therein, and more general cost functions via optimization-based controllers [5, 23]. In [12], such controllers have been employed to optimize unknown cost functions. Finally, the online optimization scenario considered here is inspired by methods such as extremum-seeking control [1, 26, 41]. However, our implementation and performance guarantees are markedly different.

Statement of contributions: We tackle the data-based optimization of an unknown vector-valued function with the goal of obtaining quantifiable performance guarantees. Starting from the assumption that the unknown function is parameterized in terms of a finite number of basis functions, we consider *set-valued regression*, i.e., we reason with the set of *all* parameters compatible with the measurements for which the noise satisfies a bound given in terms of a quadratic matrix inequality. We take a worst-case approach regarding noise realizations, leading us to provide solutions to the following *cautious optimization* problems:

(i) We provide conditions in terms of the measurements, basis functions, and noise model which guarantee that either the $2-$norm or a linear functional of the unknown function is upper bounded by a given $\delta \in \mathbb{R}$ for a given $z \in \mathbb{R}^n$. Moreover, these results make it possible for us to find the smallest worst-case bounds;

(ii) We also provide data-based conditions for convexity of either the true function or its worst-case norm bounds. These allow us to introduce a gradient-based method for finding the minimizer of the upper bound;

(iii) We also identify data-based conditions for Lipschitzness which, using interpolation, allows us to derive upper bounds guaranteed to hold over compact sets.

We illustrate the versatility of the results in two application scenarios: contraction analysis of nonlinear systems (for which we provide data-based tests to determine one-sided Lipschitz constants and use them to guarantee contractivity) and suboptimal regulation of unknown plants with unknown cost functions (for which we employ convexity and Lipschitzness to find a fixed input such that the value of the cost function at the corresponding fixed point can be explicitly bounded above). Our final contribution is to online data-based optimization. We describe a framework for incorporating new measurements of the unknown function which includes online local descent, a method that iteratively collects measurements locally and then minimizes an upper bound of the form described above. We prove that under mild assumptions on the collection of data, the set of parameters consistent with all collected measurements shrinks. This allows us to conclude that the upper bounds found with online local descent converge to the true optimizer of the unknown function.

Preliminary results of this paper were presented in the conference article [17], whose focus was restricted to scalar functions and provided upper bounds and analyzed convexity on the basis of data. The conference paper also presented a simplified version of online local descent for the scalar case. All of these are special cases of the present work. The generalization here from scalar to vector-valued functions is instrumental in expanding the range of possibilities for application in analysis and control of the proposed data-based optimization framework.

Notation: Throughout the paper, we use the following notation. We denote by $\mathbb{N}$ and $\mathbb{R}$ the sets of nonnegative integers and real numbers, respectively. We let $\mathbb{R}^{n\times m}$ denote the space of $n \times m$ real matrices and $\mathbb{S}^n$ the space of symmetric $n \times n$ matrices. For a vector $x \in \mathbb{R}^n$ we denote the standard

Preprint submitted to IEEE Open Journal of Control Systems. Received: May 9, 2025 08:56:43 Pacific Time

$p-$norm by $\|x\|_p$. Given an invertible matrix $R \in \mathbb{R}^{n\times n}$, we define the weighted $p-$norm by $\|x\|_{p,R} := \|Rx\|_p$. The i^{th} standard basis vector of $\mathbb{R}^n$ is denoted e_i. The induced matrix norm w.r.t. to the $p-$norm is denoted $\|A\|_p$ for $A \in \mathbb{R}^{n\times m}$. For vectors $v \in \mathbb{R}^n$, we write $v \geqslant 0$ (resp. $v > 0$) for elementwise nonnegativity (resp. positivity). The sets of such vectors are denoted $\mathbb{R}^n_{\geqslant 0} := \{v \in \mathbb{R}^n | v \geqslant 0\}$ and $\mathbb{R}^n_{>0} := \{v \in \mathbb{R}^n | v > 0\}$. For $P \in \mathbb{S}^n$, $P \geqslant 0$ (resp. $P > 0$) denotes that P is symmetric positive semi-definite (resp. definite). For a symmetric matrix M, we denote $\lambda_{\min}(M)$ and $\lambda_{\max}(M)$ for the smallest and largest eigenvalue of M. We denote the smallest singular value of $M \in \mathbb{R}^{n\times m}$ by $\sigma_-(M)$ and its Moore-Penrose pseudoinverse by $M^\dagger$. The convex hull and interior of $\mathcal{S} \subseteq \mathbb{R}^n$ are denoted by $\operatorname{conv}(\mathcal{S})$ and $\operatorname{int}(\mathcal{S})$, resp. Given a symmetric matrix $M \in \mathbb{S}^{u+v}$, when u and v are clear from context, we partition it as:

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix},$$

where $M_{11} \in \mathbb{S}^u$ and $M_{22} \in \mathbb{S}^v$. If $M_{22} \leqslant 0$ and $\ker M_{22} \subseteq \ker M_{12}$, we denote the (generalized) Schur complement of M with respect to M_{22} by $M|M_{22} := M_{11} - M_{12}M_{22}^\dagger M_{21}$. Lastly, such a partitioned matrix gives rise to a quadratic matrix inequality (QMI), whose set of solutions is denoted

$$\mathcal{Z}(M) := \left\{ Z \in \mathbb{R}^{v\times u} \mid \begin{bmatrix} I_u \\ Z \end{bmatrix}^\top M \begin{bmatrix} I_u \\ Z \end{bmatrix} \geqslant 0 \right\}.$$

II. Problem formulation

Consider a collection of known *basis functions* (or *features*), denoted $\phi_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \dots, k$. We collect the basis functions into a vector-function as

$$b : \mathbb{R}^n \to \mathbb{R}^k, \quad b(z) := \begin{bmatrix} \phi_1(z)^\top & \cdots & \phi_k(z)^\top \end{bmatrix}^\top. \tag{1}$$

Using these basis functions, we define linear combinations as parameterized by $\theta \in \mathbb{R}^{k\times m}$:

$$\phi^\theta : \mathbb{R}^n \to \mathbb{R}^m, \quad \phi^\theta(z) = \theta^\top b(z).$$

Our starting point is an unknown function $\hat{\phi} : \mathbb{R}^n \to \mathbb{R}^m$ that can be expressed in this form: i.e., there exists $\hat{\theta} \in \mathbb{R}^{k\times m}$ such that $\hat{\phi}(z) = \hat{\theta}^\top b(z)$.

Our goal is to investigate properties and find optimal values of the function $\hat{\phi}$ on the basis of measurements at points $\{z_i\}_{i=1}^T$. We assume the measurements are *noisy*, that is, we collect $y_i = \hat{\phi}(z_i) + w_i$, where w_i denotes an unknown disturbance vector for each i. To express this in compact form, define the matrices $Y, W \in \mathbb{R}^{m\times T}$ and $\Phi \in \mathbb{R}^{k\times T}$ by

$$\begin{aligned} Y &:= \begin{bmatrix} y_1 & \cdots & y_T \end{bmatrix}, \quad W := \begin{bmatrix} w_1 & \cdots & w_T \end{bmatrix}, \\ \Phi &:= \begin{bmatrix} b(z_1) & \cdots & b(z_T) \end{bmatrix} = \begin{bmatrix} \phi_1(z_1) & \dots & \phi_1(z_T) \\ \vdots & & \vdots \\ \phi_k(z_1) & \dots & \phi_k(z_T) \end{bmatrix}. \end{aligned} \tag{2}$$

Then, for the true value of $\hat{\theta}$, we have $Y = \hat{\theta}^\top \Phi + W$. In this equation, the matrices Y and Φ are known, and $\hat{\theta}$ and W are unknown.

A common line of reasoning (see e.g., [22]) to approximate the unknown function $\hat{\phi}$ is the following. Assuming that small noise samples are more probable than large noise samples, we attempt to find a solution θ to $Y = \theta^\top \Phi + W$ for which the Frobenius norm of W, denoted $\|W\|_F$, is as small as possible. This value is attained for $\theta = (Y\Phi^\dagger)^\top$.

Here instead we consider bounded noise and reason with the set of functions consistent with the measurements. Formally, we assume that the noise conforms to a model defined by a QMI given in terms of a partitioned matrix. The following assumption describes the noise model.

Assumption 1 (Noise model)**.** *The noise satisfies* $W^\top \in \mathcal{Z}(\Pi)$*, where* $\Pi \in \mathbb{S}^{m+T}$ *is such that* $\Pi_{22} < 0$ *and* $\Pi|\Pi_{22} \geqslant 0$.

According to [45, Thm. 3.2], under Assumption 1, the set $\mathcal{Z}(\Pi)$ is nonempty, convex, and bounded. A common example of such noise models is the case where $WW^\top = \sum_{i=1}^T w_i w_i^\top \leqslant Q$ for some $Q \geqslant 0$. This choice is analogous to assuming that w has bounded energy. If we, moreover, take $Q = \gamma T I_n$ or $Q = \gamma YY^\top$ we obtain bounds in terms of the horizon, or signal-to-noise bounds. Lastly, as shown in [45, Sec. 5.4], these noise models can be employed as confidence intervals corresponding to a given probability in the setting of Gaussian noise.

Under Assumption 1, one can describe the set of parameters consistent with the measurements as

$$\Theta = \{\theta \in \mathbb{R}^{k\times m} \mid Y = \theta^\top \Phi + W, W^\top \in \mathcal{Z}(\Pi)\}. \tag{3}$$

Note that we can write

$$Y = \theta^\top \Phi + W \iff \begin{bmatrix} I_m & W \end{bmatrix} = \begin{bmatrix} I_m & \theta^\top \end{bmatrix} \begin{bmatrix} I_m & Y \\ 0 & -\Phi \end{bmatrix}.$$

Thus, if we define

$$N := \begin{bmatrix} I_m & Y \\ 0 & -\Phi \end{bmatrix} \Pi \begin{bmatrix} I_m & Y \\ 0 & -\Phi \end{bmatrix}^\top, \tag{4}$$

it follows immediately that $\Theta = \mathcal{Z}(N)$. We refer to the procedure of obtaining the set of parameters Θ from the measurements as *set-valued regression*.

Remark II.1. (Richness of the measurements) Note that the set Θ is compact if and only if $N_{22} < 0$. Since $N_{22} = \Phi\Pi_{22}\Phi^\top$ and $\Pi_{22} < 0$, this holds if and only if Φ has full row rank. In turn, this requires that the basis functions are not identical and that the set of points $\{z_i\}_{i=1}^T$ is 'rich' enough to distinguish them. If the unknown function represents the vector field of a dynamical system, such rank properties are often referred to as (persistent) 'excitation', cf. [47]. •

Remark II.2. (Least-squares estimate) One can check that, under Assumption 1, $\theta^{\text{lse}} := -N_{22}^\dagger N_{21} \in \Theta$. This means that the function $(\theta^{\text{lse}})^\top b(z)$ is consistent with the data. In fact,

$$N|N_{22} = \begin{bmatrix} I_m \\ -N_{22}^\dagger N_{21} \end{bmatrix}^\top N \begin{bmatrix} I_m \\ -N_{22}^\dagger N_{21} \end{bmatrix} \geqslant \begin{bmatrix} I_m \\ \theta \end{bmatrix}^\top N \begin{bmatrix} I_m \\ \theta \end{bmatrix},$$

for any $\theta^\top \in \mathcal{Z}(N)$. Therefore, θ^{lse} is the value that maximizes the value of the quadratic inequality. This leads us to refer to the function $\phi^{\text{lse}}(z;\Theta) := (\theta^{\text{lse}})^\top b(z) = -N_{12}N_{22}^\dagger b(z)$ as the *least-squares estimate* of $\hat{\phi}(z)$. •

Note that, without further assumptions on the data, we cannot distinguish $\hat{\phi}$ from any of the other functions ϕ^θ with $\theta \in \Theta$. Hence, we can *only* conclude that a given optimality criterion of $\hat{\phi}$ holds if it holds for all functions ϕ^θ with $\theta \in \Theta$. Hence, a reasonable approach would be to optimize some

Preprint submitted to IEEE Open Journal of Control Systems. Received: May 9, 2025 08:56:43 Pacific Time

criterion for all ϕ^θ with $\theta \in \Theta$ *simultaneously*. However, changes in the parameter θ might lead to changes in the quantitative behavior and in particular the location of optimal values of the corresponding function ϕ^θ. This means that a small error in estimating the true parameter corresponding to the unknown function $\hat{\phi}$ might lead to significant error. This motivates us to focus on suboptimization problems instead. To be precise, we investigate whether, for a given $\delta \in \mathbb{R}$ and $c \in \mathbb{R}^m$, we can ensure that

$$\|\hat{\phi}(z)\|_2 \leqslant \delta \quad \text{or} \quad c^\top \hat{\phi}(z) \leqslant \delta. \tag{5}$$

We formalize this next.

Problem 1 (Cautious optimization)**.** Consider an unknown function $\hat{\phi} : \mathbb{R}^n \to \mathbb{R}^m$, a noise model Π such that Assumption 1 holds, and data (Y, Φ) of the true function:

(i) (Verification of suboptimality) given $z \in \mathbb{R}^n$, determine the minimal value of δ as a function of the data (Y, Φ), for which either inequality in (5) holds;

(ii) (One-shot cautious suboptimization) given a set $\mathcal{S} \subseteq \mathbb{R}^n$, find $z \in \mathcal{S}$ for which Problem 1(i) yields the minimal value of δ;

(iii) (Set-wise verification of suboptimality) given a set $\mathcal{S} \subseteq \mathbb{R}^n$, find the minimal δ as a function of the data (Y, Φ), for which either inequality in (5) holds for all $z \in \mathcal{S}$.

Given a set of basis functions and a noise model, being able to resolve any of these problems can be viewed as a property of the measurements. This is essentially the viewpoint taken with respect to data-driven control within the data informativity framework, e.g., [43, 45]: given $\delta \in \mathbb{R}$ and $\mathcal{S} \subseteq \mathbb{R}^n$, one could say that the data (Y, Φ) is *informative for δ-suboptimization on $\mathcal{S}$* if there exists $z \in \mathcal{S}$ such that $\|\hat{\phi}(z)\|_2 \leqslant \delta$.

We rely on the data informativity interpretation throughout our technical discussion in Section III to solve Problem 1. Then, in Section IV, we illustrate how to apply our results to solve various problems in data-driven analysis and control. First, we consider data-based contraction analysis for a class of nonlinear systems parameterized in terms of given basis functions. This essentially requires us to characterize one-sided Lipschitz constants for the unknown dynamics in terms of measured data. In terms of data-driven control, we consider the problem of data-based suboptimal regulation of unknown linear systems. Finally, in Section V we analyze online suboptimization, where we iteratively optimize and collect new measurements to find the optimizer.

III. Cautious optimization

This section describes our solutions to Problem 1. Recall that, based on the data we can perform set-valued regression to obtain the set Θ. In fact, we cannot distinguish $\hat{\phi}$ from any of the functions ϕ^θ as long as $\theta \in \Theta$. This means that we can only guarantee that $\|\hat{\phi}(z)\|_2 \leqslant \delta$ holds if $\sup_{\theta\in\Theta} \|\phi^\theta(z)\|_2 \leqslant \delta$. Similarly, given a vector $c \in \mathbb{R}^m$, we can only conclude that $c^\top \hat{\phi}(z) \leqslant \delta$ if $\sup_{\theta\in\Theta} c^\top \phi^\theta(z) \leqslant \delta$. This motivates the following definitions:

$$g(z;\Theta) := \sup_{\theta\in\Theta} \|\phi^\theta(z)\|_2, \quad g_c(z;\Theta) := \sup_{\theta\in\Theta} c^\top \phi^\theta(z). \tag{6}$$

These functions correspond to the elementwise worst-case realization of the unknown parameter $\hat{\theta}$. Therefore, we refer to $g(z;\Theta)$ as the *worst-case norm bound* and $g_c(z;\Theta)$ as the *worst-case linear bound*.

To draw conclusions regarding the true function $\hat{\phi}$, we investigate the functions $g(z;\Theta)$ and $g_c(z;\Theta)$. In fact, problems (i)-(iii) in Problem 1 can be recast as follows. Resolving Problem 1(i) (in Section A) boils down to finding function values of $g(z;\Theta)$ or $g_c(z;\Theta)$. Resolving Problem 1(ii) (in Section C) is equivalent to finding

$$\min_{z\in\mathcal{S}} g(z;\Theta) \quad \text{or} \quad \min_{z\in\mathcal{S}} g_c(z;\Theta). \tag{7}$$

Finally, resolving Problem 1(iii) (in Section D) amounts to finding the minimal δ for which

$$\max_{z\in\mathcal{S}} g(z;\Theta) \leqslant \delta \quad \text{or} \quad \max_{z\in\mathcal{S}} g_c(z;\Theta) \leqslant \delta. \tag{8}$$

Given the basis functions, each of the previous problems is completely specified in terms of the set Θ. In turn, this set is determined by the matrix N, containing information on the noise model and the measured data.

A. Pointwise verification of suboptimality

In order to resolve Problem 1(i), we are interested in obtaining values of the functions in (6). As a first observation, note that if the set Θ is compact (cf. Remark II.1), then the supremum is attained, and we can replace it with a maximum. In this case, the worst-case norm bound and worst-case linear bound are finite-valued functions.

We rely on the following characterization to obtain function values of $g(z;\Theta)$.

Lemma III.1. (Data-based conditions for bounds on the norm) *Given data (Y, Φ), let $\Theta = \mathcal{Z}(N)$, with N as in (4). Suppose that N has at least one positive eigenvalue. Let $\delta \geqslant 0$ and $z \in \mathbb{R}^n$. Then,*

$$\|\phi^\theta(z)\|_2 \leqslant \delta \quad \textit{for all} \quad \theta \in \Theta \tag{9}$$

if and only if there exists $\alpha \geqslant 0$ such that

$$\left[\begin{array}{cc:c} \delta^2 I_m & 0 & 0 \\ 0 & 0 & b(z) \\ \hdashline 0 & b(z)^\top & 1 \end{array}\right] - \alpha \left[\begin{array}{c:c} N & 0 \\ 0 & 0 \end{array}\right] \geqslant 0. \tag{10}$$

Proof:
Note that $\|\phi^\theta(z)\|_2^2 = b(z)^\top \theta\theta^\top b(z)$. This allows us to write (9) equivalently as

$$b(z)^\top \theta\theta^\top b(z) \leqslant \delta^2 \iff \theta^\top b(z) b(z)^\top \theta \leqslant \delta^2 I_m,$$

which can also be expressed as

$$\begin{bmatrix} I_m \\ \theta \end{bmatrix}^\top \begin{bmatrix} \delta^2 I_m & 0 \\ 0 & -b(z)b(z)^\top \end{bmatrix} \begin{bmatrix} I_m \\ \theta \end{bmatrix} \geqslant 0. \tag{11}$$

Thus (9) holds if and only if (11) holds for all $\theta \in \Theta$. Given that $\Theta = \mathcal{Z}(N)$, where N has at least one positive eigenvalue, the statement follows from applying [45, Thm 4.7] to (11) and using a Schur complement. ∎

The necessary and sufficient conditions of Lemma III.1 allow us to provide an alternative expression for $g(z;\Theta)$ that allows us to compute its value efficiently.

Corollary III.2. (Values of the worst-case norm bound) *Given data (Y, Φ), let $\Theta = \mathcal{Z}(N)$, with N as in (4). Suppose*

Preprint submitted to IEEE Open Journal of Control Systems. Received: May 9, 2025 08:56:43 Pacific Time

that N has at least one positive eigenvalue. Then, for $z \in \mathbb{R}^n$, we have

$$g(z;\Theta) = \min\{\delta \geqslant 0 \mid \exists \alpha \geqslant 0 \text{ s.t. (10) holds}\}. \quad (12)$$

Technically the statement (10) is not an LMI, since it depends quadratically on δ. However, function values of $g(z;\Theta)$ can be found efficiently via Corollary III.2 by minimizing δ^2 over (10) instead. Similar situations will arise without further mention in the remainder of the paper.

Remark III.3. (Weakening assumptions on data lead to conservative upper bound) If N does *not* have positive eigenvalues, i.e., $N \leqslant 0$, then the 'if'-side of Lemma III.1 remains true. Therefore, we can conclude that (12) in Corollary III.2 holds with '$\leqslant$' replacing '$=$'. This means that we obtain an upper bound for $g(z;\Theta)$, leading to a conservative upper bound of $\hat{\phi}(z)$. •

Moreover, we can use Corollary III.2 to find bounds independent of the choice of z.

Remark III.4. (Relaxations of the conditions for bounds on the norm) If M is such that $b(z)b(z)^\top \leqslant M$, then

$$\begin{bmatrix} \delta^2 I_m & 0 \\ 0 & -b(z)b(z)^\top \end{bmatrix} \geqslant \begin{bmatrix} \delta^2 I_m & 0 \\ 0 & -M \end{bmatrix}.$$

This implies that

$$\begin{bmatrix} \delta^2 I_m & 0 \\ 0 & -M \end{bmatrix} - \alpha N \geqslant 0 \quad (13)$$

implies (10). In turn, this means that for any M such that $b(z)b(z)^\top \leqslant M$ for all z, we have

$$g(z;\Theta) \leqslant \min\{\delta \geqslant 0 \mid \exists \alpha \geqslant 0 \text{ s.t. (13) holds}\},$$

which yields a bound independent from z. •

We rely on the following characterization to obtain function values of $g_c(z;\Theta)$.

Lemma III.5. (Data-based conditions for scalar upper bounds) *Given data (Y,Φ), let $\Theta = \mathcal{Z}(N)$, with N as in (4). Let $\delta \in \mathbb{R}$, $z \in \mathbb{R}^n$, and $c \in \mathbb{R}^m$. Suppose $c^\top(N|N_{22})c > 0$ and let*

$$N_c := \begin{bmatrix} c^\top N_{11} c & c^\top N_{12} \\ N_{21} c & N_{22} \end{bmatrix}. \quad (14)$$

Then it holds that

$$c^\top \phi^\theta(z) \leqslant \delta \quad \text{for all} \quad \theta \in \Theta \quad (15)$$

if and only if there exists $\alpha \geqslant 0$ such that

$$\begin{bmatrix} 2\delta & -b(z)^\top \\ -b(z) & 0 \end{bmatrix} - \alpha N_c \geqslant 0. \quad (16)$$

Proof:
We can rewrite $c^\top \phi^\theta(z)$ as $c^\top \theta^\top b(z) + b(z)^\top \theta c \leqslant 2\delta$. Thus, (15) holds if and only if

$$\gamma^\top b(z) + b(z)^\top \gamma \leqslant 2\delta \quad \text{for all} \quad \gamma \in \mathcal{Z}(N)c,$$

where $\mathcal{Z}(N)c := \{Zc \mid Z \in \mathcal{Z}(N)\}$. Since $c \neq 0$, by [45, Thm. 3.4], we have $\mathcal{Z}(N)c = \mathcal{Z}(N_c)$. Therefore, (15) is equivalent to:

$$\begin{bmatrix} 1 \\ \gamma \end{bmatrix}^\top \begin{bmatrix} 2\delta & -b(z)^\top \\ -b(z) & 0 \end{bmatrix} \begin{bmatrix} 1 \\ \gamma \end{bmatrix} \geqslant 0, \text{ for all } \gamma \in \mathcal{Z}(N_c). \quad (17)$$

Given that $N_{22} \leqslant 0$, we have that $c^\top(N|N_{22})c > 0$ if and only if N_c has a positive eigenvalue. Then, the statement follows from applying [45, Thm 4.7] to (17). ■

We leverage the characterization in Lemma III.5 to provide a closed-form expression of $g_c(z;\Theta)$ in terms of the data.

Theorem III.6. (Explicit bounds in the scalar case) *Given data (Y,Φ), let $\Theta = \mathcal{Z}(N)$, with N as in (4). Let $z \in \mathbb{R}^n$ and $c \in \mathbb{R}^m$. If $b(z) \in \operatorname{im}\Phi$ or $c = 0$, then*

$$g_c(z;\Theta) = -c^\top N_{12} N_{22}^\dagger b(z) + \sqrt{c^\top(N|N_{22})cb(z)^\top(-N_{22}^\dagger)b(z)},$$

and, if $c \neq 0$ and $b(z) \notin \operatorname{im}\Phi$, then $g_c(z;\Theta) = \infty$.

Proof:
Note that, if either $c = 0$ or $b(z) = 0$, then we have $g_c(z;\Theta) = 0$, thus the result follows. Assume then that $c \neq 0$, $b(z) \neq 0$, and $b(z) \in \operatorname{im}\Phi$. If $c^\top(N|N_{22})c = 0$, then $N_c \leqslant 0$. As such, $\gamma \in \mathcal{Z}(N_c)$ if and only if

$$N_c \begin{bmatrix} 1 \\ \gamma \end{bmatrix} = 0 \iff \gamma \in -N_{22}^\dagger N_{21} c + \ker N_{22}.$$

Using (17), we see that $c^\top \phi^\theta \leqslant \delta$ for all $\theta \in \Theta$ if and only if $\gamma^\top b(z) \leqslant \delta$ for all $\gamma \in \mathcal{Z}(N_c)$. Combining the previous, we see that this holds if and only if $\delta \geqslant -c^\top N_{12} N_{22}^\dagger b(z)$ (where we have used that $b(z) \in \operatorname{im}\Phi$). Minimizing δ reveals that the statement holds for this case.

Consider the case when $c^\top(N|N_{22})c > 0$. By Lemma III.5, (15) holds if and only if (16) holds for some $\alpha \geq 0$. Since $b(z) \neq 0$, it must be that $\alpha > 0$. Using the Schur complement on (16), we obtain

$$\begin{aligned} 0 \leqslant & 2\delta - \alpha c^\top N_{11} c + \alpha^{-1}(b(z)^\top + \alpha c^\top N_{12}) N_{22}^\dagger (b(z) + \alpha N_{21} c) \\ &= 2\delta + c^\top N_{12} N_{22}^\dagger b(z) + b(z)^\top N_{22}^\dagger N_{21} c \\ &\quad - \alpha c^\top (N|N_{22})c + \alpha^{-1} b(z)^\top N_{22}^\dagger b(z). \end{aligned}$$

Clearly, there exists α for which this holds if and only if it holds for the value of α that maximizes the expression on the right. Note that $\ker N_{22} = \ker \Phi^\top$. Thus, since $b(z) \in \operatorname{im}\Phi$ and $b(z) \neq 0$, we see that $b(z)^\top N_{22}^\dagger b(z) < 0$. The expression is then maximized over $\alpha \geqslant 0$ with

$$\alpha = \sqrt{\frac{-b(z)^\top N_{22}^\dagger b(z)}{c^\top(N|N_{22})c}}.$$

This yields that (16) holds if and only if

$$\begin{aligned} &2\delta + c^\top N_{12} N_{22}^\dagger b(z) + b(z)^\top N_{22}^\dagger N_{21} c \\ &\quad - 2\sqrt{c^\top(N|N_{22})cb(z)^\top(-N_{22}^\dagger)b(z)} \geqslant 0. \end{aligned}$$

To obtain $g_c(z;\Theta)$, we minimize this over δ, which yields the expression in the statement.

Finally, consider the case when $c \neq 0$ and $b(z) \notin \operatorname{im}\Phi$. It is straightforward to check that if $\theta \in \Theta$, then $\theta + vw^\top \in \Theta$, for all $w \in \mathbb{R}^m$ and $v \in \ker \Phi^\top \subseteq \mathbb{R}^k$. Since $b(z) \notin \operatorname{im}\Phi$ and $\operatorname{im}\Phi$ is orthogonal to $\ker \Phi^\top$, we can take v such that $v^\top b(z) > 0$. Now, for arbitrary $\beta \in \mathbb{R}$, take $w = \beta c$. Then,

$$\sup_{\theta \in \Theta} \phi^\theta(z) \geqslant c^\top(\theta + vw^\top)^\top b(z) = c^\top \theta^\top b(z) + \beta c^\top c v^\top b(z).$$

Since $c^\top c > 0$ and $v^\top b(z) > 0$, and this is valid for any $\beta \in \mathbb{R}$, the last part of the statement follows. ■

Remark III.7 (The special case of $N_{22} < 0$)**.** Note that when Θ is compact (or equivalently, Φ has full row rank, cf. Remark II.1), it is immediate that $b(z) \in \operatorname{im}\Phi$ for any $z \in \mathbb{R}^n$, and hence $g_c(z;\Theta)$ is always finite valued. •

Preprint submitted to IEEE Open Journal of Control Systems. Received: May 9, 2025 08:56:43 Pacific Time

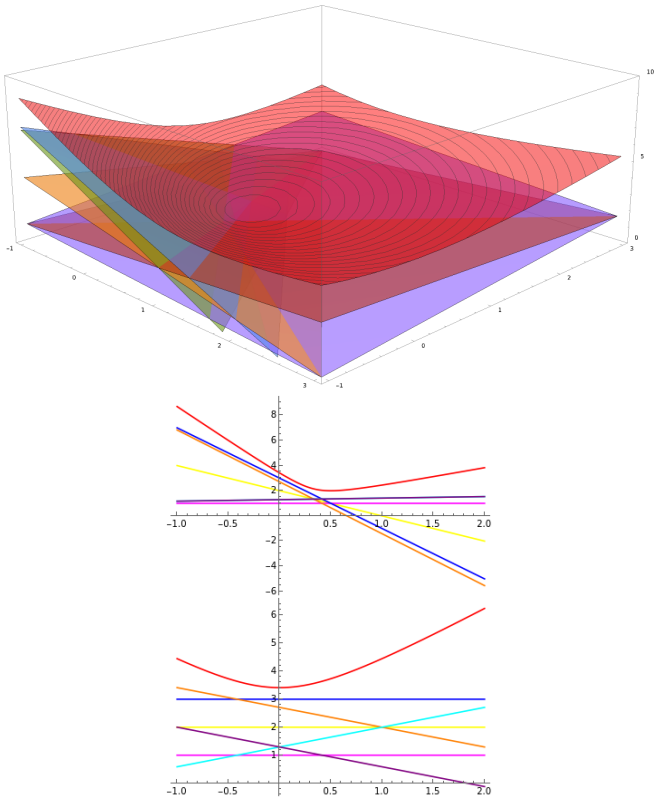

FIGURE 1: A plot depicting the situation of Example 1. For $c = \binom{1}{1}$, we plot a number function $c^\top \phi^\theta$, with θ consistent with the data. Any of these functions could be the unknown function. The least squares estimate is shown in yellow. If we are interested in bounds on $\hat{\phi}$, we can employ $g_c(z;\Theta)$ (shown in red). In particular note that this is not a linear map. Below the plot, we show the slices $z_1 = z_2$ and $z_1 = -z_2$ respectively.

The following illustrates the introduced concepts in the familiar setting of linear regression.

Example 1 (Linear regression)**.** Suppose that $\hat{\phi} : \mathbb{R}^2 \to \mathbb{R}^2$ and $k = 3$. We consider basis functions $\phi_1(z) = 1$ and $\phi_2(z) = \begin{pmatrix}1 & 0\end{pmatrix} z$, $\phi_3(z) = \begin{pmatrix}0 & 1\end{pmatrix} z$. In other words, we take $\theta \in \mathbb{R}^{3\times 2}$, and let $\phi^\theta(z) = \theta^\top \begin{bmatrix}1 & z^\top\end{bmatrix}^\top$. We collect three measurements,

$$\binom{1}{1} = \hat{\phi}(0) + w_1, \binom{0}{1} = \hat{\phi}(e_1) + w_2, \binom{1}{0} = \hat{\phi}(e_2) + w_3,$$

where we assume that $WW^\top \leqslant I_2$. This leads to $\Theta = \mathcal{Z}(N)$ with N such that

$$N_{11} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, N_{12} = \begin{bmatrix}2 & 0 & 1\\ 2 & 1 & 0\end{bmatrix}, N_{22} = \begin{bmatrix}-3 & -1 & -1\\ -1 & -1 & 0\\ -1 & 0 & -1\end{bmatrix}.$$

Now, the least-squares estimate is equal to

$$\theta^{\mathrm{lse}} = \begin{bmatrix}1 & 1\\ -1 & 0\\ 0 & -1\end{bmatrix}, \quad \phi^{\mathrm{lse}}(z;\Theta) = \binom{1-z_1}{1-z_2}.$$

In fact, using Theorem III.6 we can conclude that

$$g_c(z;\Theta) = c^\top \binom{1-z_1}{1-z_2} + \sqrt{c^\top c}\sqrt{(1-z_1-z_2)^2 + z_1^2 + z_2^2}.$$

As such, on the basis of the data we can guarantee that, for instance, $c^\top \hat{\phi}(0) \leqslant \begin{pmatrix}1 & 1\end{pmatrix} c + \|c\|_2$. •

B. Uncertainty on function values

The results in Section A allow us to find bounds for the unknown function $\hat{\phi}$, but do not consider how much these bounds deviate from its true value. For this, note that the result of Theorem III.6 can be interpreted as quantifying the distance between the true function value and the least-squares estimate ϕ^{lse} (defined in Remark II.2) in terms of the basis functions and the data. This observation leads us to define the *uncertainty* corresponding to (6) at z by

$$U(z;\Theta) := \sup_{\theta\in\Theta} \|\phi^\theta(z) - \phi^{\mathrm{lse}}(z;\Theta)\|_2, \tag{18a}$$

$$U_c(z;\Theta) := \sup_{\theta\in\Theta} c^\top \big(\phi^\theta(z) - \phi^{\mathrm{lse}}(z;\Theta)\big). \tag{18b}$$

Thus, if the uncertainty $U(z;\Theta)$ at z is small, then $\phi^{\mathrm{lse}}(z;\Theta)$ is quantifiably a good estimate of $\hat{\phi}(z)$.

The uncertainty also allows us describe a useful variant to Problem 1(ii). In order to balance the demands of a low upper bound on the value of the true function with an associated low uncertainty, it is reasonable to consider the following generalization of the cautious suboptimization problem in (7): for $\lambda \geqslant 0$, consider

$$\min_{z\in\mathcal{S}} g(z;\Theta) + \lambda U(z;\Theta), \tag{19a}$$

$$\min_{z\in\mathcal{S}} g_c(z;\Theta) + \lambda U_c(z;\Theta). \tag{19b}$$

The following result provides a closed-form expression for the objective function in (19a).

Lemma III.8. (Explicit expression of uncertainty in norms) *Given data (Y,Φ), let $\Theta = \mathcal{Z}(N)$, with N as in (4). For $z \in \mathbb{R}^n$, if $b(z) \in \operatorname{im}\Phi$, then*

$$U(z;\Theta) = \sqrt{\lambda_{\max}(N|N_{22})b(z)^\top(-N_{22}^\dagger)b(z)}. \tag{20}$$

If $b(z) \notin \operatorname{im}\Phi$, then $U(z;\Theta) = \infty$.

Proof:
First suppose that N has at least one positive eigenvalue. Similar to the proof of Lemma III.1, we can conclude that that $\|\phi^\theta(z) - \phi^{\mathrm{lse}}(z;\Theta)\|_2 \leqslant \varepsilon$ for $\varepsilon \geqslant 0$ and for all $\theta \in \Theta$ if and only if there exists $\alpha \geqslant 0$ such that

$$\begin{bmatrix}\varepsilon^2 I & 0\\ 0 & -b(z)b(z)^\top\end{bmatrix} - \alpha \begin{bmatrix}N|N_{22} & 0\\ 0 & N_{22}\end{bmatrix} \geqslant 0. \tag{21}$$

Thus, in line with Corollary III.2, we obtain

$$U(z;\Theta) = \min\{\varepsilon \geqslant 0 \mid \exists \alpha \geqslant 0 \text{ s.t. } (21)\}. \tag{22}$$

If $b(z) \in \operatorname{im}\Phi$ then we can minimize ε by taking $\alpha = b(z)^\top(-N_{22}^\dagger)b(z)$, leading to (20). If N does not have at least one positive eigenvalue, then $N|N_{22} = 0$. As in Remark III.3, we can conclude (22) with an inequality instead. Moreover, it immediately follows that $U(z;\Theta) \leqslant 0$, proving that (20) holds. The second part of the statement can be concluded in a similar fashion as the last part of the proof of Theorem III.6. ■

In addition to expressing the uncertainty, this result also gives rise to a useful explicit bound on the function $g(z;\Theta)$.

Corollary III.9. (Explicit upper bound of the worst-case norm bound) *Given data (Y,Φ), let $\Theta = \mathcal{Z}(N)$, with N as in (4). For $z \in \mathbb{R}^n$, if $b(z) \in \operatorname{im}\Phi$, then*

Preprint submitted to IEEE Open Journal of Control Systems. Received: May 9, 2025 08:56:43 Pacific Time

$$g(z;\Theta) \leqslant \|\phi^{\text{lse}}(z;\Theta)\|_2 + \sqrt{\lambda_{\max}(N|N_{22})b(z)^\top(-N_{22}^\dagger)b(z)}. \tag{23}$$

The following result provides a closed-form expression for the objective function in (19b).

Lemma III.10. (Explicit expression of scalar uncertainty) *Given data (Y,Φ), let $\Theta = \mathcal{Z}(N)$, with N as in (4). Let $z \in \mathbb{R}^n$ and $c \in \mathbb{R}^m$. If $b(z) \in \operatorname{im}\Phi$ or $c = 0$, then*

$$U_c(z;\Theta) = \sqrt{(c^\top(N|N_{22})c)(b(z)^\top(-N_{22}^\dagger)b(z))}.$$

If $b(z) \notin \operatorname{im}\Phi$ and $c \neq 0$, then $U_c(z;\Theta) = \infty$. Regardless, if $\lambda \geqslant 0$ and we define

$$N_\lambda := \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} + \begin{bmatrix} \lambda(2+\lambda)(N|N_{22}) & 0 \\ 0 & 0 \end{bmatrix},$$

then $g_c(z;\Theta) + \lambda U_c(z;\Theta) = g_c(z;\mathcal{Z}(N_\lambda))$.

Proof:
The first part follows immediately from Theorem III.6. Now note that

$$N_\lambda = \begin{bmatrix} (1+\lambda)^2(N|N_{22}) + N_{12}N_{22}^{-1}N_{21} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}.$$

Therefore $N_\lambda|N_{22} = (1+\lambda)^2(N|N_{22})$. Thus, the result follows from the application of Theorem III.6. ∎

Lemma III.10 means that, even though problem (19b) is more general than the corresponding cautious suboptimization problem in (7), both problems can be resolved in the same fashion.

C. Convexity and suboptimization

Solving the one-shot cautious suboptimization problem, cf. Problem 1(ii), in an efficient manner, amounts to minimizing the functions $g(\cdot;\Theta)$ or $g_c(\cdot;\Theta)$, respectively. For this, any standard optimization method can be used, which we illustrate below using simple gradient descent. We discuss how to determine values of the Jacobians of $g(\cdot;\Theta)$ or $g_c(\cdot;\Theta)$ in Appendix A, where we give a simple characterization in terms of properties of the basis and data, cf. Corollary A.2. Under suitable regularity conditions, these Jacobians can be used in a gradient descent scheme, which converges to a local minimum.

We can also resolve Problem 1(ii) *globally* in this manner if $g(\cdot;\Theta)$ or $g_c(\cdot;\Theta)$ are convex, respectively. To provide efficient tests for convexity, we will first assume that θc is elementwise nonnegative for all $\theta \in \Theta$. A test for this property can be found in the appendix, in Lemma A.1. Using this, we can identify the following conditions which ensure the convexity of $\|\hat{\phi}\|_2$ or $c^\top\hat{\phi}$ and their upper bounds.

Proposition III.11. (Convexity of the true function) *Given data (Y,Φ), let $\Theta = \mathcal{Z}(N)$, with N is as in (4). Assume the basis functions ϕ_i are convex. Let $c \neq 0$ with $\theta c \in \mathbb{R}^k_{\geqslant 0}$ for all $\theta \in \Theta$. Then,*

- *$c^\top\phi^\theta$ is convex for all $\theta \in \Theta$ and $g_c(\cdot;\Theta)$ is a finite-valued convex function;*
- *If, in addition, the functions ϕ_i are strictly convex and $0 \notin \mathcal{Z}(N_c)$, then $c^\top\phi^\theta$ is strictly convex for all $\theta \in \Theta$ and $g_c(\cdot;\Theta)$ is strictly convex.*

Moreover, $\|\phi^\theta(\cdot)\|_2$ is convex if $e_i^\top\phi^\theta$ is convex and nonnegative for all $i = 1,\dots m$.

Proof:
Lemma A.1 shows that, if $\theta c \in \mathbb{R}^k_{\geqslant 0}$ for all $\theta \in \Theta$, then Φ has full row rank or, equivalently, $N_{22} < 0$. This implies, cf. Remark II.1, that $\mathcal{Z}(N)$ is compact and therefore $g_c(\cdot;\Theta)$ is finite-valued. The first result now follows from the facts that nonnegative combinations of convex functions are convex and that the maximum of any number of convex functions is also convex. The last part follows from standard composition rules for convexity, see e.g., [7, Example 3.14]. ∎

The results of Lemma A.1 and Proposition III.11 taken together mean that, if the basis functions are convex, we can test for convexity on the basis of data. To guarantee strict convexity, Proposition III.11 requires that $0 \notin \mathcal{Z}(N_c)$, which is easy to check in terms of data.

Looking back at Example 1, we can make two interesting observations. First, the conditions of Proposition III.11 do *not* hold in the example, since for instance $\theta^{\text{lse}}c \in \mathbb{R}^3_{\geqslant 0}$ only if $c = 0$. Yet $c^\top\phi^\theta$ is convex for all $\theta \in \Theta$. This is because the basis functions ϕ_i are linear, and therefore the coefficients θ_i are not required to be nonnegative. Second, $g_c(\cdot;\Theta)$ in Example 1 is strictly convex for $c \neq 0$ even though the basis functions are not. Recall that we are interested in the optimization problem (7), and thus not necessarily in properties of the true function $\hat{\phi}$, but of the worst-case linear bound $g_c(\cdot;\Theta)$. These observations motivate our ensuing discussion to provide conditions for convexity of the upper bound instead. Under the assumptions of Theorem III.6:

$$g_c(z;\Theta) = c^\top\phi^{\text{lse}}(z;\Theta) + U_c(z;\Theta).$$

Thus, if (i) $c^\top\phi^{\text{lse}} = -c^\top N_{12}N_{22}^\dagger b(\cdot)$ is convex and (ii) U_c is convex, then so is g_c. Moreover, if in addition either is strictly convex, then so is g_c. Condition (i) can be checked directly if all basis functions ϕ_i are twice continuously differentiable by computing the Hessian of $c^\top\phi^{\text{lse}}$. Here, we present a simple criterion, also derived from composition rules for convexity (again, see e.g., [7, Example 3.14]), to test for condition (ii).

Corollary III.12. (Convexity of the uncertainties) *Given data (Y,Φ), let $\Theta = \mathcal{Z}(N)$ and assume $N_{22} < 0$. Then both $U(\cdot;\Theta)$ and $U_c(\cdot;\Theta)$ are convex if each basis function ϕ_i is convex and $-N_{22}^{-1}b(z) \geqslant 0$ for all $z \in \mathbb{R}^n$.*

In addition to guaranteeing that $g_c(\cdot;\Theta)$ is convex, this result can be used to find a convex upper bound when optimizing the norm, by bounding $g(\cdot;\Theta)$ using Corollary III.9.

Equipped with the results of this section, one can solve Problem 1(ii) efficiently using the following steps:

(i) Under the assumptions of Theorem III.6, we can write a closed-form expression for $g_c(\cdot;\Theta)$;
(ii) We can apply gradient descent using values of the Jacobian found with Corollary A.2, which, under suitable regularity conditions yields a local minimum;

Preprint submitted to IEEE Open Journal of Control Systems. Received: May 9, 2025 08:56:43 Pacific Time

(iii) We can test whether this closed form is (strictly) convex using e.g., Proposition III.11 or Corollary III.12, and, if so, conclude that the obtained minimum is global.

D. Set-wise verification of suboptimality

Here we solve Problem 1(iii) and provide upper (and lower) bounds as in (5) which are guaranteed to hold for all $z \in \mathcal{S} \subseteq \mathbb{R}^n$. In terms of g and g_c, this holds only if (8) holds.

To begin, we consider methods on the basis of convexity and concavity. First, we see that $g_c(z;\Theta)$ is concave if and only if $g_{-c}(\cdot;\Theta) = -g_c(\cdot;\Theta)$ is convex. Moreover,

$$\max_{z\in\mathcal{S}} g_c(z;\Theta) = -\min_{z\in\mathcal{S}} g_{-c}(z;\Theta).$$

Therefore, we can test for concavity of g_c and apply the minimization results to g_{-c} of Section C. Beyond the case of concavity, we can also employ convexity in order to efficiently provide upper bounds over the convex hull of a finite set, as the following result shows.

Proposition III.13. (Convexity and maximal values) *Given data (Y,Φ), let $\Theta = \mathcal{Z}(N)$. Let $\mathcal{F} \subseteq \mathbb{R}^n$ be a finite set and let $\mathcal{S} = \operatorname{conv}\mathcal{F}$. Suppose that $g(z;\Theta)$ is convex. Then (9) for all $z \in \mathcal{S}$ if and only if $g(z;\Theta) \leqslant \delta \quad \forall z \in \mathcal{F}$. Similarly, if $g_c(z;\Theta)$ is convex, then*

$$\|\phi^\theta(z)\|_2 \leqslant \delta \quad \text{for all } \theta \in \Theta \text{ and all } z \in \mathcal{S}$$

if and only if $g_c(z;\Theta) \leqslant \delta \quad \forall z \in \mathcal{F}$.

The proof of this result follows immediately from the fact that any convex function attains its maximum over $\mathcal{S}$ on $\mathcal{F}$. Thus, convexity (resp., concavity) allows us to find minimal upper bounds (resp., maximal lower bounds) over a set, leading to a solution of Problem 1.(iii).

To solve Problem 1(iii) in scenarios where convexity or concavity cannot be guaranteed, we employ arguments based on Lipschitz continuity and coverings. We are thus interested in the question of whether, for all $z, z^\star \in \mathcal{S} \subseteq \mathbb{R}^n$,

$$\|\hat{\phi}(z) - \hat{\phi}(z^\star)\|_2 \leqslant L\|z - z^\star\|_2,$$

holds. We give a characterization of Lipschitz constants in terms of data in Appendix A, cf. Lemma A.3. In addition, we say that set $\mathcal{G} \subset \mathbb{R}^n$ is an *ε-covering* of $\mathcal{S} \subseteq \mathbb{R}^n$ if, for every $z \in \mathcal{S}$, there exists $z^\star \in \mathcal{G}$ such that $\|z - z^\star\|_2 \leqslant \varepsilon$. If $\mathcal{S}$ is a bounded set, then for any $\varepsilon > 0$, there exists an ε-covering with finitely many elements.

Having access to a Lipschitz constant and an ε-covering allows us to bound the unknown function $\hat{\phi}$ as follows.

Theorem III.14. (Coverings and bounds) *Given data (Y,Φ), let $\Theta = \mathcal{Z}(N)$. For $\varepsilon > 0$, let $\mathcal{G}$ be a finite ε-covering of $\mathcal{S} \subseteq \mathbb{R}^n$. Suppose that $g(z;\Theta)$ is Lipschitz with constant L on $\mathcal{S}$. Then, for all $z \in \mathcal{S}$,*

$$\min_{z^\star\in\mathcal{G}} g(z^\star;\Theta) - \varepsilon L \leqslant g(z;\Theta) \leqslant \max_{z^\star\in\mathcal{G}} g(z^\star;\Theta) + \varepsilon L.$$

If $g_c(z;\Theta)$ is Lipschitz with constant L, then for all $z \in \mathcal{S}$,

$$\min_{z^\star\in\mathcal{G}} g_c(z^\star;\Theta) - \varepsilon L \leqslant g_c(z;\Theta) \leqslant \max_{z^\star\in\mathcal{G}} g_c(z^\star;\Theta) + \varepsilon L.$$

We omit the proof, which follows directly from combining the definitions of Lipschitz continuity and $\varepsilon-$coverings.

This result means that we can find guaranteed upper and lower bounds of either $\|\hat{\phi}(z)\|$ or $c^\top\hat{\phi}(z)$ over the bounded set $\mathcal{S}$ in terms of a finite number of evaluations of $g(z;\Theta)$ or $g_c(z;\Theta)$, resp. In turn, recall that Corollary III.2 allows us to efficiently find function values of $g(z;\Theta)$ on the basis of measurements and, similarly, Theorem III.6 allows us to directly calculate values of $g_c(z;\Theta)$.

IV. Applications to system analysis and control

In this section, we exploit the proposed solutions to Problem 1 in two applications: contraction analysis of unknown nonlinear systems and regulation of an unknown linear system to a suboptimal point of an unknown cost function.

A. Data-based contraction analysis for nonlinear systems

Consider the autonomous discrete- or continuous-time system given by either

$$z_{k+1} = \hat{\phi}(z_k), \quad \text{or} \quad \dot{z}(t) = \hat{\phi}(z(t)), \tag{24}$$

where $\hat{\phi} : \mathbb{R}^n \to \mathbb{R}^n$ is an unknown function. We consider noisy measurements (2) as described in Section II. Performing set-valued regression with this data, we obtain $\hat{\phi} = \phi^\theta$ for some $\theta \in \Theta = \mathcal{Z}(N)$. In the continuous-time case, this means that we collect noisy measurements of the derivative of the state at a finite set of states. In most applications, these derivative measurements need to be determined from collected state measurements (see e.g. [**?**] for a discussion on this assumption).

The discrete-time system in (24) is *strongly contracting* (cf. [10, Sec. 3.4]) with respect to the weighted norm $\|\cdot\|_{2,P^{1/2}}$ if $\hat{\phi}$ admits a Lipschitz constant $L < 1$, that is,

$$\|\hat{\phi}(z) - \hat{\phi}(z^\star)\|_{2,P^{1/2}} \leqslant L\|z - z^\star\|_{2,P^{1/2}}.$$

One can now directly employ Lemma A.3 in order to obtain data-based conditions under which this property holds.

To deal with the situation of continuous-time systems, we first require a number of prerequisites. Given $P > 0$, a function $f : \mathbb{R}^n \to \mathbb{R}^n$ is *one-sided Lipschitz with respect to* $\|\cdot\|_{2,P^{1/2}}$ if there exists $\gamma \in \mathbb{R}$ such that, for all $z, z^\star \in \mathbb{R}^n$,

$$(z - z^\star)^\top P(f(z) - f(z^\star)) \leqslant \gamma\|z - z^\star\|^2_{2,P^{1/2}}. \tag{25}$$

Such γ is called a *one-sided Lipschitz constant* and the smallest such γ is denoted $\operatorname{osLip}(f)$. The autonomous system $\dot{z}(t) = f(z(t))$ is *strictly contracting* with rate $|b|$ if $\operatorname{osLip}(f) \leqslant b < 0$. Given any two trajectories $x(t), \bar{x}(t)$ of a strictly contracting system, one has $\|x(t) - \bar{x}(t)\|_{2,P^{1/2}} \leqslant e^{b(t-s)}\|x(s) - \bar{x}(s)\|_{2,P^{1/2}}$ for any $t \geqslant s \geqslant 0$.

The following result, derived from Theorem III.6, establishes a test for strict contractivity of the continuous-time system in (24) on the basis of a set of measurements of $\hat{\phi}$.

Theorem IV.1. (Data-based test for strict contractivity) *Given data (Y,Φ), let $\Theta = \mathcal{Z}(N)$, with N as in (4). Let $P > 0$ and $z, z^\star \in \mathbb{R}^n$. Assume Φ has full row rank. Then,*

$$(z - z^\star)^\top P\theta^\top(b(z) - b(z^\star)) \leqslant \gamma\|z - z^\star\|^2_{2,P^{1/2}}$$

for all $\theta \in \Theta$ if and only if

Preprint submitted to IEEE Open Journal of Control Systems. Received: May 9, 2025 08:56:43 Pacific Time

$$(z-z^\star)^\top P(-N_{12}N_{22}^{-1})(b(z)-b(z^\star)) + \sqrt{(z-z^\star)^\top P(N|N_{22})P(z-z^\star)} \cdot \sqrt{(b(z)-b(z^\star))^\top(-N_{22}^{-1})(b(z)-b(z^\star))} \leqslant \gamma\|z-z^\star\|^2_{2,P^{1/2}}.$$

As a consequence of this result, we can provide a test to establish strict contractivity in terms only of the least-squares estimate ϕ^{lse}, the data matrix N, and a Lipschitz constant for the basis functions.

Corollary IV.2. (Strict contractivity in terms of the least-squares estimate) *Given data (Y,Φ), let $\Theta = \mathcal{Z}(N)$, with N as in (4). Assume Φ has full row rank. Let L be such that $\|b(z)-b(z^\star)\|_{2,(-N_{22})^{-1/2}} \leqslant L\|z-z^\star\|_2$, for all $z, z^\star \in \mathbb{R}^n$. Then, for $P \succ 0$,*

$$(z-z^\star)^\top P\theta^\top(b(z)-b(z^\star)) \leqslant \gamma\|z-z^\star\|^2_{2,P^{1/2}}$$

for all $\theta \in \Theta$ and $z, z^\star \in \mathbb{R}^n$ if

$$\operatorname{osLip}(\phi^{\text{lse}}) < \gamma - L\sqrt{\lambda_{\max}(N|N_{22})}\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}. \tag{26}$$

Remark IV.3. (Contraction w.r.t. different norms) The results of this section can be readily generalized to contractivity (and thus one-sided Lipschitzness) with respect to $\|\cdot\|_{p,R}$ for $p \in [1,\infty)$. To do this, one replaces the one-sided Lipschitz condition (25) with the respective condition from [14, Table I] and adjust the statements accordingly. Note that the case of $\|\cdot\|_{\infty,R}$ is not amenable to this treatment because the corresponding one-sided Lipschitz condition $\operatorname{osLip}(f)$ is not linear in f. •

B. Suboptimal regulation of unknown systems

Consider the problem of regulating an unknown linear system to a suboptimal point of the norm of an unknown cost function $\hat{\phi} : \mathbb{R}^n \to \mathbb{R}^m$ on the basis of measurements. We consider noisy measurements (2) as described in Section II. In short, $\hat{\phi} = \phi^\theta$ for some $\theta \in \Theta = \mathcal{Z}(N)$. Moreover, let

$$x_{k+1} = \hat{A}x_k + \hat{B}u_k, \tag{27}$$

with state $x_k \in \mathbb{R}^n$ and input $u_k \in \mathbb{R}^r$. Here, $\hat{A} \in \mathbb{R}^{n\times n}$ and $\hat{B} \in \mathbb{R}^{n\times r}$. We are interested in regulating the system (27) to an equilibrium $x^\star$ for which we can guarantee that $\|\hat{\phi}(x^\star)\|_2 \leqslant \delta$, with a value of δ as small as possible.

If $\hat{A}$ and $\hat{B}$ are known and $\hat{A}$ is stable, any fixed point (i.e., equilibrium) of the dynamics (27) corresponds to a constant input $u^\star$. Indeed, if we apply the constant input $u_k = u^\star$, the state asymptotically converges to the fixed point

$$\lim_{k\to\infty} x_k = x^\star := (I-\hat{A})^{-1}\hat{B}u^\star,$$

regardless of the initial condition x_0. This means that we can find a static input to regulate the system towards any state in $\operatorname{im}(I-\hat{A})^{-1}\hat{B} \subseteq \mathbb{R}^n$. Hence, if the matrices $\hat{A}$ and $\hat{B}$ are known, the problem corresponds to Problem 1(ii) with $\mathcal{S} = \operatorname{im}(I-\hat{A})^{-1}\hat{B}$.

In the following, we assume that we do *not* have access to the matrices $\hat{A}$ and $\hat{B}$. Instead, we assume that we have not only measurements of $\hat{\phi}$, but additionally measurements corresponding to the unknown system. To be precise, we collect measurements of the states and input of the system

$$x_{k+1} = \hat{A}x_k + \hat{B}u_k + w_k. \tag{28}$$

Using the basis functions

$$\psi_i(x,u) = \begin{cases} x_i & \text{for } i = 1,\dots,n \\ u_i & \text{for } i = n+1,\dots n+r, \end{cases}$$

we can perform a linear version of set-valued regression (cf. Section II) to obtain a set Σ such that $(\hat{A},\hat{B})^\top \in \Sigma = \mathcal{Z}(M)$, where $M \in \mathbb{S}^{n+(n+r)}$ is defined analogously to N in (4). In fact, this is equivalent to the setup of e.g., [45].

In line with all our previous reasoning, we can formalize the problem:

Problem 2 (Suboptimal regulation)**.** *Consider an unknown function $\hat{\phi} = \phi^\theta$ for some $\theta \in \Theta = \mathcal{Z}(N)$ and an unknown linear system $(A,B) \in \Sigma = \mathcal{Z}(M)$, where both corresponding noise models satisfy Assumption 1. Find an input $u^\star$ such that each system $(A,B) \in \Sigma$ converges to a (potentially different) equilibrium which is a suboptimal point for each function ϕ^θ with $\theta \in \Theta = \mathcal{Z}(N)$.*

Our approach to this problem takes the following steps:

- First, we provide conditions for the unknown system matrix to be stable;
- Next, we provide bounds on the set of all equilibria resulting from the application of a given input to the set of systems Σ;
- Lastly, we leverage structure of these results to solve the suboptimal regulation problem.

To be able to regulate the unknown system in the same manner as before, recall that we require that $\hat{A}$ is stable. Given that, on the basis of measurements we can not distinguish $\hat{A}$ from other matrices A such that $(A,B)^\top \in \Sigma$, we therefore require all such A to be stable. In the following, we describe a test for the existence of a shared quadratic Lyapunov function, which is a stronger stability condition. For a detailed discussion on the conservativeness of this assumption, see e.g., [44]. Now we can easily adapt e.g., [45, Thm. 5.1] to obtain the following.

Lemma IV.4 (Informativity for stability)**.** *Let $\Sigma = \mathcal{Z}(M)$, where the corresponding noise model satisfies Assumption 1. Then, there exists $P \in \mathbb{S}^n$ with $P \succ 0$ and $APA^\top < P$ for all $(A,B)^\top \in \Sigma$ if and only if there exist $\bar{P} \in \mathbb{S}^n$ with $\bar{P} \succ 0$ and $\beta > 0$ such that*

$$\begin{bmatrix} \bar{P}-\beta I_n & 0 & 0 \\ 0 & -\bar{P} & 0 \\ 0 & 0 & 0 \end{bmatrix} - M \succcurlyeq 0. \tag{29}$$

Remark IV.5. (Stability and contractivity) Applying a fixed input $u^\star$ to the unknown system yields an autonomous discrete-time system. We know that for any initial condition this converges to a fixed point. Moreover, if $P \succ 0$ and $APA^\top < P$ for all $(A,B)^\top \in \Sigma$, then $A^\top P^{-1}A < P^{-1}$. Therefore, the system $x_{k+1} = \hat{A}x_k + \hat{B}u^\star$ is strongly contracting with respect to $\|\cdot\|_{2,P^{-1/2}}$. •

Preprint submitted to IEEE Open Journal of Control Systems. Received: May 9, 2025 08:56:43 Pacific Time

When each system matrix A in the set Σ of systems consistent with the data is stable, we can characterize the fixed points resulting from applying the same input to each.

Lemma IV.6 (Characterizing fixed points)**.** *Let $\Sigma = \mathcal{Z}(M)$, where the corresponding noise model satisfies Assumption 1 and M has at least one positive eigenvalue. Assume A is stable for all $(A,B)^\top \in \Sigma$. Given $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^r$, $\|x-(I-A)^{-1}Bu\|_2 \leqslant \varepsilon$ for all $(A,B)^\top \in \Sigma$ if and only if there exists $\gamma \geqslant 0$ such that*

$$\left[\begin{array}{ccc:c} I_n & -I_n & 0 & x \\ -I_n & I_n & 0 & -x \\ 0 & 0 & 0 & -u \\ \hdashline x^\top & -x^\top & -u^\top & \varepsilon^2 \end{array}\right] - \gamma \left[\begin{array}{c:c} M & 0 \\ \hdashline 0 & 0 \end{array}\right] \geqslant 0 \tag{30}$$

Proof:
Let $(A,B)^\top \in \Sigma$. First, note that we have

$$\begin{aligned} &\|x-(I_n-A)^{-1}Bu\|_2 \leqslant \varepsilon \\ \Leftrightarrow\ &(x-(I_n-A)^{-1}Bu)(x-(I_n-A)^{-1}Bu)^\top \leqslant \varepsilon^2 I. \end{aligned}$$

Since A is stable, we know that $I-A$ is nonsingular. Thus, the previous holds if and only if

$$((I-A)x-Bu)((I-A)x-Bu)^\top \leqslant \varepsilon^2 (I-A)(I-A)^\top,$$

or equivalently,

$$\begin{bmatrix} I_n \\ A^\top \\ B^\top \end{bmatrix}^\top \left(\begin{bmatrix} \varepsilon^2 I_n & -\varepsilon^2 I_n & 0 \\ -\varepsilon^2 I_n & \varepsilon^2 I_n & 0 \\ 0 & 0 & 0 \end{bmatrix} - \begin{pmatrix} x \\ -x \\ -u \end{pmatrix} \begin{pmatrix} x \\ -x \\ -u \end{pmatrix}^\top \right) \begin{bmatrix} I_n \\ A^\top \\ B^\top \end{bmatrix} \geqslant 0.$$

The set of (A,B) that satisfy the condition can be written as the solution set of a QMI. By assumption $\Sigma = \mathcal{Z}(M)$. We can then apply the matrix S-Lemma [45, Thm 4.7] and use a Schur complement to prove the statement. ■

Motivated by Lemma IV.6, we define

$$\varepsilon^-(x) := \min\{\varepsilon \mid \exists \gamma \geqslant 0, u \in \mathbb{R}^r \text{ s.t. } (30) \text{ holds}\}. \tag{31}$$

Given $x \in \mathbb{R}^N$, the function $\varepsilon^-(x)$ thus gives the minimal radius of a ball around x, for which there exists a $u^\star$ such that the corresponding set of fixed points of each system in Σ is contained in this ball. The next result describes useful properties of this function.

Lemma IV.7. (Properties of the minimal radius function) *Let $\Sigma = \mathcal{Z}(M)$, where the corresponding noise model satisfies Assumption 1 and M has at least one positive eigenvalue. Assume A is stable for all $(A,B)^\top \in \Sigma$. Then $\varepsilon^-(x) \leqslant \|x\|_2$ for any $x \in \mathbb{R}^n$ and ε^- is convex.*

Proof:
The statement $\varepsilon^-(x) \leqslant \|x\|_2$ follows by noting that, for any x and M, (30) holds with $u=0$, $\gamma=0$ and $\varepsilon = \|x\|_2$. Next, suppose that for all $(A,B)^\top \in \Sigma$,

$$\|x-(I-A)^{-1}Bu\|_2 \leqslant \varepsilon, \quad \|\bar{x}-(I-A)^{-1}B\bar{u}\|_2 \leqslant \bar{\varepsilon}.$$

Let $\lambda \in [0,1]$ and define $x_\lambda := \lambda x + (1-\lambda)\bar{x}$ and similarly for u_λ and ε_λ. Then,

$$\begin{aligned} &\|x_\lambda - (I-A)^{-1}Bu_\lambda\|_2 \leqslant \lambda \|x-(I-A)^{-1}Bu\|_2 \\ &+ (1-\lambda)\|\bar{x}-(I-A)^{-1}B\bar{u}\|_2 \leqslant \varepsilon_\lambda. \end{aligned}$$

Therefore, $\varepsilon^-(x_\lambda) \leqslant \varepsilon_\lambda$, proving that ε^- is convex in x. ■

The pieces are now in place for our solution to Problem 2 in the following result.

Theorem IV.8. (Suboptimal regulation) *Let $\Sigma = \mathcal{Z}(M)$ and $\Theta = \mathcal{Z}(N)$, where the corresponding noise models satisfy Assumption 1. Assume A is stable for all $(A,B)^\top \in \Sigma$ and ϕ^θ is Lipschitz with constant L for all $\theta \in \Theta$. Then*

$$\min_u \|\hat{\phi}((I-\hat{A})^{-1}\hat{B}u)\|_2 \leqslant \min_x \ \big(g(x;\Theta) + L\varepsilon^-(x)\big). \tag{32}$$

Proof:
Using the triangle inequality and Lipschitzness,

$$\|\hat{\phi}((I-\hat{A})^{-1}\hat{B}u)\|_2 \leqslant \|\hat{\phi}(x)\|_2 + L\|x-(I-\hat{A})^{-1}\hat{B}u\|_2,$$

for any $x \in \mathbb{R}^n$. Note that this separates the unknown parameter $\hat{\theta}$ from the unknown pair $(\hat{A},\hat{B})$. From here, we deduce

$$\begin{aligned} &\min_u \|\hat{\phi}((I-\hat{A})^{-1}\hat{B}u)\|_2 \\ &\leqslant \min_{x,u} \Big(\max_{\theta\in\Theta} \|\phi^\theta(x)\| + L\Big(\max_{(A,B)\in\Sigma} \|x-(I-A)^{-1}Bu\| \Big) \Big). \end{aligned}$$

Since only the second term is dependent on u, we can use the definition of $\varepsilon^-(x)$ and Lemma IV.6 to conclude the statement. ■

Similar to Problem 1(ii), we can now resolve the optimization problem on the right hand side under various regularity conditions. To illustrate this, recall that ε^- is convex, and thus we can resolve this problem efficiently if so is $g(\cdot;\Theta)$ (see Corollary III.12). We conclude this section by noting that Theorem IV.8 requires only stability of each system matrix A. Under the stronger assumption of quadratic stability, which can be determined using Lemma IV.4, we can provide a transient guarantee on the values of the unknown function.

Theorem IV.9. (Transient values of the cost function) *Let $\Sigma = \mathcal{Z}(M)$ and $\Theta = \mathcal{Z}(N)$, where the corresponding noise models satisfy Assumption 1. Let $P > 0$ such that $APA^\top < P$ for all $(A,B)^\top \in \Sigma$ and assume ϕ^θ is Lipschitz with constant L for all $\theta \in \Theta$. Let $\delta > 0$ and $u^\star$ be a fixed input such that $\|\hat{\phi}((I-\hat{A})^{-1}\hat{B}u^\star)\|_2 \leqslant \delta$. Given an initial condition $\hat{x}_0 \in \mathbb{R}^n$, let $\hat{x}_k$ denote the trajectory of the unknown system $(\hat{A},\hat{B})$ corresponding to the input $u^\star$ and starting from $\hat{x}_0$. Then, there exists $\lambda \in [0,1)$ such that*

$$\|\hat{\phi}(\hat{x}_k)\|_2 \leqslant \delta + \lambda^k L \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} \|\hat{x}_0 - x^\star\|_2.$$

Proof:
Let $x^\star$ be the fixed point corresponding to the input $u^\star$ for the system $(\hat{A},\hat{B})$. If $APA^\top < P$ for all $(A,B)^\top \in \Sigma$, then the set $\{A \mid (A,B)^\top \in \Sigma\}$ is bounded. Since the set is closed by definition, we can conclude that $APA^\top < P$ for all $(A,B)^\top \in \Sigma$ if and only if there exists $\lambda \in [0,1)$ such that $APA^\top < \lambda P$ for all $(A,B)^\top \in \Sigma$. In particular, this implies that the *affine* system resulting from the application of input $u^\star$ to $(\hat{A},\hat{B})$ is strongly contracting (cf. Remark IV.5). Thus, for any $k \geqslant 1$,

$$\|\hat{x}_k - x^\star\|_{2,P^{-1/2}} \leqslant \lambda \|\hat{x}_{k-1} - x^\star\|_{2,P^{-1/2}}.$$

By applying this k times repeatedly, we can conclude that $\|\hat{x}_k - x^\star\|_{2,P^{-1/2}} \leqslant \lambda^k \|\hat{x}_0 - x^\star\|_{2,P^{-1/2}}$, and thus

$$\|\hat{x}_k - x^\star\|_2 \leqslant \lambda^k \frac{\lambda_{\max}(P^{-1})}{\lambda_{\min}(P^{-1})} \|\hat{x}_0 - x^\star\|_2.$$

Preprint submitted to IEEE Open Journal of Control Systems. Received: May 9, 2025 08:56:43 Pacific Time

Using the triangle inequality and the fact that $\hat{\phi}$ is Lipschitz with constant L,

$$\|\hat{\phi}(\hat{x}_k)\|_2 \leqslant \|\hat{\phi}(x^\star)\|_2 + L\|\hat{x}_k - x^\star\|_2.$$

Combining this previous with the fact that $\lambda_{\max}(P^{-1}) = 1/\lambda_{\min}(P)$ and $\lambda_{\min}(P^{-1}) = 1/\lambda_{\max}(P)$ yields

$$\|\hat{\phi}(\hat{x}_k)\|_2 \leqslant \delta + \lambda^k L \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}\|\hat{x}_0 - x^\star\|_2.$$

This proves the statement. ■

Remark IV.10. (Suboptimal regulation of unstable systems) Our discussion above assumes that the unknown system and all of those consistent with the measurements are stable. When this is not the case, we can instead proceed by first testing if the data is sufficiently informative for *quadratic stabilization*, i.e., to guarantee the existence of a static feedback K such that $(A+BK)P(A+BK)^\top < P$ for all $(A,B)^\top \in \Sigma$, cf [45, Thm. 5.1]. Then, we can characterize the behavior of the fixed points of the resulting systems by considering

$$\|x - (I - A + BK)^{-1}Bu\|_2 \leqslant \varepsilon.$$

Adapting Lemma IV.6 and the definition of ε^- accordingly is then straightforward. This gives rise to a two-step approach: (i) determine a gain K which stabilizes all systems in Σ and (ii) minimize $g(x;\Theta) + L\varepsilon^-(x)$. Note that the choice of K among potentially many stabilizing gains influences the resulting ε^- in a nontrivial way. We leave the analysis of jointly minimizing $g(x;\Theta) + L\varepsilon^-(x)$ over all x and all possible choices K as an open problem. •

V. Online cautious optimization

Our exposition so far has considered one-shot optimization settings where, given a set of measurements, we determine regularity properties and address suboptimization of the unknown function. In this section, we consider *online* scenarios, where one can repeatedly combine optimization of the unknown function with the collection of additional measurements, aiming at reducing the optimality gap. Consistent with our exposition, this means that we are interested in finding guaranteed upper bounds on $c^\top\hat{\phi}$ on the basis of initial measurements, then refining this bound on the basis of newly collected measurements[1]. We aim at doing this in a monotonic fashion, that is, in a way that has the obtained upper bounds do not increase as time progresses.

Consider repeated measurements of the form

$$Y_k = \hat{\theta}^\top\Phi_k + W_k, \text{ with } W_k^\top \in \mathcal{Z}(\Pi), \tag{33}$$

where Y_k and W_k are as in (2) and $k \in \mathbb{N}$. Let $\Theta_k := \mathcal{Z}(N_k)$ be the set of of parameters which are compatible with the k^{th} set of measurements, where

$$N_k := \begin{bmatrix} I & Y_k \\ 0 & -\Phi_k \end{bmatrix}\Pi\begin{bmatrix} I & Y_k \\ 0 & -\Phi_k \end{bmatrix}^\top.$$

Our online optimization strategy is described as follows.

Algorithm 1. (Online cautious optimization) *Consider an initial candidate optimizer $z_0 \in \mathbb{R}^n$, an initial set of data (Y_0, Φ_0), and $c \in \mathbb{R}^m$. Define the initial set of compatible parameters as $\bar{\Theta}_0 = \mathcal{Z}(N_0)$, where N_0 is given as in (4). For $k \geqslant 1$, we alternate two steps:*

(i) Employ the set $\bar{\Theta}_{k-1}$ to improve the candidate optimizer by finding z_k such that

$$g_c(z_k;\bar{\Theta}_{k-1}) \leqslant g_c(z_{k-1};\bar{\Theta}_{k-1}). \tag{34}$$

(ii) Measure the function as in (33) and update the set of parameters by finding a set $\bar{\Theta}_k$ such that $\hat{\theta} \in \bar{\Theta}_k$ and

$$g_c(z_k;\bar{\Theta}_k) \leqslant g_c(z_k;\bar{\Theta}_{k-1}). \tag{35}$$

In the following, we enforce (35) by considering parameters consistent with *all* previous measurements,

$$\bar{\Theta}_k := \Theta_0 \cap \ldots \cap \Theta_k. \tag{36}$$

Note that Algorithm 1 provides a sequence of upper bounds to true function values, on the basis of measurements. This means that after *any* number of iterations, we obtain a worst-case estimate of the function value $c^\top\hat{\phi}(z_k)$ and, as such, for the minimum of $c^\top\hat{\phi}$. To be precise, for each $k \geqslant 1$,

$$\min_{z\in\mathbb{R}^n} c^\top\hat{\phi}(z) \leqslant c^\top\hat{\phi}(z_k) \leqslant g_c(z_k;\bar{\Theta}_{k-1}), \tag{37}$$

and these bounds are monotonically nonincreasing

$$g_c(z_{k+1};\bar{\Theta}_k) \leqslant g_c(z_k;\bar{\Theta}_{k-1}) \tag{38}$$

Therefore this sequence of upper bounds is nonincreasing and bounded below, and we can conclude that the algorithm converges. However, without further assumptions, one cannot guarantee convergence of the upper bounds to the minimal value of $c^\top\hat{\phi}$, or respectively of z_k to the global minimum of $c^\top\hat{\phi}$. This is because, even though when repeatedly collecting measurements, it is reasonable to assume that the set of consistent parameters would decrease in size, this is not necessarily the case in general. In particular, a situation might arise where repeated measurements corresponding to a worst-case, or adversarial, noise signal give rise to convergence to a fixed bound with nonzero uncertainty. This is the point we address next by considering random, stochastic noise realizations.

A. Collection of uniformly distributed measurements

Recall that, so far we have assumed that the noise, as collected in the matrix W, satisfies Assumption 1, i.e., we have access to $\Pi \in \mathbb{S}^{m+T}$ with $\Pi_{22} < 0$ and $\Pi|\Pi_{22} \geqslant 0$, such that $W^\top \in \mathcal{Z}(\Pi)$. In particular, this implies that $\mathcal{Z}(\Pi)$ is bounded. In this section, we consider a scenario where the set $\mathcal{Z}(\Pi)$ has nonempty interior (equivalent to $\Pi|\Pi_{22} > 0$, [45, Thm. 3.2]) and the noise samples are not only bounded but *distributed uniformly randomly* over the set $\mathcal{Z}(\Pi)$. To formalize this, consider a measure μ on $\mathbb{R}^{m\times T}$, and define a probability density function p by

$$p(W) := \begin{cases} \frac{1}{\mu(\mathcal{Z}(\Pi))} & W^\top \in \mathcal{Z}(\Pi), \\ 0 & \text{otherwise.} \end{cases}$$

As a notational shorthand, we write $W^\top \sim \text{Uni}(\mathcal{Z}(\Pi))$ if the distribution of W follows the probability density function p. The following result shows that, under such uniformly

[1] To keep the presentation contained, we do not consider minimization of $\|\hat{\phi}\|_2$, but the results can be adapted to this case.

Preprint submitted to IEEE Open Journal of Control Systems. Received: May 9, 2025 08:56:43 Pacific Time

distributed noise samples, the size of the set of parameters indeed shrinks with increasing k.

Theorem V.1. (Repeated measurements leading to shrinking set of consistent parameters) *Suppose that the measurements* (Y_k, Φ_k) *are collected such that* $W_k^\top \sim \mathrm{Uni}(\mathcal{Z}(\Pi))$ *and* $\sigma_-(\Phi_k(z)) \geqslant a$ *for some* $a \in \mathbb{R}_{>0}$ *and all* $z \in \mathbb{R}^n$. *Then, for any probability* $0 < \pi < 1$ *and any* $\varepsilon > 0$, *there exists* k *such that if* $B_\varepsilon(\hat{\theta})$ *denotes the ball of radius* $\varepsilon > 0$ *centered at* $\hat{\theta}$ *in* $\mathbb{R}^{k\times m}$, *then*

$$p(\Theta_0 \cap \ldots \cap \Theta_{k-1} \subseteq B_\varepsilon(\hat{\theta})) > \pi. \tag{39}$$

Proof:
Let $k \geqslant 0$ and consider the set $\Theta_k = \mathcal{Z}(N_k)$. In terms of the true parameter $\hat{\theta}$, we obtain $Y_k = \hat{\theta}^\top \Phi_k + W_k$, with $W_k^\top \sim \mathrm{Uni}(\mathcal{Z}(\Pi))$. This implies that

$$Y_k^\top \sim \mathrm{Uni}(\mathcal{Z}(\Pi) + \Phi_k^\top \hat{\theta}).$$

For $\theta \in \mathbb{R}^k$, we know that $\theta \in \Theta_k$ if and only if $Y_k^\top - \Phi_k^\top \theta \in \mathcal{Z}(\Pi)$. Therefore, $Y_k^\top - \Phi_k^\top \theta \sim \mathrm{Uni}(\mathcal{Z}(\Pi) + \Phi_k^\top(\hat{\theta} - \theta))$. We can now calculate the probability of $\theta \in \Theta_k$ to be

$$p(\theta \in \Theta_k) = \frac{\mu((\Phi_k^\top(\hat{\theta} - \theta) + \mathcal{Z}(\Pi)) \cap \mathcal{Z}(\Pi))}{\mu(\mathcal{Z}(\Pi))}.$$

Recall, however, that the true parameter $\hat{\theta}$ is unknown. Let $\theta \notin B_\varepsilon(\hat{\theta})$, then, since by assumption $\sigma_-(\Phi_k) \geqslant a$, we can derive that

$$p(\theta \in \Theta_k) \leqslant \max_{\|V\|_2=1} \frac{\mu((a\varepsilon V + \mathcal{Z}(\Pi)) \cap \mathcal{Z}(\Pi))}{\mu(\mathcal{Z}(\Pi))} =: \eta(a\varepsilon).$$

Note that, since $\mathcal{Z}(\Pi)$ is bounded and convex, we have in addition that $\eta(\lambda) < 1$ for $\lambda > 0$. This means that, regardless of the values of the measurements, the probability that $\theta \in \Theta_k$ is strictly less than 1. Now, by repeating such measurements, we see that for $\theta \notin B_\varepsilon(\hat{\theta})$, we have

$$p(\theta \in \Theta_0 \cap \ldots \cap \Theta_{k-1}) = p(\theta \in \Theta_0) \cdots p(\theta \in \Theta_{k-1}) \leqslant \eta(a\varepsilon)^k.$$

From this, if k is large enough, then (39) holds. ∎

Remark V.2. (Extension to other probability distributions) Note that the proof of Theorem V.1 essentially depends on the fact that $p(W) \neq 0$ for all $W^\top \in \mathcal{Z}(\Pi)$. Therefore, we can draw the same conclusion for other probability distributions that satisfy this assumption, such as truncated Gaussian distributions. •

Thus, if the noise is uniformly random, Theorem V.1 guarantees that the set of parameters converges. This allows us to conclude that, under Algorithm 1, the sequence $g_c(z_k; \bar{\Theta}_k)$ converges to the sequence $c^\top \hat{\phi}(z_k)$ pointwise.

B. Methods to improve the candidate optimizer

In this section we discuss methods to guarantee that (34) holds. A direct way of enforcing this is by simply optimizing the expression, that is, updating z_k as

$$z_k \in \arg\min_{z\in\mathbb{R}^n} g_c(z; \bar{\Theta}_{k-1}), \tag{40}$$

where $\arg\min$ denotes the set of arguments of the minimization. This requires us to find a global optimizer of the function $g_c(z; \bar{\Theta}_{k-1})$, which might be difficult to obtain depending on the scenario. As an alternative, we devise a procedure where the property (34) is guaranteed by *locally* updating the candidate optimizer along with collecting new measurements *near* the candidate optimizer. We show that, using only local information in this manner, the algorithm converges to the true optimizer under appropriate technical assumptions.

In order to formalize this notion of 'local', we both measure and optimize on a polyhedral set around the candidate optimizer. For this, let $\mathcal{F} = \{f_i\}_{i=1}^T \subseteq \mathbb{R}^n$ be a finite set such that $0 \in \mathrm{int}(\mathrm{conv}\,\mathcal{F})$. Let $\mathcal{S}(z) := z + \mathrm{conv}\,\mathcal{F}$. Then, (34) holds if we update the optimizer by

$$z_k \in \arg\min_{z\in\mathcal{S}(z_{k-1})} g_c(z; \Theta_0 \cap \ldots \cap \Theta_{k-1}). \tag{41}$$

We measure the function at all points in $z_k + \mathcal{F}$, that is, we take $\Phi_k = \Phi^{\mathcal{F}}(z_k)$, where

$$\Phi^{\mathcal{F}}(z) := \begin{bmatrix} \phi_1(z+f_1) & \ldots & \phi_1(z+f_T) \\ \vdots & & \vdots \\ \phi_k(z+f_1) & \ldots & \phi_k(z+f_T) \end{bmatrix}.$$

The online optimization procedure then incrementally incorporates these measurements to refine the computation of the candidate optimizer. The following result investigates the properties of the resulting *online local descent*.

Theorem V.3 (Online local descent)**.** *Let* $\mathcal{F}$ *be a finite set such that* $0 \in \mathrm{int}(\mathrm{conv}\,\mathcal{F})$. *Let* $\mathcal{S}(z) := z + \mathrm{conv}\,\mathcal{F}$. *Consider Algorithm 1, employing local update rules* (41) *and* (36), *where* $\theta_k = \mathcal{Z}(N_k)$ *and* $\Phi_k = \Phi^{\mathcal{F}}(z_k)$. *Suppose that the initial point* $z_0 \in \mathbb{R}^n$ *and the data* $(Y_0, \Phi^{\mathcal{F}}(z_0))$ *are such that* $c^\top \phi^\theta$ *is strictly convex for all* $\theta \in \Theta_0$. *Then the following hold:*

(i) For any $k \geqslant 1$, *the problem* (41) *is strictly convex;*
(ii) If $z_k \neq z_{k+1}$, *then*

$$g_c(z_{k+1}; \bar{\Theta}_k) < g_c(z_k; \bar{\Theta}_{k-1}),$$

that is, (38) *holds with a strict inequality.*

Proof:
Statement (i) follows from Proposition III.11. Then, note that if $z_k \neq z_{k+1}$, we have that

$$g_c(z_{k+1}; \Theta_0 \cap \ldots \cap \Theta_k) < g_c(z_k; \Theta_0 \cap \ldots \cap \Theta_k),$$

thus proving (ii). ∎

Remark V.4. (Reduction of the complexity) Most optimization schemes require a large number of function evaluations in order to find an optimizer of e.g., (41). In turn, finding values of $g_c(\cdot; \bar{\Theta}_k)$ requires us to resolve another optimization problem. This nested nature of the problem means that reducing the number of evaluations can speed up computation significantly. This can be done by relaxing (41) as follows. Assume a Lipschitz constant of $c^\top \hat{\phi}$ is known or obtained from measurements. Since $\mathcal{F}$ is a finite set, there exists $\nu \in \mathbb{R}_{\geqslant 0}$ such that each $z \in \mathcal{S}(z_{k-1})$ is at most a distance ν from a point in $z_{k-1} + \mathcal{F}$. As such, we can instead find

$$z_k \in \arg\min_{z\in z_{k-1}+\mathcal{F}} g_c(z; \Theta_0 \cap \ldots \cap \Theta_{k-1}),$$

and employ the value of ν and the Lipschitz constant to approximate the optimal value of (41). •

Recall that, without making further assumptions, we can only guarantee that the algorithm converges, but not that it

Preprint submitted to IEEE Open Journal of Control Systems. Received: May 9, 2025 08:56:43 Pacific Time

converges to the true optimizer. Moreover, even when the assumptions of Theorem V.1 hold, we do not yet have a way of concluding that the algorithm has converged. The following result employs the uncertainty near the optimizer, to provide a criterion for convergence.

Lemma V.5 (Stopping criterion)**.** *Let Θ be compact and $\mathcal{S} \subseteq \mathbb{R}^n$ closed. Define*

$$\bar{z} \in \arg\min_{z \in \mathcal{S}} g_c(z;\Theta), \quad \hat{z} \in \arg\min_{z \in \mathcal{S}} c^\top \hat{\phi}(z).$$

Then $g_c(\bar{z};\Theta) \geqslant c^\top \hat{\phi}(\hat{z}) \geqslant g_c(\bar{z};\Theta) - 2\max_{z\in\mathcal{S}} U_c(z;\Theta)$.

Proof:
For the first inequality, note that by definition,

$$g_c(\bar{z};\Theta) \geqslant c^\top \hat{\phi}(\bar{z}) \geqslant c^\top \hat{\phi}(\hat{z}).$$

To show the second inequality, note that $g_c(\bar{z};\Theta) \leqslant g_c(\hat{z};\Theta)$, and

$$2\max_{z\in\mathcal{S}} U_c(z;\Theta) \geqslant g_c(\hat{z};\Theta) - c^\top \hat{\phi}(\hat{z};\Theta).$$

Combining these, we obtain

$$g_c(\bar{z};\Theta) - 2\max_{z\in\mathcal{S}} U(z;\Theta) \leqslant c^\top \hat{\phi}(\hat{z};\Theta),$$

which proves the result. ∎

Importantly, we note that the upper and lower bounds in Lemma V.5 can be computed purely in terms of data.

If the maximal uncertainty on the set $\mathcal{S}$ is equal to zero, then the local minima of g_c and $c^\top \hat{\phi}$ coincide. In addition, if $c^\top \hat{\phi}$ is strictly convex, we have that any local minimum in the *interior* of $\mathcal{S}$ is a global minimum.

Corollary V.6. (Uncertainty under repetition) *Under the assumptions of Theorem V.3, suppose in addition that for $k \geqslant 1$, the measurements $(Y_k, \Phi^{\mathcal{F}}(z_k))$ are collected such that $W_k^\top \sim \mathrm{Uni}(\mathcal{Z}(\Pi))$ and $\sigma_-(\Phi^{\mathcal{F}}(z_k)) \geqslant a$ for some $a \in \mathbb{R}_{>0}$ and all k. Then, for any $z \in \mathbb{R}^n$, the expected value of the uncertainty monotonically converges to 0, that is,*

$$U_c(z;\Theta_0 \cap \ldots \cap \Theta_{k-1}) \geqslant U_c(z;\Theta_0 \cap \ldots \cap \Theta_k),$$

and $\lim_{k\to\infty} \mathbb{E}(U_c(z;\Theta_0 \cap \ldots \cap \Theta_k)) = 0$.

Proof:
The monotonic decrease of the uncertainty readily follows from its definition. If $\Theta_0 \cap \ldots \cap \Theta_{k-1} \subseteq B_\varepsilon(\hat{\theta})$, then it is immediate from the definitions that

$$U_c(z;\Theta_0 \cap \ldots \cap \Theta_{k-1}) \leqslant U_c(z;B_\varepsilon(\hat{\theta})).$$

Given that the map $\theta \mapsto c^\top \phi^\theta(z)$ is linear in θ, the right-hand side can be seen to converge to 0 for $\epsilon \to 0$. Combining these pieces with Theorem V.1 proves the statement. ∎

As a consequence of Lemma V.5 and Corollary V.6, we conclude that the expected difference between the optimal value $\min_{z\in\mathcal{S}} c^\top \hat{\phi}(z)$ of the unknown function and the optimal value $\min_{z\in\mathcal{S}} g_c(z;\Theta_0 \cap \cdots \cap \Theta_k)$ provided by online local descent both converge to zero.

VI. Simulation examples

Here we provide two simulation examples to illustrate the proposed framework and our results. We consider an application to data-based contraction analysis based on Section A, and the data-based suboptimal regulation of an unmanned aerial vehicle, using Section B. We refer the interested reader to our conference paper [17, Section V] for an illustration of a simplified version of online cautious suboptimization.

A. Application to data-based contraction analysis

Consider continuous-time systems $\dot{z} = \phi^\theta(z)$, where $z \in \mathbb{R}^2$, the parameter $\theta \in \mathbb{R}^{6\times 2}$, and with basis functions

$$\phi_1(z) = z_1, \quad \phi_3(z) = \sin(z_1) - 1, \quad \phi_5(z) = \cos(z_1),$$
$$\phi_2(z) = z_2, \quad \phi_4(z) = \sin(z_2) - 1, \quad \phi_6(z) = \cos(z_2).$$

These define the vector-function $b : \mathbb{R}^2 \to \mathbb{R}^6$, cf. (1). Suppose that the parameter of the unknown function $\hat{\phi}$ is

$$\hat{\theta}^\top = \begin{bmatrix} -6 & 1 & -1 & 1 & -1 & 1 \\ 0 & -6 & 1 & -1 & 1 & -1 \end{bmatrix}.$$

We are interested in determining whether the system $\dot{z} = \hat{\phi}(z)$ is strictly contracting with respect to the unweighted $2-$norm on the basis of noisy measurements.

As a first step, note that for the Jacobian of b, we have $\|J(b)\|_2 = \sqrt{2}$. Then writing $\hat{\phi}(z) = -6z + h(z)$, we can conclude that

$$\begin{aligned} \mathrm{osLip}(\hat{\phi}) &\leqslant -6 + \mathrm{osLip}(h) \leqslant -6 + \|J(h)\|_2 \\ &\leqslant -6 + \| \begin{bmatrix} 0 & 1 & -1 & 1 & -1 & 1 \\ 0 & 0 & 1 & -1 & 1 & -1 \end{bmatrix} \|_2 \cdot \|J(b)\|_2 \\ &\approx -1.8694. \end{aligned} \tag{42}$$

Thus, the system $\dot{z} = \hat{\phi}(z)$ is indeed strictly contracting.

However, as before we assume that we do not have access to the value of $\hat{\theta}$ and can only assess contractivity based on measurements of $\hat{\phi}$. To be precise, we assume that we have access to samples of a single continuous-time trajectory starting from $z(0) = (10, -20)^\top$ at the time instances

$$t \in \{0, 0.01, \ldots, 0.25\}.$$

The noise is assumed to satisfy $WW^\top \leqslant 10 \cdot I_2$, giving rise to a set $\Theta = \mathcal{Z}(N)$. Two examples of parameters compatible with the measurements are given by

$$\begin{bmatrix} -5.2588 & 1.3807 & -1.4810 & 0.8087 & -0.8825 & 1.5444 \\ -0.2166 & -6.1490 & 1.1110 & -0.5087 & 0.4157 & -0.6826 \end{bmatrix},$$
$$\begin{bmatrix} -6.3111 & 0.8091 & -0.6433 & 1.0725 & -0.5355 & 0.9938 \\ 0.7750 & -5.6377 & 0.9785 & -1.1375 & 1.3976 & -0.8905 \end{bmatrix}.$$

To illustrate the different systems compatible with the measurements, Figure 2 shows the trajectories emanating from the origin for a number of such systems.

In order to guarantee strict contraction in terms of the measurements, we employ the condition of Corollary IV.2. With these measurements Φ has full row rank and thus $\Theta = \mathcal{Z}(N)$ is compact. For the last assumption, from the fact that $\|J(b)\|_2 = \sqrt{2}$, we obtain that for any $z, z^\star \in \mathbb{R}^2$,

$$\|b(z) - b(z^\star)\|_{2,(-N_{22})^{-1/2}} \leqslant \frac{\sqrt{2}}{\sqrt{-\lambda_{\max}(N_{22})}} \|z - z^\star\|_2.$$

This means that we can use Corollary IV.2, which shows that the unknown system is strictly contracting if

Preprint submitted to IEEE Open Journal of Control Systems. Received: May 9, 2025 08:56:43 Pacific Time

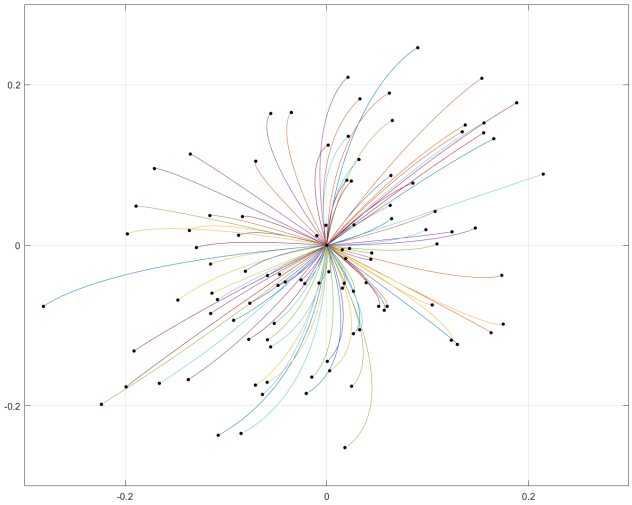

FIGURE 2: Trajectories of $\dot{z} = \phi^\theta(z)$ from the origin for 100 randomly generated realizations of the parameter $\theta \in \Theta$. The black dots indicate the resulting equilibria.

$$\operatorname{osLip}(\phi^{\text{lse}}) < -\frac{\sqrt{2\lambda_{\max}(N|N_{22})}}{\sqrt{-\lambda_{\max}(N_{22})}} \approx -1.78.$$

Using a line of reasoning as in (42) allows us to conclude that this indeed holds. As such, the system $\dot{z} = \phi^\theta$ is strictly contracting for all $\theta \in \Theta$. We plot a number of trajectories of the systems corresponding to 4 different realizations of the parameter θ in Figure 3. It can be seen that each is strictly contracting.

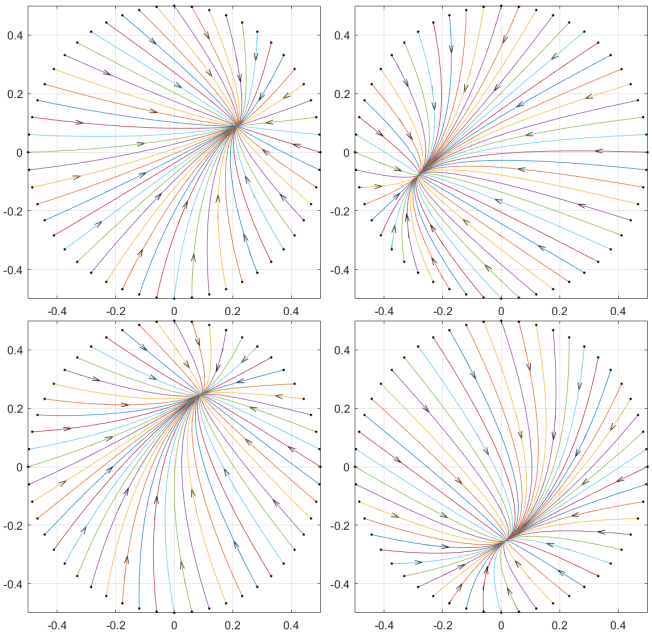

FIGURE 3: Trajectories of $\dot{z} = \phi^\theta(z)$ starting from a number of different initial conditions on the circle with radius $\frac{1}{2}$ for four different realizations of $\theta \in \Theta$.

B. Suboptimal regulation of unmanned aerial vehicle

This example considers the suboptimal regulation of a fixed-wing unmanned aerial vehicle (UAV). We consider a model of the longitudinal motion for which the relevant equations of motion are derived in e.g., [6, Ch. 2]. For the values of the parameters and the linearization we follow [3], which derives a benchmark model on the basis of 10 different models of real-world UAV's of different specifications. The derived system is a state-space model

$$\dot{x}(t) = A_{\text{cont}}x(t) + B_{\text{cont}}u(t),$$

given in [3, Eq. (8)-(10)], with states $x(t) \in \mathbb{R}^4$ and inputs $u(t) \in \mathbb{R}$. Here, $x_1(t)$ denotes the *deviation from the nominal forward velocity*, $x_2(t)$ the *deviation from the nominal angle of attack*, $x_3(t)$ the *pitch angle*, and $x_4(t)$ the *pitch rate*. The input $u(t)$ represents the *elevator deflection*. We use the parameters derived in [3, Eq. (13)-(14)], and thus

$$A_{\text{cont}} = \begin{bmatrix} -0.240 & 0.345 & -0.411 & 0 \\ -1.905 & -10.695 & 0 & 0.941 \\ 0 & 0 & 0 & 1 \\ 0.457 & -250.513 & 0 & -8.844 \end{bmatrix}, B_{\text{cont}} = \begin{bmatrix} 0 \\ -0.301 \\ 0 \\ -98.658 \end{bmatrix}.$$

We consider a discretization of this model with stepsize $\frac{1}{20}$s, thus arriving at a system of the form (28), where $\hat{A} = I_4 + \frac{1}{20}A_{\text{cont}}$ and $\hat{B} = \frac{1}{20}B_{\text{cont}}$. These matrices are unknown to us, and instead we have access to a single second-long [2] trajectory of measurements of the state $x(t)$ with $x(0) = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}^\top$, and fixed input $u(t) = -4$. Thus, $T = 20$. Denoting the matrix collecting the noise samples by W_s, we assume that the noise on the measurements is such that $W_s W_s^\top \leqslant 10^{-4} I_4$. These measurements give rise to a set of systems consistent with the data given by $\Sigma = \mathcal{Z}(M)$.

Following Section B, we aim at finding suboptimal values of the norm of an unknown cost function. In this example, we consider a simple cost function $\hat{\phi} : \mathbb{R}^4 \to \mathbb{R}^4$ of the form $\hat{\phi}(x) = x - \hat{x}$, where $\hat{x} = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix}^\top$. This situation might arise, for instance, when the UAV is tasked to mirror the orientation and speed of another UAV on the basis of noisy measurements. If the cost function were known, the problem of minimizing $\|\hat{\phi}\|_2$ would be to simply regulate x as close to $\hat{x}$ as possible. However, we assume that we only know that $\hat{\phi}$ is affine in x, and thus of the form

$$\phi^\theta(x) := \theta^\top \begin{pmatrix} 1 \\ x \end{pmatrix}, \text{ where } \hat{\theta} = \begin{bmatrix} -\hat{x}^\top \\ I_4 \end{bmatrix}.$$

We assume that the measurements available to us are collected concurrently with the collection of the measurements of the unknown system. In line with Section II, this means that we measure at $z_i := x(i-1)$ for $i = 1, \ldots, T = 20$. We assume a noise model of the form $WW^\top \leqslant I_4$, and obtain a set of parameters consistent with the measurements given by $\Theta = \mathcal{Z}(N)$.

We are interested in finding upper bounds for the quantity

$$\min_u \|\hat{\phi}((I - \hat{A})^{-1}\hat{B}u)\|_2.$$

[2] Note that, in the simulation we applied a fixed input. In reality, one would not want the UAV to be uncontrolled for a full second. It should be noted that the results are also valid for data collected in shorter bursts.

Preprint submitted to IEEE Open Journal of Control Systems. Received: May 9, 2025 08:56:43 Pacific Time

Before applying Theorem IV.8, we need to check its assumptions. First, we employ Lemma IV.4 in order to check whether each A consistent with the measurements is stable. Solving the LMI (29) yields

$$\bar{P} \approx \begin{bmatrix} 0.0072 & 0.0015 & 0.0004 & -0.0102 \\ 0.0015 & 0.1665 & 0.1218 & 0.1496 \\ 0.0004 & 0.1218 & 0.1648 & -1.3489 \\ -0.0102 & 0.1496 & -1.3489 & 45.00437 \end{bmatrix}.$$

Next, we employ Lemma A.3 to obtain a Lipschitz constant. Given that the basis functions are linear, we can easily minimize L over (45) to obtain $L = 2.0171$. Thus, L is a Lipschitz constant for ϕ^θ for all $\theta \in \mathcal{Z}(N)$.

Therefore, the required pieces are in place, and we can employ Theorem IV.8 to show that

$$\min_u \|\hat{\phi}((I - \hat{A})^{-1}\hat{B}u)\|_2 \leqslant 2.3347. \tag{43}$$

This bound corresponds to (32) with minimizers $u^\star = 0.2321$ and $x^\star = \begin{pmatrix} 0.809 & 0.054 & -0.355 & -0.006 \end{pmatrix}^\top$ on the left- and right-hand sides, respectively.

Figure 4 shows the results of applying this input to different realizations of the system matrices compatible with the measurements. We make the following observations. First, one can see that each of the systems converges to a fixed point, and each of these fixed points are within a distance $\varepsilon^-(x) = 0.4491$ of x. Second, for certain systems the transients, especially in states x_1 and x_2, converge quite slowly. This holds because the true system is marginally stable and the same holds for elements of set $\mathcal{Z}(M)$ (to be precise, the upper bound of Theorem IV.9 can only be shown to hold for $0.999 < \lambda < 1$).

Corresponding to these trajectories, Figure 5 shows the resulting values of $\|\phi^\theta(x_k)\|_2$, for both the true value and for different realizations of $\theta \in \mathcal{Z}(N)$. The plot reveals that, with the true parameter $\hat{\theta}$, the values at the fixed points are well below the bound (43). In fact, resolving the problem with known $\hat{\phi}$ yields $u^\star = 0.1820$ and an upper bound of 0.7963. Indeed, as the left plot shows, the unknown nature of $\hat{\phi}$ has a large effect and therefore has to be taken into account when providing formal guarantees.

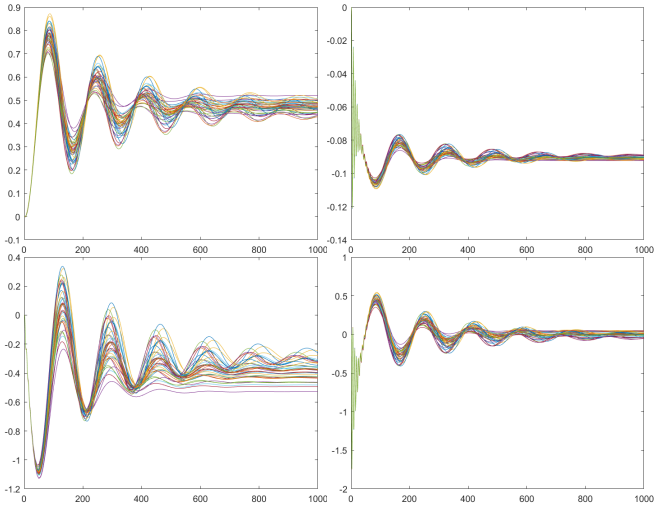

FIGURE 4: State trajectory corresponding to the found input $u^\star = 0.2321$ for 40 different realizations of the matrices (A, B) consistent with the measurements.

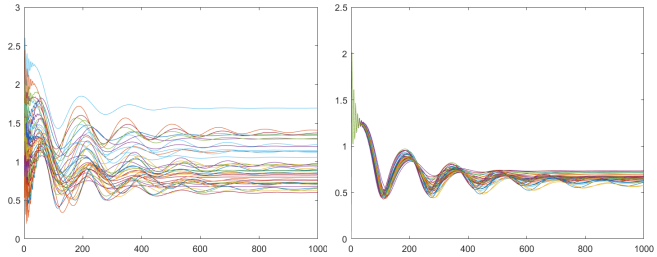

FIGURE 5: Values of $\|\phi^\theta(x_k)\|_2$ for the state trajectories of Figure 4. On the left, we pick for each system a different realization of $\theta \in \mathcal{Z}(N)$. On the right, we use the true value of $\hat{\theta}$.

VII. Conclusions

We have developed a set-valued regression approach to data-based optimization of an unknown vector-valued function. Taking a worst-case perspective, we have provided a range of guarantees in terms of measurements on the minimization of least conservative upper bounds of both the norm and any linear combination of the components of the unknown function. Our analysis has yielded closed-form expressions for the proposed upper bounds, conditions to ensure their convexity (as an enabler for the use of gradient-based methods for their optimization), and Lipschitzness characterizations to facilitate simplifications based on interpolation.

We have illustrated the applicability of the proposed cautious optimization approach in systems and control scenarios with unknown dynamics: first, by providing conditions on the data which guarantee that the dynamics is strongly contracting and, second, by providing conditions under which we can regulate an unknown system to a point with a guaranteed maximal value of an unknown cost function. Finally, we have considered online scenarios, where one repeatedly combines the optimization of the unknown function with the collection of additional data. Under mild assumptions, we show that repeated measurements lead to a strictly shrinking set of compatible parameters and we build on this result to design an online procedure that provides sequential upper bounds which converge to a true optimizer.

Future work will investigate conditions to ensure other regularity properties of the unknown function and its least conservative upper bounds. Moreover, we envision an investigation of the impact of the choice of basis functions on the guarantees and computational efficiency of the proposed cautious optimization methods (particularly, the use of polynomial bases coupled with sum-of-squares optimization techniques). The links between choices of basis and the performance on out-of-basis functions under certain regularity conditions is an interesting direction. Lastly, many applications in the field of control remain open, such as the certification of Lyapunov stability and controller design for nonlinear systems.

Preprint submitted to IEEE Open Journal of Control Systems. Received: May 9, 2025 08:56:43 Pacific Time

REFERENCES

[1] K. B. Ariyur and M. Krstić. *Real-Time Optimization by Extremum-Seeking Control*. Wiley, New York, 2003.

[2] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Applied Mathematics Series. Princeton University Press, Princeton, NJ, 2009.

[3] E. Bertran, P. Tercero, and A. Sànchez-Cerdà. UAV generalized longitudinal model for autopilot controller designs. *Aircraft Engineering and Aerospace Technology*, 94(3):380–391, 2022.

[4] D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

[5] G. Bianchin, M. Vaquero, J. Cortés, and E. Dall'Anese. Online stochastic optimization of unknown linear dynamical systems: data-driven controller synthesis and analysis. *IEEE Transactions on Automatic Control*, 69(7), 2024. 4411-4426.

[6] J. H. Blakelock. *Automatic Control of Aircraft and Missiles*. John Wiley & Sons, 1991.

[7] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2009. ISBN 0521833787.

[8] S. L. Brunton, J. L. Proctor, and J. N. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15): 3932–3937, 2016.

[9] M. Budišić, R. Mohr, and I. Mezić. Applied Koopmanism. *Chaos*, 22 (4):047510, 2012.

[10] F. Bullo. *Contraction Theory for Dynamical Systems*. Kindle Direct Publishing, 1.1 edition, 2023. ISBN 979-8836646806.

[11] J. Calliess, S. J. Roberts, C. E. Rasmussen, and J. Maciejowski. Lazily adapted constant kinky inference for nonparametric regression and model-reference adaptive control. *Automatica*, 122:109216, 2020.

[12] L. Cothren, G. Bianchin, and E. Dall'Anese. Data-enabled gradient flow as feedback controller: Regulation of linear dynamical systems to minimizers of unknown functions. In *Learning for Dynamics and Control Conference*, volume 168 of *Proceedings of Machine Learning Research*, pages 234–247. PMLR, 2022.

[13] G. Cybenko. Dynamic load balancing for distributed memory multiprocessors. *Journal of Parallel and Distributed Computing*, 7(2): 279–301, 1989.

[14] A. Davydov, S. Jafarpour, and F. Bullo. Non-Euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, 67(12):6667–6681, 2022.

[15] C. De Persis, M. Rotulo, and P. Tesi. Learning controllers from data via approximate nonlinearity cancellation. *IEEE Transactions on Automatic Control*, 68(10):6082–6097, 2023.

[16] F. Dörfler, P. Tesi, and C. De Persis. On the certainty-equivalence approach to direct data-driven LQR design. *IEEE Transactions on Automatic Control*, 68(12):7989–7996, 2023.

[17] J. Eising and J. Cortés. Set-valued regression and cautious suboptimization: from noisy data to optimality. In *IEEE Conf. on Decision and Control*, pages 5319–5324, Singapore, 2023.

[18] J. Eising and J. Cortés. When sampling works in data-driven control: informativity for stabilization in continuous time. *IEEE Transactions on Automatic Control*, 70(1):1558–2523, 2025.

[19] Michel Gevers. Identification for control: From the early achievements to the revival of experiment design. *European Journal of Control*, 11 (4):335–352, 2005.

[20] M. Guo, C. De Persis, and P. Tesi. Data-driven stabilization of nonlinear polynomial systems with noisy data. *IEEE Transactions on Automatic Control*, 67(8):4210–4217, 2021.

[21] M. Haseli and J. Cortés. Generalizing dynamic mode decomposition: balancing accuracy and expressiveness in Koopman approximations. *Automatica*, 153:111001, 2023.

[22] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, New York, 2nd edition, 2013.

[23] A. Hauswirth, Z. He, S. Bolognani, G. Hug, and F. Dörfler. Optimization algorithms as robust feedback controllers. *Annual Reviews in Control*, 57:100941, 2024.

[24] M. R. Jovanović, P. J. Schmid, and J. W. Nichols. Sparsity-promoting dynamic mode decomposition. *Physics of Fluids*, 26(2):024103, 2014.

[25] M. Korda and I. Mezic. Optimal construction of Koopman eigenfunctions for prediction and control. *IEEE Transactions on Automatic Control*, 65(12):5114–5129, 2020.

[26] M. Krstić and H-H. Wang. Stability of extremum seeking feedback for general nonlinear dynamic systems. *Automatica*, 36(4):595–601, 2000.

[27] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*, volume 149 of *Other Titles in Applied Mathematics*. SIAM, Philadelphia, PA, 2016.

[28] Y. Li, J. Yu, L. Conger, T. Kargin, and A. Wierman. Learning the uncertainty sets of linear control systems via set membership: A non-asymptotic analysis. *arXiv preprint arXiv:2309.14648*, 2023.

[29] W. Lohmiller and J.-J. E. Slotine. On contraction analysis for nonlinear systems. *Automatica*, 34(6):683–696, 1998.

[30] V. G. Lopez and M. A. Müller. On a continuous-time version of Willems' lemma. In *IEEE Conf. on Decision and Control*, pages 2759–2764, 2022.

[31] I. Markovsky. Data-driven simulation of generalized bilinear systems via linear time-invariant embedding. *IEEE Transactions on Automatic Control*, 68(2):1101–1106, 2022.

[32] M. Milanese and C. Novara. Set membership estimation of nonlinear regressions. In *IFAC World Congress*, volume 35, pages 7–12, Barcelona, Spain, 2002.

[33] M. Milanese and A. Vicino. Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *Automatica*, 27(6):997–1009, 1991.

[34] O. Nelles. *Nonlinear System Identification*. Springer Cham, 2020.

[35] I. Pólik and T. Terlaky. A survey of the S-Lemma. *SIAM Review*, 49 (3):371–418, 2007.

[36] Paolo Rapisarda, M Kanat Camlibel, and HJ van Waarde. A persistency of excitation condition for continuous-time systems. *IEEE Control Systems Letters*, 7:589–594, 2022.

[37] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA, 2005.

[38] P. J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010.

[39] P. J. Schmid. Dynamic mode decomposition and its variants. *Annual Review of Fluid Mechanics*, 54(1):225–254, 2022.

Preprint submitted to IEEE Open Journal of Control Systems. Received: May 9, 2025 08:56:43 Pacific Time

[40] P. Tabuada and B. Gharesifard. Universal approximation power of deep residual neural networks through the lens of control. *IEEE Transactions on Automatic Control*, 68(5):2715–2728, 2023.

[41] A.R. Teel and D. Popovic. Solving smooth and nonsmooth multivariable extremum seeking problems by the methods of nonlinear programming. In *American Control Conference*, pages 2394–2399, Arlington, VA, 2001.

[42] R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B*, 58(1):267–288, 1996.

[43] H. J. van Waarde, J. Eising, H. L. Trentelman, and M. K. Camlibel. Data informativity: a new perspective on data-driven analysis and control. *IEEE Transactions on Automatic Control*, 65(11):4753–4768, 2020.

[44] H. J. van Waarde, M. K. Camlibel, and H. L. Trentelman. Data-driven analysis and design beyond common Lyapunov functions. In *IEEE Conf. on Decision and Control*, pages 2783–2788, 2022.

[45] H. J. van Waarde, M. K. Camlibel, J. Eising, and H. L. Trentelman. Quadratic matrix inequalities with applications to data-based control. *SIAM Journal on Control and Optimization*, 61(4):2251–2281, 2023.

[46] H. J. van Waarde, J. Eising, H. L. Trentelman, and M. K. Camlibel. The informativity approach: To data-driven analysis and control. *IEEE Control Systems*, 43(6):32–66, 2023.

[47] J. C. Willems, P. Rapisarda, I. Markovsky, and B. L. M. De Moor. A note on persistency of excitation. *Systems & Control Letters*, 54(4): 325–329, 2005.

Appendix

A. Nonnegativity of the parameters

The following result provides a test in terms of the data for θc to be elementwise nonnegative for each θ compatible with the data.

Lemma A.1. (Nonnegativity of parameters) *Given data (Y, Φ), let $\Theta = \mathcal{Z}(N)$, with N is as in (4). Let $c \neq 0$. Then, $\theta c \in \mathbb{R}^k_{\geqslant 0}$ for all $\theta \in \Theta$ if and only if Φ has full row rank and one of the following conditions hold*

(i) $c^\top(N|N_{22})c = 0$ *and* $-N_{22}^{-1}N_{21}c \in \mathbb{R}^k_{\geqslant 0}$, *or*

(ii) $c^\top(N|N_{22})c > 0$, $-N_{22}^{-1}N_{21}c \in \mathbb{R}^k_{>0}$ *and for all* $i = 1, \dots, k$,

$$c^\top(N|N_{22})c + \frac{(e_i^\top N_{22}^{-1}N_{21}c)^2}{e_i^\top N_{22}^{-1}e_i} \leqslant 0.$$

Proof:
Recall the definition of N_c from (14). Since $c \neq 0$ we have that $\mathcal{Z}(N)c = \mathcal{Z}(N_c)$. Thus, $\theta c \in \mathbb{R}^k_{\geqslant 0}$ for all $\theta \in \Theta$ if and only if $\mathcal{Z}(N_c) \subseteq \mathbb{R}^k_{\geqslant 0}$.

($\Rightarrow$): Suppose that $\mathcal{Z}(N_c) \subseteq \mathbb{R}^k_{\geqslant 0}$. This implies that the set $\mathcal{Z}(N_c)$ contains no nontrivial subspace, which implies that $N_{22} < 0$. In turn, this holds only if Φ has full row rank.

Recall that $N|N_{22} \geqslant 0$ by assumption. Therefore we have either $c^\top(N|N_{22})c = 0$ or $c^\top(N|N_{22})c > 0$. If $c^\top(N|N_{22})c = 0$, then $N_c \leqslant 0$. As such, $\mathcal{Z}(N_c) = \{-N_{22}^{-1}N_{21}\}$, that is, *only* the least-squares estimate is compatible with the data. Hence, *(i)* follows.

Next, consider $c^\top(N|N_{22})c > 0$, that is, N_c has one positive eigenvalue. Note that the Slater condition holds, and we can apply the S-Lemma [35, Thm 2.2] to see that $\mathcal{Z}(N_c) \subseteq \mathbb{R}^k_{\geqslant 0}$ is equivalent to the existence of $\alpha_1, \dots, \alpha_k \geqslant 0$ with

$$\begin{bmatrix} 0 & e_i^\top \\ e_i & 0 \end{bmatrix} - \alpha_i N_c \geqslant 0,$$

for each standard basis vector e_i. This requires $\alpha_i \neq 0$ for each $i \in \{1, \dots, k\}$. Thus, this can equivalently written as

$$N_c - \frac{1}{\alpha_i}\begin{bmatrix} 0 & e_i^\top \\ e_i & 0 \end{bmatrix} \leqslant 0.$$

Since $N_{22} < 0$, we can define $\beta_i = \frac{1}{\alpha_i}$ and conclude that the latter holds if and only if

$$\begin{aligned} &c^\top N_{11}c - (c^\top N_{12} - \beta_i e_i^\top)N_{22}^{-1}(N_{21}c - \beta_i e_i) \\ &\quad = c^\top(N|N_{22})c + \beta_i(c^\top N_{12}N_{22}^{-1}e_i + e_i^\top N_{22}^{-1}N_{21}c) \\ &\qquad - \beta_i^2 e_i^\top N_{22}^{-1}e_i \leqslant 0. \end{aligned}$$

In turn, there exist such β_i if and only if the minimal value over all $\beta_i > 0$ satisfies this. Since this is a scalar expression, we can explicitly minimize it to prove that *(ii)* holds.

($\Leftarrow$): Similar to before, if Φ has full row rank and *(i)* holds, then $\mathcal{Z}(N_c) = \{-N_{22}^{-1}N_{21}\}$. Hence $\mathcal{Z}(N_c) \subseteq \mathbb{R}^k_{\geqslant 0}$ follows immediately. If, instead, Φ has full row rank and *(ii)* holds, reversing the steps of the previous part of the proof yields that $\mathcal{Z}(N_c) \subseteq \mathbb{R}^k_{\geqslant 0}$. ■

B. Bounding the Jacobian

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable and denote its Jacobian,

$$J(f) := \begin{bmatrix} \frac{df}{dz_1} & \cdots & \frac{df}{dz_n} \end{bmatrix} = \begin{bmatrix} \frac{df_1}{dz_1} & \cdots & \frac{df_1}{dz_n} \\ \vdots & & \vdots \\ \frac{df_m}{dz_1} & \cdots & \frac{df_m}{dz_n} \end{bmatrix}.$$

Clearly, $J(\phi^\theta) = \theta^\top J(b)$. The following result, consequence of Theorem III.6, provides an expression for the Jacobian of the worst-case linear bound function.

Corollary A.2. (Jacobians of the worst-case linear bound) *Given data (Y, Φ), let $\Theta = \mathcal{Z}(N)$ and assume $N_{22} < 0$. Let the basis functions $\phi_i(z)$ be differentiable and such that $b(z) \neq 0$ for all z. Then*

$$J(g_c(z;\Theta)) = \left(c^\top N_{12} + \frac{\sqrt{c^\top(N|N_{22})c}}{\sqrt{b(z)^\top(-N_{22}^{-1})b(z)}}b(z)^\top\right)(-N_{22}^{-1})J(b(z)).$$

In a similar fashion, we can also obtain gradients of the bound of $g(z;\Theta)$ given in (23). Note that, if $N_{22} < 0$ this requires that both $b(z) \neq 0$ and $\phi^{\text{lse}}(z;\Theta) \neq 0$ for all z.

C. Lipschitz continuity

We investigate Lipschitz continuity for an unknown function $\hat{\phi}$. In line with the rest of the paper, we guarantee this by providing conditions for the function ϕ^θ for *all* parameters $\theta \in \Theta$. Without added difficulty, the following result considers the slightly more general case of weighted norms.

Lemma A.3. (Lipschitz constants of functions consistent with the measurements) *Given data (Y, Φ), let $\Theta = \mathcal{Z}(N)$,*

Preprint submitted to IEEE Open Journal of Control Systems. Received: May 9, 2025 08:56:43 Pacific Time

with N as in (4), and assume N has at least one positive eigenvalue. Let $P \in \mathbb{S}^m$ and $Q \in \mathbb{S}^n$ with $P > 0$ and $Q > 0$. For $L \geqslant 0$ and $z, z^\star \in \mathcal{S} \subseteq \mathbb{R}^n$, we have

$$\|\phi^\theta(z) - \phi^\theta(z^\star)\|_{2,P^{1/2}} \leqslant L\|z - z^\star\|_{2,Q^{1/2}} \quad \textit{for all} \quad \theta \in \Theta$$

if and only if there exists $\alpha \geqslant 0$ such that

$$\left[\begin{array}{cc:c} L^2P^{-1} & 0 & 0 \\ 0 & 0 & \frac{b(z)-b(z^\star)}{\|z-z^\star\|_{2,Q^{1/2}}} \\ \hdashline 0 & \frac{b(z)^\top - b(z^\star)^\top}{\|z-z^\star\|_{2,Q^{1/2}}} & 1 \end{array}\right] - \alpha \left[\begin{array}{c:c} N & 0 \\ \hdashline 0 & 0 \end{array}\right] \geqslant 0. \quad (44)$$

If, in addition the basis functions ϕ_i are differentiable, then

$$\|J(\phi^\theta)\|_2 \leqslant L \quad \textit{for all} \quad \theta \in \Theta$$

if and only if there exists $\alpha \geqslant 0$ such that

$$\left[\begin{array}{cc:c} L^2 I_m & 0 & 0 \\ 0 & 0 & J(b) \\ \hdashline 0 & J(b)^\top & I_n \end{array}\right] - \alpha \left[\begin{array}{c:c} N & 0 \\ \hdashline 0 & 0 \end{array}\right] \geqslant 0. \quad (45)$$

Proof:
Note that $\phi^\theta(z) - \phi^\theta(z^\star) = \theta^\top(b(z) - b(z^\star))$. Similarly, $J(\phi^\theta) = \theta^\top J(b)$. The result can then be proven by following the steps of the proof of Lemma III.1. ■

Remark A.4. (Single check for establishing Lipschitz constant) To check for Lipschitzness, we can employ the method of Remark III.4 to avoid checking (44) for all $z, z^\star \in \mathbb{R}^n$. Suppose the basis satisfies a Lipschitz condition of the form

$$\|b(z) - b(z^\star)\|_2 \leqslant L_b\|z - z^\star\|^2_{2,Q^{1/2}},$$

for all $z, z^\star \in \mathbb{R}^n$. Then,

$$\frac{(b(z) - b(z^\star))(b(z) - b(z^\star))^\top}{\|z - z^\star\|^2_{2,Q^{1/2}}} \leqslant L_b^2 I,$$

and thus, as in Remark III.4 we have

$$\begin{bmatrix} L^2P^{-1} & 0 \\ 0 & -L_b^2 I_k \end{bmatrix} - \alpha N \geqslant 0 \quad (46)$$

implies (44). This gives a single condition to check to establish a Lipschitz constant valid for all $z, z^\star \in \mathbb{R}^n$. •

Preprint submitted to IEEE Open Journal of Control Systems. Received: May 9, 2025 08:56:43 Pacific Time