Automatica 145 (2022) 110537

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

# Data-driven control via Petersen's lemma[☆]

Andrea Bisoffi [a,*], Claudio De Persis [a], Pietro Tesi [b]

[a] ENTEG, University of Groningen, 9747 AG Groningen, The Netherlands
[b] DINFO, University of Florence, 50139 Florence, Italy

ARTICLE INFO

*Article history:*
Received 28 September 2021
Received in revised form 2 April 2022
Accepted 13 June 2022
Available online 29 August 2022

*Keywords:*
Data-based control
Optimization-based controller synthesis
Analysis of systems with uncertainty
Robust control of nonlinear systems
Linear matrix inequalities
Sum-of-squares

ABSTRACT

We address the problem of designing a stabilizing closed-loop control law directly from input and state measurements collected in an experiment. In the presence of a process disturbance in data, we have that a set of dynamics could have generated the collected data and we need the designed controller to stabilize such set of data-consistent dynamics robustly. For this problem of data-driven control with noisy data, we advocate the use of a popular tool from robust control, Petersen's lemma. In the cases of data generated by linear and polynomial systems, we conveniently express the uncertainty captured in the set of data-consistent dynamics through a matrix ellipsoid, and we show that a specific form of this matrix ellipsoid makes it possible to apply Petersen's lemma to all of the mentioned cases. In this way, we obtain necessary and sufficient conditions for data-driven stabilization of linear systems through a linear matrix inequality. The matrix ellipsoid representation enables insights and interpretations of the designed control laws. In the same way, we also obtain sufficient conditions for data-driven stabilization of polynomial systems through alternate (convex) sum-of-squares programs. The findings are illustrated numerically.



## 1. Introduction

### 1.1. Motivation and Petersen's lemma

Data-driven control design is a relevant methodology to tune controllers whenever modeling from first principles is challenging, the model parameters are possibly redundant and cannot be unambiguously identified through suitable experiments, while (possibly large) datasets can be obtained from the process to be controlled. Thanks to the technological trend that measurements are increasingly easier to access and retrieve, using data to directly design controllers has witnessed a renewed surge in interest in recent years (Baggio, Katewa, & Pasqualetti, 2019; Berberich, Romer, Scherer, & Allgöwer, 2020; Coulson, Lygeros, & Dörfler, 2019; Dai & Sznaier, 2018; De Persis & Tesi, 2020; Recht, 2019; van Waarde, Camlibel, & Mesbahi, 2021).

These recent developments have been drawing results from classical areas of control theory such as behavioral theory (Coulson et al., 2019; De Persis & Tesi, 2020; Dörfler, Coulson, & Markovsky, 2022), set-membership system identification, and robust control (Berberich, Romer, Scherer, & Allgöwer, 2020; Dai & Sznaier, 2018). A pivotal role in many of these developments has been played by the so-called fundamental lemma by Willems, Rapisarda, Markovsky, and De Moor (2005, Thm. 1); qualitatively speaking, this result shows that for a linear system, controllability and persistence of excitation ensure that its representation through matrices $(A, B)$ is equivalent to a representation through a finite-length trajectory; however, such trajectory is assumed not to be affected by noise. Then, the inevitable presence of noise in data prevents from representing equivalently the actual system and induces rather a set of systems that could have generated the noisy data for a given bound on the noise, i.e., the set of systems consistent with data. This set, which we call $\mathcal{C}$, plays a central role since control design must therefore target all systems in $\mathcal{C}$, which are indistinguishable from each other based on data. We consider noise in data in the form of process disturbance, but the approach could be extended to genuine measurement noise (cf. Remark 3).

A natural way to address this uncertainty induced by noisy data is via robust control tools: e.g., system level synthesis (Anderson, Doyle, Low, & Matni, 2019; Dean, Mania, Matni, Recht, & Tu, 2020; Xue & Matni, 2021), Young's inequality for matrices (De Persis & Tesi, 2020), matrix generalizations of the S-procedure (Ferizbegovic, Umenberger, Hjalmarsson, & Schön, 2019; van Waarde et al., 2021), Farkas's lemma (Dai & Sznaier, 2018, 2021), linear fractional transformations (Berberich, Romer, Scherer, & Allgöwer, 2020; Berberich, Scherer, & Allgöwer, 2020).

☆ This research is partially supported by NWO, The Netherlands, project no. 15472 and by a Marie Skłodowska-Curie COFUND grant, no. 754315. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Valery Ugrinovskii under the direction of Editor Ian R. Petersen.

* Corresponding author.
*E-mail addresses:* a.bisoffi@rug.nl (A. Bisoffi), c.de.persis@rug.nl (C. De Persis), pietro.tesi@unifi.it (P. Tesi).

https://doi.org/10.1016/j.automatica.2022.110537
0005-1098/

We advocate here the use of another robust control tool for data-driven control, Petersen's lemma (Petersen, 1987; Petersen & Hollot, 1986). This lemma, whose strict and nonstrict versions we report later in Facts 1 and 2, can be seen as a matrix elimination method since, instead of verifying for all matrices bounded in norm a certain inequality, one can *equivalently* verify another inequality where such matrices do not appear. The utility of Petersen's lemma in the realm of robust control has been featured in Ji and Su (2016), Khlebnikov and Shcherbakov (2008) and Shcherbakov and Topunov (2008). Petersen's lemma underpins the data-based results of this work and its main appealing feature is its broad applicability to different classes of systems, such as linear and polynomial ones; it also provides conceptual insights on the rationale of the designed control law and its relation with least-squares approaches and certainty equivalence (cf. Section 4.2).

### 1.2. Contributions

Our main contributions are the following. *(C1)* We bring Petersen's lemma to the attention as a powerful tool for data-driven control. *(C2)* For linear systems, we provide by it necessary and sufficient conditions for quadratic stabilization, which are alternative to those in van Waarde et al. (2021). These conditions take the convenient form of linear matrix inequalities. *(C3)* We give several insights on the design conditions and, in particular, establish connections with certainty equivalence and robust indirect control, which have been extensively investigated for stochastic noise models, e.g., Dean et al. (2020), Ferizbegovic et al. (2019) and Treven, Curi, Mutnỳ, and Krause (2021). *(C4)* For polynomial systems, we obtain new sufficient conditions for data-driven control with respect to Dai and Sznaier (2021) and Guo, De Persis, and Tesi (2021). These conditions are tractably relaxed into alternate (convex) sum-of-squares programs.

### 1.3. Relations with the literature

We assume an upper bound on the norm of the sequence of process disturbances, which is the so-called *unknown-but-bounded* disturbance paradigm (Hjalmarsson & Ljung, 1993). This makes our approach different from those considering stochastic noise descriptions (Dean et al., 2020; Ferizbegovic et al., 2019; Recht, 2019; Treven et al., 2021) and similar in nature to set-membership identification and control (Fogel, 1979; Milanese & Novara, 2004; Tanaskovic, Fagiano, Novara, & Morari, 2017). The use of robust control tools to counteract the uncertainty induced by unknown-but-bounded noise is quite natural and has been pursued in Berberich, Romer, Scherer, and Allgöwer (2020), Berberich, Scherer, and Allgöwer (2020), Dai and Sznaier (2021), De Persis and Tesi (2020), Guo et al. (2021) and van Waarde et al. (2021). Next, we compare with these works referring to our aforementioned contributions *(C1)–(C4)*.

*(C1)* The use of Petersen's lemma differentiates our approach from those in Berberich, Romer, Scherer, and Allgöwer (2020), Berberich, Scherer, and Allgöwer (2020), De Persis and Tesi (2020) and van Waarde et al. (2021), which also address data-driven stabilization of linear systems (besides $\mathcal{H}_2$, $\mathcal{H}_\infty$ or quadratic performance). In Bisoffi, De Persis, and Tesi (2020), we used Petersen's lemma only as a sufficient condition (Bisoffi et al., 2020, Fact 1) to obtain a data-driven controller for structurally different bilinear systems.

*(C2)* For linear systems in discrete time, van Waarde et al. (2021) provided necessary and sufficient conditions for data-based stabilization as we do here. The differences are illustrated in detail in Section 4.3. In a nutshell, here we operate under an easy-to-enforce condition stemming from persistence of excitation instead of under a generalized Slater condition, and the former (but not the latter) can be seamlessly satisfied also in the relevant special case of ideal data (i.e., without noise).

*(C3)* For the considered noise setting, the uncertainty set $\mathcal{C}$ consists in a *matrix ellipsoid*, whose center is the (ordinary) least-squares estimate of the system dynamics, and whose size depends on the noise bound. This justifies why *certainty-equivalence* control can be expected to work well in regimes of small uncertainty (small noise), which agrees with recent works on performance of certainty-equivalence control for linear quadratic control (Dörfler, Tesi, & De Persis, 2021; Mania, Tu, & Recht, 2019). On the other hand, this also explains why robust design is generally needed to have stability guarantees, which is also the main idea behind the *robust indirect* control approaches (Dean et al., 2020; Ferizbegovic et al., 2019; Treven et al., 2021) under a stochastic noise description. On a related note, we introduced the notion of matrix ellipsoid in Bisoffi, De Persis, and Tesi (2021c), which had however a quite different focus and research question.

*(C4)* Data-driven control of polynomial systems was proposed also in Dai and Sznaier (2021) and Guo et al. (2021). As in Guo et al. (2021), we use Lyapunov methods to obtain sufficient conditions for data-based global asymptotic stabilization. Whereas Guo et al. (2021) parametrizes the Lyapunov function in a specific way to obtain a convex sum-of-squares program, the present data-based conditions parallel naturally the classical model-based ones in Khalil (2002) since they correspond to enforcing those model-based conditions (through Petersen's lemma) for all systems consistent with data, which leads to succinct derivations. Due to this natural parallel, the present approach appears to be extendible with appropriate modifications to other cases where Lyapunov(-like) conditions occur, as we do for *local* asymptotic stabilization in Corollary 3. On the other hand, Dai and Sznaier (2021) follows a radically different approach. Instead of Lyapunov functions, it uses *density functions* by Rantzer (2001) to give a necessary and sufficient condition for data-based stabilization, which however needs to be relaxed into a quadratically-constrained quadratic program through sum of squares and then into a semidefinite program through moment-based techniques for tractability.

### 1.4. Structure

In Section 2, we recall Petersen's lemma, formulate the problem and derive some properties of the set $\mathcal{C}$. In Section 3 we provide our main results for linear systems and we comment the results in Section 4. In Section 5 we provide our main result for polynomial systems. All results are exemplified numerically in Section 6.

## 2. Preliminaries and problem setting

### 2.1. Notation and Petersen's lemma

For a vector $a$, $|a|$ denotes its 2-norm. For a matrix $A$, $\|A\|$ denotes its induced 2-norm, which is equivalent to the largest singular value of $A$; moreover, for a scalar $a \geq 0$, $\|A\| \leq a$ if and only if $A^\top A \preceq a^2 I$ where $I$ is the identity matrix. For matrices $A$, $B$ and $C$ of compatible dimensions, we abbreviate $ABC(AB)^\top$ to $AB \cdot C[\star]^\top$, where the dot in the second expression clarifies unambiguously that $AB$ are the terms to be transposed. For matrices $A = A^\top$, $B$, $C = C^\top$, we also abbreviate the symmetric matrix $\left[\begin{smallmatrix} A & B \\ B^\top & C \end{smallmatrix}\right]$ as $\left[\begin{smallmatrix} A & B \\ \star & C \end{smallmatrix}\right]$ or $\left[\begin{smallmatrix} A & \star \\ B^\top & C \end{smallmatrix}\right]$. For a positive semidefinite matrix $A$, $A^{1/2}$ denotes the unique positive semidefinite root of $A$. For a matrix $A$, $A^\dagger$ denotes the Moore–Penrose generalized inverse of $A$, which is uniquely determined by certain axioms (Horn & Johnson, 2013, p. 453, 7.3.P7). For positive integers $n$, $r$ and the set $\mathcal{P}$ of polynomials $p:\mathbb{R}^n \to \mathbb{R}$ (resp., the set $\mathcal{P}_{\mathrm{m}}$ of matrix polynomials $P:\mathbb{R}^n \to \mathbb{R}^{r\times r}$), the set $\mathcal{S} \subset \mathcal{P}$ (resp. $\mathcal{S}_{\mathrm{m}} \subset \mathcal{P}_{\mathrm{m}}$) denotes the set of sum-of-squares polynomials (resp., the set of sum-of-squares matrix polynomials) in the variable $x \in \mathbb{R}^n$; see Chesi (2010) and references therein for more details on these and other sum-of-squares notions.

Petersen's lemma is the essential tool we use to address data-driven control design. First, we present in the next fact a version where inequalities are strict.

**Fact 1** (*Strict Petersen's Lemma*). *Consider matrices* $\mathbf{C} \in \mathbb{R}^{n\times n}$, $\mathbf{E} \in \mathbb{R}^{n\times p}$, $\bar{\mathbf{F}} \in \mathbb{R}^{q\times q}$, $\mathbf{G} \in \mathbb{R}^{q\times n}$ *with* $\mathbf{C} = \mathbf{C}^\top$ *and* $\bar{\mathbf{F}} = \bar{\mathbf{F}}^\top \succeq 0$, *and let* $\mathcal{F}$ *be*

$$\mathcal{F} := \{\mathbf{F} \in \mathbb{R}^{p\times q} : \mathbf{F}^\top\mathbf{F} \preceq \bar{\mathbf{F}}\}. \tag{1}$$

*Then,*

$$\mathbf{C} + \mathbf{E}\mathbf{F}\mathbf{G} + \mathbf{G}^\top\mathbf{F}^\top\mathbf{E}^\top \prec 0 \text{ for all } \mathbf{F} \in \mathcal{F} \tag{2a}$$

*if and only if there exists* $\lambda > 0$ *such that*

$$\mathbf{C} + \lambda\mathbf{E}\mathbf{E}^\top + \lambda^{-1}\mathbf{G}^\top\bar{\mathbf{F}}\mathbf{G} \prec 0. \tag{2b}$$

For $\bar{\mathbf{F}} = I$, one obtains the original version by I. R. Petersen in Petersen (1987), Petersen and Hollot (1986), and the version in Fact 1 proposes a slight extension where the bound $\bar{\mathbf{F}}$ is any positive semidefinite matrix. For this version, then, we give the proof in the Appendix for completeness. Although one could prove Fact 1 with S-procedure arguments as some authors do for *non-strict* versions (Khlebnikov & Shcherbakov, 2008; Shcherbakov & Topunov, 2008), we follow the original proof strategy of Petersen (1987) and Petersen and Hollot (1986).

Second, we present in the next fact a version of Petersen's lemma where inequalities are nonstrict.

**Fact 2** (*Nonstrict Petersen's Lemma*). *Consider matrices* $\mathbf{C} \in \mathbb{R}^{n\times n}$, $\mathbf{E} \in \mathbb{R}^{n\times p}$, $\bar{\mathbf{F}} \in \mathbb{R}^{q\times q}$, $\mathbf{G} \in \mathbb{R}^{q\times n}$ *with* $\mathbf{C} = \mathbf{C}^\top$ *and* $\bar{\mathbf{F}} = \bar{\mathbf{F}}^\top \succeq 0$, *and let* $\mathcal{F}$ *be defined as in (1). Suppose additionally* $\mathbf{E} \neq 0$, $\bar{\mathbf{F}} \succ 0$ *and* $\mathbf{G} \neq 0$. *Then,*

$$\mathbf{C} + \mathbf{E}\mathbf{F}\mathbf{G} + \mathbf{G}^\top\mathbf{F}^\top\mathbf{E}^\top \preceq 0 \text{ for all } \mathbf{F} \in \mathcal{F} \tag{3a}$$

*if and only if there exists* $\lambda > 0$ *such that*

$$\mathbf{C} + \lambda\mathbf{E}\mathbf{E}^\top + \lambda^{-1}\mathbf{G}^\top\bar{\mathbf{F}}\mathbf{G} \preceq 0. \tag{3b}$$

*Moreover, (3b) implies (3a)* without *the assumption* $\mathbf{E} \neq 0$, $\bar{\mathbf{F}} \succ 0$ *and* $\mathbf{G} \neq 0$.

For $\bar{\mathbf{F}} = I$, one obtains precisely the nonstrict versions of Petersen's lemma in Khlebnikov and Shcherbakov (2008, §2) and Shcherbakov and Topunov (2008, §2); for completeness we then report the proof of Fact 2 in Bisoffi, De Persis, and Tesi (2021a). The additional assumption with respect to Fact 1 (i.e., $\mathbf{E} \neq 0$, $\bar{\mathbf{F}} \succ 0$ and $\mathbf{G} \neq 0$) is due to having *nonstrict* inequalities and is needed to obtain the specific form (3b), see Bisoffi et al. (2021a).

### 2.2. Problem formulation

Consider a discrete-time linear time-invariant system

$$x^+ = A_\star x + B_\star u + d \tag{4}$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ is the input, $d \in \mathbb{R}^n$ is a disturbance, and the matrices $A_\star$ and $B_\star$ are *unknown* to us. At the same time and with the same meaning for the quantities $x$, $u$ and $d$, consider the continuous-time linear time-invariant system

$$\dot{x} = A_\star x + B_\star u + d. \tag{5}$$

The modifications required for the continuous-time case are limited, and this allows us to treat it in parallel to the discrete-time case. Instead of relying on model knowledge given by $A_\star$ and $B_\star$, we perform an experiment on the system by applying an input sequence $u(t_0)$, $u(t_1)$, …, $u(t_{T-1})$ of $T$ samples, so that by (4)/(5)

$$x(t_{i+1})/\dot{x}(t_i) = A_\star x(t_i) + B_\star u(t_i) + d(t_i)$$

for $i = 0, \dots, T-1$. We measure the state response $x(t_0)$, $x(t_1)$, …, $x(t_{T-1})$, and, in discrete time, the shifted state response $x(t_1)$, $x(t_2)$, …, $x(t_T)$ or, in continuous time, the state-derivative response $\dot{x}(t_0)$, $\dot{x}(t_1)$, …, $\dot{x}(t_{T-1})$. The disturbance sequence $d(t_0)$, $d(t_1)$, …, $d(t_{T-1})$ affects the evolution of the system and is *unknown*, hence data are *noisy*. We collect the noisy data in the matrices

$$U_0 := \begin{bmatrix} u(t_0) & u(t_1) & \cdots & u(t_{T-1}) \end{bmatrix} \tag{6a}$$

$$X_0 := \begin{bmatrix} x(t_0) & x(t_1) & \cdots & x(t_{T-1}) \end{bmatrix} \tag{6b}$$

$$X_1 := \begin{bmatrix} x(t_1) & x(t_2) & \cdots & x(t_T) \end{bmatrix} \text{ in discrete time, or} \tag{6c}$$

$$X_1 := \begin{bmatrix} \dot{x}(t_0) & \dot{x}(t_1) & \cdots & \dot{x}(t_{T-1}) \end{bmatrix} \text{ in continuous time.} \tag{6d}$$

We can also arrange the unknown disturbance sequence as $D_0 := \begin{bmatrix} d(t_0) & d(t_1) & \cdots & d(t_{T-1}) \end{bmatrix}$, so that $D_0$ and data in (6) satisfy

$$X_1 = A_\star X_0 + B_\star U_0 + D_0 \tag{7}$$

since (4) (in discrete time) or (5) (in continuous time) is the underlying data generation mechanism. In the former case, we have $t_0$, $t_1$, …, $t_T$ equal to, respectively, 0, 1, …, $T$; in the latter case, $t_0, t_1, \dots, t_{T-1}$ are sampled periodically at $0, T_s, \dots, (T-1)\cdot T_s$ for some sampling time $T_s$, although this is not necessary.

We operate under a certain disturbance model. Specifically, we assume that the disturbance sequence $D_0$ has bounded energy, i.e., $D_0 \in \mathcal{D}$ where, for some matrix $\Delta$,

$$\mathcal{D} := \{D \in \mathbb{R}^{n\times T} : DD^\top \preceq \Delta\Delta^\top\}. \tag{8}$$

As we said, $D_0$ is unknown to us and the only a-priori knowledge on it is given by the set $\mathcal{D}$, and in particular the knowledge of the positive semidefinite bound $\Delta\Delta^\top$. This disturbance model enforces an energy bound on the disturbance since it constrains the whole disturbance sequence, unlike an instantaneous disturbance bound (Bisoffi et al., 2021c). Energy bounds are used in Berberich, Romer, Scherer, and Allgöwer (2020), Berberich, Scherer, and Allgöwer (2020), De Persis and Tesi (2020), van Waarde et al. (2021) and many other works. In fact, model (8) is quite general as it can capture signal-to-noise-ratio conditions (De Persis & Tesi, 2020), over-approximate instantaneous bounds (Bisoffi et al., 2021c), and can also be used to have probabilistic bounds for Gaussian noise (De Persis & Tesi, 2021).

With data (6) and set $\mathcal{D}$ in (8), we introduce the set $\mathcal{C}$ of matrices *consistent with data*

$$\mathcal{C} := \left\{ \begin{bmatrix} A & B \end{bmatrix} : X_1 = AX_0 + BU_0 + D, D \in \mathcal{D} \right\}, \tag{9}$$

i.e., the set of all pairs $\begin{bmatrix} A & B \end{bmatrix}$ of matrices that could generate data $X_1$, $X_0$ and $U_0$ based on (4) or (5) while keeping the disturbance sequence in the set $\mathcal{D}$. This is elucidated by comparing (9) with the similar (7). We note that $D_0 \in \mathcal{D}$ is precisely equivalent to $\begin{bmatrix} A_\star & B_\star \end{bmatrix} \in \mathcal{C}$.

**Remark 1.** In the language of set-membership identification (Milanese & Novara, 2004), we have two *prior assumptions*, the first one on the class of dynamical systems (4) or (5) and the second one on the noise (8). The set $\mathcal{C}$ in (9) corresponds to the *feasible systems set* (Milanese & Novara, 2004, Def. 1). We noted that $\begin{bmatrix} A_\star & B_\star \end{bmatrix} \in \mathcal{C}$. This corresponds to *validation of prior assumptions* (Milanese & Novara, 2004, Def. 2).

Our objective is to design a state feedback controller

$$u = Kx$$

that makes the closed-loop matrix $A_\star + B_\star K$ Schur stable (i.e., all its eigenvalues have magnitude less than 1) in discrete time, or Hurwitz stable (i.e., all its eigenvalues have real part less than 0) in continuous time. However, we lack the knowledge of $\begin{bmatrix} A_\star & B_\star \end{bmatrix}$ and the disturbance $d$ induces uncertainty in data, which results into a *set* $\mathcal{C}$ of matrices consistent with data. Our objective becomes then to stabilize robustly all matrices $A + BK$ for $\begin{bmatrix} A & B \end{bmatrix} \in \mathcal{C}$; in other words, in discrete time,

$$\text{find} \quad K,\ P = P^\top \succ 0 \tag{10a}$$

$$\text{s. t.} \quad (A+BK)P(A+BK)^\top - P \prec 0 \text{ for all } \begin{bmatrix} A & B \end{bmatrix} \in \mathcal{C} \tag{10b}$$

or, in continuous time,

$$\text{find} \quad K,\ P = P^\top \succ 0 \tag{11a}$$

$$\text{s. t.} \quad (A+BK)P + P(A+BK)^\top \prec 0 \text{ for all } \begin{bmatrix} A & B \end{bmatrix} \in \mathcal{C}. \tag{11b}$$

Both (10) and (11) are quadratic stabilization problems. Achieving the objective of robust stabilization of all matrices $A + BK$ for $\begin{bmatrix} A & B \end{bmatrix} \in \mathcal{C}$ (hence, also of $A_\star + B_\star K$) guarantees bounded-input bounded-state stability of $x^+ = (A_\star + B_\star K)x + d$ or $\dot{x} = (A_\star + B_\star K)x + d$ by Antsaklis and Michel (2006, Thm. 9.5).

### 2.3. Reformulations of set $\mathcal{C}$ and properties

We perform some rearrangements of $\mathcal{C}$. We substitute in (9) the definition of set $\mathcal{D}$ in (8) and obtain

$$\mathcal{C} = \Big\{ \begin{bmatrix} A & B \end{bmatrix} : X_1 = AX_0 + BU_0 + D, D \in \mathbb{R}^{n\times T}, \begin{bmatrix} I & D \end{bmatrix} \begin{bmatrix} -\Delta\Delta^\top & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I \\ D^\top \end{bmatrix} \preceq 0 \Big\}.$$

In this expression we substitute $D = X_1 - AX_0 - BU_0$ in the matrix inequality and collect $\begin{bmatrix} I & A & B \end{bmatrix}$ on the left and its transpose on the right of the matrix inequality; then, $\mathcal{C}$ rewrites equivalently as

$$\mathcal{C} = \Big\{ \begin{bmatrix} A & B \end{bmatrix} : \begin{bmatrix} I & A & B \end{bmatrix} \cdot \left[ \begin{array}{c|c} \mathbf{C} & \mathbf{B}^\top \\ \hline \mathbf{B} & \mathbf{A} \end{array} \right] [\star]^\top \preceq 0 \Big\} \tag{12}$$

$$= \Big\{ \begin{bmatrix} A & B \end{bmatrix} = Z^\top : \mathbf{C} + \mathbf{B}^\top Z + Z^\top \mathbf{B} + Z^\top \mathbf{A} Z \preceq 0 \Big\} \tag{13}$$

where we define

$$\left[ \begin{array}{c|c} \mathbf{C} & \mathbf{B}^\top \\ \hline \mathbf{B} & \mathbf{A} \end{array} \right] := \left[ \begin{array}{c|c} X_1 X_1^\top - \Delta\Delta^\top & -X_1 \begin{bmatrix} X_0 \\ U_0 \end{bmatrix}^\top \\ \hline -\begin{bmatrix} X_0 \\ U_0 \end{bmatrix} X_1^\top & \begin{bmatrix} X_0 \\ U_0 \end{bmatrix} \begin{bmatrix} X_0 \\ U_0 \end{bmatrix}^\top \end{array} \right]. \tag{14}$$

**Remark 2.** For given matrices $R = R^\top$, $Q = Q^\top \succ 0$, $S$, one can consider a disturbance model $\mathcal{D}' := \{D \in \mathbb{R}^{n\times T} : \begin{bmatrix} I & D \end{bmatrix} \begin{bmatrix} R & S^\top \\ S & Q \end{bmatrix} \begin{bmatrix} I \\ D^\top \end{bmatrix} \preceq 0\}$ more general than $\mathcal{D}$ in (8), as in Berberich, Romer, Scherer, and Allgöwer (2020), Berberich, Scherer, and Allgöwer (2020) and van Waarde et al. (2021). With $\mathcal{D}'$, one can still carry out the derivations for a set $\mathcal{C}'$ similar to (12), with slightly different expressions of $\mathbf{A}'$, $\mathbf{B}'$, $\mathbf{C}'$. Nonetheless, (8) is general enough to capture interesting classes of noise, see the discussion after (8).

We make the next assumption on matrix $\begin{bmatrix} X_0 \\ U_0 \end{bmatrix}$ in (14).

**Assumption 1.** Matrix $\begin{bmatrix} X_0 \\ U_0 \end{bmatrix}$ has full row rank.

Assumption 1 is related to persistence of excitation as we illustrate in Section 4.1, and can be checked directly from data. If this condition holds, it implies $T \geq n + m$; otherwise, it can typically be enforced by collecting more data points (i.e., adding more columns to $\begin{bmatrix} X_0 \\ U_0 \end{bmatrix}$). An immediate consequence of Assumption 1 for $\mathbf{A}$ in (14) is $\mathbf{A} \succ 0$.

**Remark 3.** We consider here the case of process disturbance, see (4). If, in addition, measurement noise is present, (4) needs to be combined with $\xi = x + \nu$, so that we no longer measure $x$ but $\xi$, which is corrupted by measurement noise $\nu$. By (4) and $\xi = x + \nu$, the data generation mechanism in terms of $\xi$ becomes $\xi^+ = A_\star \xi + B_\star u + d + \nu^+ - A_\star \nu$ and the data points satisfy $\Xi_1 = A_\star \Xi_0 + B_\star U_0 + D_0 + N_1 - A_\star N_0$ for $\Xi_0$ and $\Xi_1$ analogous to $X_0$ and $X_1$, and the unknown $N_1 := \begin{bmatrix} \nu(t_1) & \cdots & \nu(t_T) \end{bmatrix}$ and $N_0 := \begin{bmatrix} \nu(t_0) & \cdots & \nu(t_{T-1}) \end{bmatrix}$. By following De Persis and Tesi (2020, §V-A), one can reduce this case to the case of process disturbances. In particular, based on the relation $\Xi_1 = A_\star \Xi_0 + B_\star U_0 + D_0 + N_1 - A_\star N_0$, it is possible to construct a set $\mathcal{C}$ analogous to (9) and, under an assumption slightly more conservative than Assumption 1 and checkable from data, to pursue the same approach as in the sequel.

The set $\mathcal{C}$ in (13) can be regarded as a *matrix* ellipsoid, i.e., a natural extension of the standard (vector) ellipsoid (Boyd, El Ghaoui, Feron, & Balakrishnan, 1994, p. 42) with parameters $\mathbf{c} \in \mathbb{R}$, $\mathbf{b} \in \mathbb{R}^p$, $\mathbf{a} \in \mathbb{R}^{p\times p}$:

$$\{z \in \mathbb{R}^p : \mathbf{c} + \mathbf{b}^\top z + z^\top \mathbf{b} + z^\top \mathbf{a} z \leq 0\}.$$

In fact, if a scalar system with $n = m = 1$ is considered, $\mathcal{C}$ reduces to a standard ellipsoid with $Z^\top \in \mathbb{R}^2$. The interpretation of $\mathcal{C}$ as a matrix ellipsoid (introduced in Bisoffi et al., 2021c to compute a size for this set) proves useful here since it enables a simple reformulation of $\mathcal{C}$ as

$$\mathcal{C} = \Big\{ \begin{bmatrix} A & B \end{bmatrix} = Z^\top : (Z - \boldsymbol{\zeta})^\top \mathbf{A} (Z - \boldsymbol{\zeta}) \preceq \mathbf{Q} \Big\} \tag{15}$$

where, by $\mathbf{A} \succ 0$ from Assumption 1, we define

$$\boldsymbol{\zeta} := -\mathbf{A}^{-1}\mathbf{B}, \quad \mathbf{Q} := \mathbf{B}^\top \mathbf{A}^{-1} \mathbf{B} - \mathbf{C}, \tag{16}$$

as can be verified by substituting (16) into (15) and expanding all products to obtain (13). We will further discuss later in Section 4.2 the interpretation of some of the parameters $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\boldsymbol{\zeta}$, $\mathbf{Q}$ of $\mathcal{C}$. The matrix-ellipsoid parametrizations of $\mathcal{C}$ in (13), (15) and (19) are analogous to the parametrizations of a standard ellipsoid as, respectively, a quadratic form (Boyd et al., 1994, Eq. (3.8)), as a center and shape matrix and as a linear transformation of a unit ball (Boyd et al., 1994, Eq. (3.9)). We report the sign definiteness of $\mathbf{A}$ and $\mathbf{Q}$ in the next lemma.

**Lemma 1.** *Under Assumption 1, $\mathbf{A} \succ 0$ and $\mathbf{Q} \succeq 0$.*

**Proof.** $\mathbf{A} \succ 0$ from Assumption 1 by Horn and Johnson (2013, Thm. 7.2.7(c)). As for $\mathbf{Q}$, $\mathbf{A} \succ 0$ allows defining

$$\mathbf{Q}_\mathrm{p} := \begin{bmatrix} X_0 \\ U_0 \end{bmatrix}^\top \left( \begin{bmatrix} X_0 \\ U_0 \end{bmatrix} \begin{bmatrix} X_0 \\ U_0 \end{bmatrix}^\top \right)^{-1} \begin{bmatrix} X_0 \\ U_0 \end{bmatrix}.$$

$\mathbf{Q}_\mathrm{p}$ is a projection matrix, i.e., $\mathbf{Q}_\mathrm{p}^2 = \mathbf{Q}_\mathrm{p}$ as one verifies immediately. Then, (16) and (14) yield

$$\mathbf{Q} = X_1 \mathbf{Q}_\mathrm{p} X_1^\top - X_1 X_1^\top + \Delta\Delta^\top. \tag{17}$$

Write (7) as $X_1 = \begin{bmatrix} A_\star & B_\star \end{bmatrix} \begin{bmatrix} X_0 \\ U_0 \end{bmatrix} + D_0$; this expression and $\mathbf{Q}_\mathrm{p}$ being a projection matrix shows that $\mathbf{Q}_\mathrm{p} X_1^\top - X_1^\top = \mathbf{Q}_\mathrm{p} ( \begin{bmatrix} X_0 \\ U_0 \end{bmatrix}^\top \begin{bmatrix} A_\star & B_\star \end{bmatrix}^\top +$

$D_0^\top) - (\begin{bmatrix} X_0 \\ U_0 \end{bmatrix}^\top [A_\star \; B_\star]^\top + D_0^\top) = (\mathbf{Q}_\mathrm{p} - I)D_0^\top$. Using this expression in (17) yields

$$\mathbf{Q} = D_0(\mathbf{Q}_\mathrm{p} - I)D_0^\top + \Delta\Delta^\top \succeq -D_0 D_0^\top + \Delta\Delta^\top \succeq 0 \tag{18}$$

since $\mathbf{Q}_\mathrm{p} \succeq 0$ and $D_0 \in \mathcal{D}$ (i.e., $D_0 D_0^\top \preceq \Delta\Delta^\top$). □

From $\mathbf{A} \succ 0$, we have the next desirable property of $\mathcal{C}$.

**Lemma 2.** *Under Assumption 1, $\mathcal{C}$ is bounded with respect to any matrix norm.*

**Proof.** Consider $\mathcal{C}$ in (15), which is nonempty as $\boldsymbol{\zeta}^\top \in \mathcal{C}$. $Z^\top \in \mathcal{C}$ if and only if for all $v \in \mathbb{R}^n$, $v^\top (Z - \boldsymbol{\zeta})^\top \mathbf{A}(Z - \boldsymbol{\zeta})v \le v^\top \mathbf{Q} v$. Let $\lambda_{\min}(\mathbf{A})$ denote the minimum eigenvalue of (symmetric) $\mathbf{A}$. By Lemma 1, this implies

$$\begin{aligned}
&\sqrt{\lambda_{\min}(\mathbf{A})}|(Z - \boldsymbol{\zeta})v| \le |\mathbf{Q}^{1/2}v| \text{ for all } v : |v| = 1 \\
&\implies \sqrt{\lambda_{\min}(\mathbf{A})} \sup_{|v|=1} |(Z - \boldsymbol{\zeta})v| \le \sup_{|v|=1} |\mathbf{Q}^{1/2}v| \\
&\implies \|Z - \boldsymbol{\zeta}\| \le \lambda_{\min}(\mathbf{A})^{-1/2}\|\mathbf{Q}^{1/2}\| \\
&\implies \|Z\| \le \|\boldsymbol{\zeta}\| + \lambda_{\min}(\mathbf{A})^{-1/2}\|\mathbf{Q}^{1/2}\|
\end{aligned}$$

where we used the definition of induced 2-norm and the reverse triangle inequality in the second and third implication, respectively. All quantities on the right hand side are finite, so each $Z^\top \in \mathcal{C}$ has bounded 2-norm. Recall that any two matrix norms are equivalent (Horn & Johnson, 2013, p. 352), so for any given pair of matrix norms $\|\cdot\|_a$ and $\|\cdot\|_b$, there is a finite constant $C_{ab} > 0$ such that $\|M\|_a \le C_{ab}\|M\|_b$ for all matrices $M$. Hence, boundedness of $\mathcal{C}$ with respect to the induced 2-norm implies boundedness of $\mathcal{C}$ with respect to any other norm, as needed proving. □

## 3. Data-driven control for linear systems

So far, we have rewritten the set $\mathcal{C}$ of matrices $[A \;\; B]$ consistent with data as (15). To derive the main result from Petersen's lemma, a final reformulation of $\mathcal{C}$ is needed. We define

$$\mathcal{E} := \left\{(\boldsymbol{\zeta} + \mathbf{A}^{-1/2}\Upsilon\mathbf{Q}^{1/2})^\top : \|\Upsilon\| \le 1\right\} \tag{19}$$

and show that it coincides with $\mathcal{C}$ in the next proposition.

**Proposition 1.** *For $\mathbf{A} \succ 0$ and $\mathbf{Q} \succeq 0$, $\mathcal{C} = \mathcal{E}$.*

**Proof.** It is sufficient to prove $\mathcal{E} \subseteq \mathcal{C}$ and $\mathcal{C} \subseteq \mathcal{E}$.
($\mathcal{E} \subseteq \mathcal{C}$) Suppose $Z^\top \in \mathcal{E}$, i.e., $Z = \boldsymbol{\zeta} + \mathbf{A}^{-1/2}\Upsilon\mathbf{Q}^{1/2}$ for some matrix $\Upsilon$ with $\|\Upsilon\| \le 1$. Hence, $(Z - \boldsymbol{\zeta})^\top \mathbf{A}(Z - \boldsymbol{\zeta}) = (\mathbf{A}^{-1/2}\Upsilon\mathbf{Q}^{1/2})^\top \mathbf{A}(\mathbf{A}^{-1/2}\Upsilon\mathbf{Q}^{1/2}) = \mathbf{Q}^{1/2}\Upsilon^\top\Upsilon\mathbf{Q}^{1/2} \preceq \mathbf{Q}$. Thus $Z^\top \in \mathcal{C}$.
($\mathcal{C} \subseteq \mathcal{E}$) Suppose $Z^\top \in \mathcal{C}$, i.e.,

$$(Z - \boldsymbol{\zeta})^\top \mathbf{A}(Z - \boldsymbol{\zeta}) \preceq \mathbf{Q}. \tag{20}$$

We need to find a matrix $\Upsilon$ with $\|\Upsilon\| \le 1$ such that $Z = \boldsymbol{\zeta} + \mathbf{A}^{-1/2}\Upsilon\mathbf{Q}^{1/2}$, i.e.,

$$\Upsilon\mathbf{Q}^{1/2} = \mathbf{A}^{1/2}(Z - \boldsymbol{\zeta}). \tag{21}$$

If $\mathbf{Q}^{1/2} = 0$, we can take $\Upsilon = 0$. Otherwise, $\mathbf{Q}^{1/2}$ has $p \in \{1, \dots, n\}$ positive eigenvalues that define the diagonal matrix $\Lambda_p := \operatorname{diag}(\lambda_1, \dots, \lambda_p) \succ 0$. Since $\mathbf{Q}^{1/2}$ is symmetric, there exists a real orthogonal matrix $T$ (i.e., $T^\top T = TT^\top = I$) such that

$$\mathbf{Q}^{1/2} = T\Lambda T^\top := T \begin{bmatrix} \Lambda_p & 0 \\ 0 & 0 \end{bmatrix} T^\top, \tag{22}$$

which is an eigendecomposition of $\mathbf{Q}^{1/2}$ and admits $\Lambda = \Lambda_p$ if $p = n$ (i.e., $\mathbf{Q}^{1/2} \succ 0$). Writing $T =: [T_1 \;\; T_2]$ yields

$$\begin{bmatrix} T_1^\top T_1 & T_1^\top T_2 \\ T_2^\top T_1 & T_2^\top T_2 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \text{ and } T_1 T_1^\top + T_2 T_2^\top = I \tag{23}$$

from $T^\top T = I$ and $TT^\top = I$. Select

$$\Upsilon = \mathbf{A}^{1/2}(Z - \boldsymbol{\zeta})T_1\Lambda_p^{-1}T_1^\top \tag{24}$$

(which reduces to $\mathbf{A}^{1/2}(Z - \boldsymbol{\zeta})\mathbf{Q}^{-1/2}$ if $p = n$). We first show $\|\Upsilon\| \le 1$:

$$\begin{aligned}
\Upsilon^\top\Upsilon &= T_1\Lambda_p^{-1}T_1^\top(Z - \boldsymbol{\zeta})^\top\mathbf{A}^{1/2}\mathbf{A}^{1/2}(Z - \boldsymbol{\zeta})T_1\Lambda_p^{-1}T_1^\top \\
&\overset{(20)}{\preceq} T_1\Lambda_p^{-1}T_1^\top \cdot \mathbf{Q}[\star]^\top \overset{(22)}{=} T_1\Lambda_p^{-1}T_1^\top [T_1 \; T_2] \cdot \begin{bmatrix} \Lambda_p^2 & 0 \\ 0 & 0 \end{bmatrix} [\star]^\top \\
&= T_1\Lambda_p^{-1}T_1^\top T_1 \cdot \Lambda_p^2 [\star]^\top \overset{(23)}{=} T_1 T_1^\top \overset{(23)}{\preceq} I.
\end{aligned}$$

Then, we show that (21) holds. (21) is equivalent to

$$\begin{aligned}
&\Upsilon\mathbf{Q}^{1/2} \overset{(22)}{=} \Upsilon [T_1 \; T_2] \begin{bmatrix} \Lambda_p & 0 \\ 0 & 0 \end{bmatrix} T^\top = \mathbf{A}^{1/2}(Z - \boldsymbol{\zeta}) \\
&\iff [\Upsilon T_1 \Lambda_p \; 0] = \mathbf{A}^{1/2}(Z - \boldsymbol{\zeta})T = \mathbf{A}^{1/2}(Z - \boldsymbol{\zeta})[T_1 \; T_2] \\
&\iff \big(\Upsilon T_1\Lambda_p = \mathbf{A}^{1/2}(Z - \boldsymbol{\zeta})T_1, \; 0 = \mathbf{A}^{1/2}(Z - \boldsymbol{\zeta})T_2\big).
\end{aligned}$$

If we show the last two equalities, we have shown (21) and completed the proof. The first equality holds by the selection of $\Upsilon$ since $\Upsilon T_1\Lambda_p \overset{(24)}{=} \mathbf{A}^{1/2}(Z - \boldsymbol{\zeta})T_1\Lambda_p^{-1}T_1^\top T_1\Lambda_p \overset{(23)}{=} \mathbf{A}^{1/2}(Z - \boldsymbol{\zeta})T_1$. The second equality holds since the columns of $T_2$ are in $\ker \mathbf{Q}^{1/2}$ and $\ker \mathbf{Q}^{1/2} \subseteq \ker(\mathbf{A}^{1/2}(Z - \boldsymbol{\zeta}))$. The columns of $T_2$ are in $\ker \mathbf{Q}^{1/2}$ because $\mathbf{Q}^{1/2}T_2 \overset{(22)}{=} T \begin{bmatrix} \Lambda_p & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} T_1^\top \\ T_2^\top \end{bmatrix} T_2 \overset{(23)}{=} T \begin{bmatrix} \Lambda_p & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ I \end{bmatrix} = 0$; $\ker \mathbf{Q}^{1/2} \subseteq \ker(\mathbf{A}^{1/2}(Z - \boldsymbol{\zeta}))$ because, if $v$ satisfies $\mathbf{Q}^{1/2}v = 0$, then $0 = v^\top\mathbf{Q}v \overset{(20)}{\ge} v^\top(Z - \boldsymbol{\zeta})^\top\mathbf{A}(Z - \boldsymbol{\zeta})v = |\mathbf{A}^{1/2}(Z - \boldsymbol{\zeta})v|^2$, hence $\mathbf{A}^{1/2}(Z - \boldsymbol{\zeta})v = 0$. □

Considering $\mathbf{Q} \succeq 0$ rather than $\mathbf{Q} \succ 0$ is motivated since it allows us to include seamlessly the relevant special case of ideal data, namely, when the disturbance is not present. This corresponds indeed to $\Delta = 0$ and $\mathcal{D} = \{0\}$ in (8) and $\mathbf{Q} = 0$ in (18) by $D_0 \in \mathcal{D}$. With the equivalent parametrization $\mathcal{E}$ of set $\mathcal{C}$ and Petersen's lemma in Fact 1, we reach the next main result.

**Theorem 1.** *For data given by $U_0$, $X_0$, $X_1$ in (6) satisfying Assumption 1 and yielding $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ in (14), feasibility of (10) is equivalent to feasibility of*

$$\text{find} \quad Y, P = P^\top \succ 0 \tag{25a}$$

$$\text{s. t.} \quad \begin{bmatrix} -P - \mathbf{C} & 0 & \mathbf{B}^\top \\ 0 & -P & [P \;\; Y^\top] \\ \mathbf{B} & \begin{bmatrix} P \\ Y \end{bmatrix} & -\mathbf{A} \end{bmatrix} \prec 0. \tag{25b}$$

*If (25) is solvable, a controller gain is $K = YP^{-1}$.*

**Proof.** Thanks to Proposition 1, (10b) is equivalent to the fact that for all $[A \;\; B] \in \mathcal{E}$

$$\begin{aligned}
&(A + BK)P(A + BK)^\top - P \\
&= [A \;\; B] \begin{bmatrix} I \\ K \end{bmatrix} PP^{-1}P \begin{bmatrix} I \\ K \end{bmatrix}^\top [A \;\; B]^\top - P \prec 0.
\end{aligned}$$

Finding $P = P^\top \succ 0$, $K$ such that this matrix inequality holds for all $[A \;\; B] \in \mathcal{E}$ is equivalent to finding $P = P^\top \succ 0$, $Y$ such that

$$\begin{bmatrix} -P & -[A \; B]\begin{bmatrix} P \\ Y \end{bmatrix} \\ -\begin{bmatrix} P \\ Y \end{bmatrix}^\top [A \; B]^\top & -P \end{bmatrix} \prec 0 \text{ for all } [A \;\; B] \in \mathcal{E}, \tag{26}$$

by $P \succ 0$ and Schur complement. Note that, as claimed in the statement, $Y$ and $K$ are related by $Y = KP$, and $Y$ is preferred over $K$ as decision variable since $KP$ makes the matrix inequality nonlinear. $[A \;\; B] = Z^\top \in \mathcal{E}$ if and only if $Z = \boldsymbol{\zeta} + \mathbf{A}^{-1/2}\Upsilon\mathbf{Q}^{1/2}$ for some $\Upsilon$ with $\Upsilon^\top\Upsilon \preceq I$, by the parametrization in (19). Hence, (26) is true if and only if (27), which is displayed in Box I,

$$0 \succ \begin{bmatrix} -P & -(\zeta + \mathbf{A}^{-1/2}\Upsilon\mathbf{Q}^{1/2})^\top \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right] \\ \star & -P \end{bmatrix} = \begin{bmatrix} -P & -\zeta^\top \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right] \\ \star & -P \end{bmatrix} + \begin{bmatrix} 0 \\ -\left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right]\mathbf{A}^{-1/2} \end{bmatrix} \Upsilon \begin{bmatrix} \mathbf{Q}^{1/2} & 0 \end{bmatrix} + \begin{bmatrix} \mathbf{Q}^{1/2} \\ 0 \end{bmatrix} \Upsilon^\top \begin{bmatrix} 0 & -\mathbf{A}^{-1/2}\left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right] \end{bmatrix} \qquad (27)$$

**Box I.**

holds for all $\Upsilon$ with $\Upsilon^\top\Upsilon \preceq I$. (27) is written in a way that enables applying Petersen's lemma in Fact 1 with respect to the uncertainty $\Upsilon$. Indeed, simple computations yield that (27) holds for all $\Upsilon$ with $\Upsilon^\top\Upsilon \preceq I$ if and only if there exists $\lambda > 0$ such that

$$\begin{bmatrix} -P + \lambda^{-1}\mathbf{Q} & -\zeta^\top \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right] \\ -\left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right]^\top \zeta & -P + \lambda \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right]^\top \mathbf{A}^{-1} \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right] \end{bmatrix} \prec 0. \qquad (28)$$

In summary, we have so far that (10) is the same as

$$\text{find } Y, P = P^\top \succ 0, \lambda > 0 \text{ subject to (28)}. \qquad (29)$$

Multiply both sides of (28) by $\lambda > 0$ and "absorb" it in $P$ and $Y$, so that (29) is actually equivalent to

$$\text{find} \quad Y, P = P^\top \succ 0 \qquad (30a)$$

$$\text{s. t.} \quad \begin{bmatrix} -P + \mathbf{Q} & -\zeta^\top \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right] \\ -\left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right]^\top \zeta & -P + \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right]^\top \mathbf{A}^{-1} \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right] \end{bmatrix} \prec 0. \qquad (30b)$$

Substitute in (30b) $\zeta$ and $\mathbf{Q}$ as in (16) to obtain

$$\begin{bmatrix} -P + \mathbf{B}^\top\mathbf{A}^{-1}\mathbf{B} - \mathbf{C} & \mathbf{B}^\top\mathbf{A}^{-1} \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right] \\ \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right]^\top \mathbf{A}^{-1}\mathbf{B} & -P + \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right]^\top \mathbf{A}^{-1} \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right] \end{bmatrix}$$

$$= \begin{bmatrix} -P - \mathbf{C} & 0 \\ 0 & -P \end{bmatrix} + \begin{bmatrix} \mathbf{B}^\top \\ \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right]^\top \end{bmatrix} \mathbf{A}^{-1} \begin{bmatrix} \mathbf{B} & \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right] \end{bmatrix} \prec 0.$$

Take a Schur complement of this inequality and replace by it the one in (30b) to make (30) equivalent to (25). □

Similarly, we use the set $\mathcal{E}$ in (19) and Petersen's lemma reported in Fact 1 to resolve (11) in the next theorem.

**Theorem 2.** *For data given by $U_0$, $X_0$, $X_1$ in (6) satisfying Assumption 1 and yielding $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ in (14), feasibility of (11) is equivalent to feasibility of*

$$\textit{find} \quad Y, P = P^\top \succ 0 \qquad (31a)$$

$$\textit{s. t.} \quad \begin{bmatrix} -\mathbf{C} & \mathbf{B}^\top - \begin{bmatrix} P \\ Y \end{bmatrix}^\top \\ \mathbf{B} - \begin{bmatrix} P \\ Y \end{bmatrix} & -\mathbf{A} \end{bmatrix} \prec 0. \qquad (31b)$$

*If (31) is solvable, a controller gain is $K = YP^{-1}$.*

**Proof.** The proof follows the same reasoning of the proof of Theorem 1 and has somehow simplified steps since we do not need to first apply a Schur complement. It is thus omitted, but can be found in Bisoffi et al. (2021a). □

Suppose that the set $\mathcal{C}$ is given directly in the form (15) as a matrix-ellipsoid over-approximation of a less tractable set that is derived from data, which we discuss in Section 4.4. The next corollary suits this case and can be used instead of Theorems 1–2.

**Corollary 1.** *For the set $\mathcal{C} = \{[A \ \ B] = Z^\top : (Z-\zeta)^\top\mathbf{A}(Z-\zeta) \preceq \mathbf{Q}\}$ as in (15), assume $\mathbf{A} \succ 0$ and $\mathbf{Q} \succeq 0$. Then, feasibility of (10) (resp., (11)) is equivalent to feasibility of (25) (resp., (31)). If (25) (resp., (31)) is solvable, a controller gain is $K = YP^{-1}$.*

**Proof.** By assuming $\mathbf{A} \succ 0$ and $\mathbf{Q} \succeq 0$, one can follow the same steps used in the proofs of Theorems 1–2 to draw the same conclusions. □

**Remark 4.** Instead of parameters $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ of $\mathcal{C}$ in (12), we can write the conditions (25b) and (31b) in Theorems 1 and 2 in terms of $\mathbf{A}$, $\zeta$, $\mathbf{Q}$ of $\mathcal{C}$ in (15) as

$$\begin{bmatrix} -P+\mathbf{Q} & \star & \star \\ -\left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right]^\top \zeta & -P & \star \\ 0 & \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right] & -\mathbf{A} \end{bmatrix} \prec 0 \text{ and } \begin{bmatrix} \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right]^\top \zeta + \zeta^\top \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right] + \mathbf{Q} & \star \\ \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right] & -\mathbf{A} \end{bmatrix} \prec 0.$$

These conditions, which are obtained by Schur complement (see (30b)), are equivalent to (25b) and (31b), respectively.

Finally, (25) or (31) are feasibility programs and, when implemented, any feasible solution is returned. However, one can consider a cost criterion for the closed loop, e.g., the $\mathcal{H}_2$-norm of the transfer function from $d$ to some performance signal. This can be accommodated as in Dörfler et al. (2021, §II.A) by turning the feasibility program in (25) or (31) into an optimization program with that cost criterion.

## 4. Discussion and interpretations

This section is devoted to giving an overall interpretation of the previous developments.

### 4.1. *Assumption 1 and persistence of excitation*

Assumption 1 is intimately related to the notion of persistence of excitation, as we now motivate. With full details in De Persis and Tesi (2021, §4.2), the result (Willems et al., 2005, Cor. 2), which was given in the ideal case without disturbance $x^+ = A_\star x + B_\star u$, can show for the present case

$$x^+ = A_\star x + B_\star u + d = A_\star x + \begin{bmatrix} B_\star & I \end{bmatrix} \begin{bmatrix} u \\ d \end{bmatrix}$$

that: an input sequence and a disturbance sequence both persistently exciting of order $n+1$[1] imply that $\begin{bmatrix} X_0 \\ U_0 \\ D_0 \end{bmatrix}$ has full row rank and so has $\begin{bmatrix} X_0 \\ U_0 \end{bmatrix}$, as required in Assumption 1. In other words, Assumption 1 holds if the augmented input $\left[\begin{smallmatrix} u \\ d \end{smallmatrix}\right]$ is persistently exciting.

### 4.2. *Ellipsoidal uncertainty, least squares and certainty-equivalence control*

The discrete- and continuous-time stability conditions of Theorems 1 and 2 are equivalent, see Remark 4, to

$$\begin{bmatrix} -P & -\zeta^\top \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right] \\ -\left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right]^\top \zeta & -P \end{bmatrix} + \begin{bmatrix} \mathbf{Q} & 0 \\ 0 & \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right]^\top \mathbf{A}^{-1} \left[\begin{smallmatrix} P \\ Y \end{smallmatrix}\right] \end{bmatrix} \prec 0 \text{ and} \qquad (32)$$

$$\left( \begin{bmatrix} P \\ Y \end{bmatrix}^\top \zeta + \zeta^\top \begin{bmatrix} P \\ Y \end{bmatrix} \right) + \left( \mathbf{Q} + \begin{bmatrix} P \\ Y \end{bmatrix}^\top \mathbf{A}^{-1} \begin{bmatrix} P \\ Y \end{bmatrix} \right) \prec 0, \qquad (33)$$

[1] See De Persis and Tesi (2020, Def. 1) or Willems et al. (2005, p. 327) for a definition.

respectively, with $\mathbf{A}$ as in (14) and $\boldsymbol{\zeta}$, $\mathbf{Q}$ as in (16). The matrix $\boldsymbol{\zeta}$ appears only in the first term of the two matrix inequalities and represents the center of the uncertainty set $\mathcal{C}$, see (15). On the other hand, the matrices $\mathbf{A}$, $\mathbf{Q}$ appearing in the second term of the two matrix inequalities determine the size of the uncertainty; in particular, the size of $\mathcal{C}$ is given by $(\det \mathbf{Q})^{(n+m)/2}(\det \mathbf{A})^{-n/2}$, see Bisoffi et al. (2021c, §2.2). By Lemma 1, the second terms in (32) and (33) are positive semidefinite, and the design problem can thus be interpreted as the problem of finding a controller that *robustly* stabilizes the dynamics associated with the center $\boldsymbol{\zeta}$ of the uncertainty set $\mathcal{C}$, where the uncertainty increases with the noise bound $\Delta\Delta^\top$, see the expression of $\mathbf{Q}$ in (18). Recall from Section 2.2 that our operative setting of $D_0 \in \mathcal{D}$ implies $\begin{bmatrix} A_\star & B_\star \end{bmatrix} \in \mathcal{C}$.

Quite interestingly, the center $\boldsymbol{\zeta}$ of the uncertainty set $\mathcal{C}$ coincides with the (ordinary) least-squares estimate of the system dynamics, i.e., with the solution $\begin{bmatrix} A_{\rm ls} & B_{\rm ls} \end{bmatrix}$ to

$$\min_{[A\ B]} \|X_1 - AX_0 - BU_0\|_{\rm F}^2$$

where $\|\cdot\|_{\rm F}$ is the Frobenius norm. Indeed, see Verhaegen and Verdult (2007, §2.6),

$$\begin{aligned}\begin{bmatrix} A_{\rm ls} & B_{\rm ls} \end{bmatrix} &:= \arg\min_{[A\ B]} \|X_1 - AX_0 - BU_0\|_{\rm F}^2 \\ &= X_1 \begin{bmatrix} X_0 \\ U_0 \end{bmatrix}^\dagger = \begin{bmatrix} A_\star & B_\star \end{bmatrix} + D_0 \begin{bmatrix} X_0 \\ U_0 \end{bmatrix}^\dagger \\ &= (-\mathbf{A}^{-1}\mathbf{B})^\top = \boldsymbol{\zeta}^\top\end{aligned}$$

where we rely on Assumption 1. This justifies why *certainty-equivalence* control works well in regimes of small uncertainty (when $\Delta$ is small), in agreement with what has been recently observed in Dörfler et al. (2021) and Mania et al. (2019). On the other hand, this also explains why robust control is generally needed with noisy data, which is also the main idea behind the *robust indirect* control approaches (Dean et al., 2020; Ferizbegovic et al., 2019; Treven et al., 2021) under a stochastic noise description. Besides the noise description, a difference between our work and Dean et al. (2020), Ferizbegovic et al. (2019) and Treven et al. (2021) is that solving (25) or (31) does not require to explicitly construct any estimate of the system dynamics, which is distinctive of indirect methods. Our approach is direct in the sense that it represents an end-to-end method for controller design (once data are collected, substituted in (25) or (31) and these are solved, a controller is obtained). Note that there is a difference from other papers, e.g., Tanaskovic et al. (2017), where direct data-driven control methods tune controller parameters by imposing on them constraints depending on measured data.

We conclude emphasizing that, due to the uncertainty induced by noisy data and the impossibility of knowing the actual system, the goal is to robustly stabilize the set $\mathcal{C}$; for this goal, we provide *necessary and sufficient* conditions in Theorems 1 and 2. Although obtaining $\begin{bmatrix} A_{\rm ls} & B_{\rm ls} \end{bmatrix}$ from data is straightforward, a robust controller designed from $\begin{bmatrix} A_{\rm ls} & B_{\rm ls} \end{bmatrix}$ would not be able to "outperform" (25) or (31) in addressing the uncertainty robustly, namely, if a controller based on $\begin{bmatrix} A_{\rm ls} & B_{\rm ls} \end{bmatrix}$ stabilizes robustly the set $\mathcal{C}$, (25) or (31) would also be feasible and yield a controller. Moreover, to tune the robust controller based on $\begin{bmatrix} A_{\rm ls} & B_{\rm ls} \end{bmatrix}$, users would need to determine from data some magnitude of the uncertainty with respect to which the controller needs to be robust, possibly with conservatism and resulting in sufficient (but not necessary) conditions. On the other hand, our design takes care of embedding the data-induced uncertainty directly in (25) or (31), which are linear matrix inequalities and thus simple to implement given the many available solvers.

### 4.3. Comparison with alternative conditions in van Waarde et al. (2021)

Section 4.1 leads us to a comparison with the approach based on a matrix S-procedure in van Waarde et al. (2021). We recall its main result for data-based stabilization, van Waarde et al. (2021, Thm. 14), and rephrase it for the context of this paper in the next fact.

**Fact 3** (*van Waarde et al., 2021, Thm. 14*). *Assume that the generalized Slater condition*

$$\begin{bmatrix} I \\ \bar{Z} \end{bmatrix}^\top \begin{bmatrix} \mathbf{C} & \mathbf{B}^\top \\ \mathbf{B} & \mathbf{A} \end{bmatrix} \begin{bmatrix} I \\ \bar{Z} \end{bmatrix} \prec 0$$

*holds for some $\bar{Z} \in \mathbb{R}^{(n+m)\times n}$. Then, there exist a feedback gain $K$ and a matrix $P = P^\top \succ 0$ such that $(A+BK)P(A+BK)^\top - P \prec 0$ for all $\begin{bmatrix} A & B \end{bmatrix} \in \mathcal{C}$ if and only if the next program is feasible*

$$\begin{aligned} &\text{find} && P = P^\top \succ 0, Y, \alpha \geq 0, \beta > 0 \\ &\text{s. t.} && \begin{bmatrix} -P + \beta I & 0 \\ 0 & \begin{bmatrix} P \\ Y \end{bmatrix} P^{-1} \begin{bmatrix} P \\ Y \end{bmatrix}^\top \end{bmatrix} - \alpha \begin{bmatrix} \mathbf{C} & \mathbf{B}^\top \\ \mathbf{B} & \mathbf{A} \end{bmatrix} \preceq 0. \end{aligned}$$

*If $P$ and $Y$ are a solution to it, then $K = YP^{-1}$ is a stabilizing gain for all $\begin{bmatrix} A & B \end{bmatrix} \in \mathcal{C}$.*

Fact 3 and Theorem 1 are two alternative approaches since both propose a necessary and sufficient condition for quadratic stabilization; indeed, (25b) in Theorem 1 is equivalent, by Schur complement and changing sign to off-diagonal terms, to

$$\begin{bmatrix} -P & 0 \\ 0 & \begin{bmatrix} P \\ Y \end{bmatrix} P^{-1} \begin{bmatrix} P \\ Y \end{bmatrix}^\top \end{bmatrix} - \begin{bmatrix} \mathbf{C} & \mathbf{B}^\top \\ \mathbf{B} & \mathbf{A} \end{bmatrix} \prec 0.$$

There are some interesting differences, though. Fact 3 operates under a Slater condition, whereas Theorem 1 under Assumption 1. The Slater condition can capture the case of an unbounded set $\mathcal{C}$, which cannot occur with Assumption 1 (see Lemma 2); by contrast, the Slater condition cannot capture the case of ideal data (van Waarde & Camlibel, 2021, §II.C), which requires different arguments (van Waarde & Camlibel, 2021).

In addition, while performance specification in the form of $\mathcal{H}_2$ or $\mathcal{H}_\infty$ control are addressed through the matrix S-procedure in van Waarde et al. (2021), performance specifications in terms of, e.g., convergence rate and overshoot can be handled by Petersen's lemma as in Bisoffi, De Persis, and Tesi (2021b). Both approaches seem then to enjoy a similar degree of flexibility. Whereas Section 3 handles linear systems, the applicability of the proposed method based on Petersen's lemma extends to polynomial systems as in Section 5.

### 4.4. $\mathcal{C}$ as an ellipsoidal over-approximation

As (9) shows, we have derived the set $\mathcal{C}$ based on the disturbance bound in $\mathcal{D}$ and the relation data need to satisfy. On the other hand, the matrix-ellipsoid form (15) of set $\mathcal{C}$ can be fruitfully used as an over-approximation of sets of matrices consistent with data that are not matrix ellipsoids, since ellipsoidal sets are generally better tractable. In that case, as long as matrices $\mathbf{A}$ and $\mathbf{Q}$ in (15) satisfy $\mathbf{A} \succ 0$ and $\mathbf{Q} \succeq 0$, one can use directly Corollary 1. We describe succinctly a relevant case when this could be done based on Bisoffi et al. (2021c), to which we refer the reader for a more elaborate discussion.

With the definitions for $i = 0, 1, \ldots, T-1$

$$x_i^\circ := x(t_{i+1}) \text{ or } x_i^\circ := \dot{x}(t_i),\ x_i := x(t_i),\ u_i := u(t_i)$$

that embed discrete or continuous time, consider the disturbance model $\mathcal{D}_{\rm i} := \{d \in \mathbb{R}^n : |d|^2 \leq \delta\}$. The corresponding set of matrices consistent with all data points $i = 0, \dots, T-1$ is $\mathcal{C}_\mathrm{i} := \bigcap_{i=0}^{T-1}\{[A \ \ B] : x_i^\circ = Ax_i + Bv_i + d, d \in \mathcal{D}_\mathrm{i}\}$ and, due to the intersection, its size remains equal or decreases with $T$. $\mathcal{C}_\mathrm{i}$ is not a matrix ellipsoid and the results in Section 3 cannot be applied to it. Still, a matrix ellipsoid $\mathcal{C} \supseteq \mathcal{C}_\mathrm{i}$ as in (12) can be readily obtained; its parameters $\mathbf{A}, \mathbf{B}, \mathbf{C} := \mathbf{B}^\top\mathbf{A}^{-1}\mathbf{B} - I$ follow from the optimization problem

$$\text{min.} \quad -\log\det\mathbf{A} \quad (\text{over } \mathbf{A}, \mathbf{B}, \tau_0, \dots, \tau_{T-1}) \tag{34a}$$

$$\text{s. t.} \quad \begin{bmatrix} -I - \sum_{i=0}^{T-1}\tau_i\boldsymbol{\gamma}_i & \star & \star \\ \mathbf{B} - \sum_{i=0}^{T-1}\tau_i\boldsymbol{\beta}_i & \mathbf{A} - \sum_{i=0}^{T-1}\tau_i\boldsymbol{\alpha}_i & \star \\ \mathbf{B} & 0 & -\mathbf{A} \end{bmatrix} \preceq 0 \tag{34b}$$

$$\mathbf{A} \succ 0, \ \tau_i \geq 0 \text{ for } i = 0, 1, \dots, T-1 \tag{34c}$$

with data-related quantities

$$\boldsymbol{\gamma}_i := x_i^\circ x_i^{\circ\top} - \delta I, \ \boldsymbol{\beta}_i := -\begin{bmatrix} x_i \\ v_i \end{bmatrix} x_i^{\circ\top}, \ \boldsymbol{\alpha}_i := \begin{bmatrix} x_i \\ v_i \end{bmatrix}\begin{bmatrix} x_i \\ v_i \end{bmatrix}^\top \tag{35}$$

for $i = 0, \dots, T-1$. (This optimization problem is the natural extension to matrix ellipsoids of the one in Boyd et al., 1994, §3.7.2 for classical ellipsoids.) A feasible solution to (34) guarantees by construction $\mathbf{A} \succ 0$ and $\mathbf{Q} = \mathbf{B}^\top\mathbf{A}^{-1}\mathbf{B} - \mathbf{C} = I \succeq 0$ (by the selection of $\mathbf{C}$); hence, Corollary 1 can be applied to this $\mathcal{C}$. A very desirable feature of this $\mathcal{C}$, inherited from $\mathcal{D}_\mathrm{i}$, is that its size generally decreases with $T$, and this requires, in turn, a lesser degree of robustness in the design of the controller if one collects more data. In summary, when an instantaneous disturbance model $\mathcal{D}_\mathrm{i}$ is given, the results of Section 3 cannot be applied to the corresponding set $\mathcal{C}_\mathrm{i}$ but can be to the set $\mathcal{C}$ obtained by (34). The tightness of the over-approximation is problem-dependent, and it might be convenient to work directly with $\mathcal{C}_\mathrm{i}$ at the expense of an increase in the computational complexity (Bisoffi et al., 2021c).

## 5. Data-driven control for polynomial systems

We illustrate in this section that Petersen's lemma proves useful also for polynomial systems, if applied pointwise. As an important class of nonlinear input-affine systems, consider the polynomial systems

$$\dot{x} = f_\star(x) + g_\star(x)u + d = A_\star Z(x) + B_\star W(x)u + d \tag{36}$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ is the input, $d \in \mathbb{R}^n$ is the disturbance; $x \mapsto Z(x) \in \mathbb{R}^N$ is a *known* regressor vector of monomials of $x$ and $x \mapsto W(x) \in \mathbb{R}^{M\times m}$ is a *known* regressor matrix of monomials of $x$; the rectangular matrices $A_\star \in \mathbb{R}^{n\times N}$ and $B_\star \in \mathbb{R}^{n\times M}$ with the coefficients of the regressors are *unknown* to us. In (36), $\dot{x} = f_\star(x) + g_\star(x)u + d$ represents the actual system, so we emphasize that, by virtue of how (36) is written, we assume that if a monomial is present in the actual $f_\star$ or $g_\star$, then it is also listed in $Z$ or $W$, respectively. We allow that $Z$ or $W$ list *more* monomials than those in $f_\star$ or $g_\star$. In fact, the typical case is that, due to lack of knowledge, one includes in $Z$ or $W$ more monomials than necessary. The selection of the regressors $Z$ and $W$ is a key aspect for feasibility of the optimization-based control law, and we comment this in detail in Section 6.3. We will handle data-driven control conditions for (36) through a sum-of-squares relaxation; since sum-of-squares tools are most commonly used for continuous-time systems, we consider directly the continuous-time case in (36).

As in Section 2.2, we perform an experiment on the system by applying an input sequence $u(t_0), \dots, u(t_{T-1})$ of $T$ samples and measure the state and state-derivative sequences $x(t_0), \dots, x(t_{T-1})$ and $\dot{x}(t_0), \dots, \dot{x}(t_{T-1})$. The unknown disturbance sequence $d(t_0), \dots, d(t_{T-1})$ affects the evolution of the system, leading to noisy data. We collect the data points in the matrices

$$V_0 := \begin{bmatrix} W(x(t_0))u(t_0) & \cdots & W(x(t_{T-1}))u(t_{T-1}) \end{bmatrix} \tag{37a}$$

$$Z_0 := \begin{bmatrix} Z(x(t_0)) & \cdots & Z(x(t_{T-1})) \end{bmatrix} \tag{37b}$$

$$X_1 := \begin{bmatrix} \dot{x}(t_0) & \cdots & \dot{x}(t_{T-1}) \end{bmatrix}. \tag{37c}$$

With the unknown disturbance sequence in $D_0 := \begin{bmatrix} d(t_0) & \cdots & d(t_{T-1}) \end{bmatrix}$, data satisfy

$$X_1 = A_\star Z_0 + B_\star V_0 + D_0.$$

As in Section 2.2, the set of matrices consistent with data $X_1$, $Z_0$, $V_0$ and disturbance model $\mathcal{D}$ in (8) is

$$\tilde{\mathcal{C}} := \{[A \ \ B] : X_1 = AZ_0 + BV_0 + D, D \in \mathcal{D}\}$$

and we have $D_0 \in \mathcal{D}$. As in the linear case, $D_0 \in \mathcal{D}$ is precisely equivalent to $[A_\star \ \ B_\star] \in \tilde{\mathcal{C}}$. We can then follow closely the rationale of Section 2.3, and we briefly outline only the key steps. The set $\tilde{\mathcal{C}}$ can be reformulated as

$$\tilde{\mathcal{C}} = \left\{[A \ \ B] : [I \ \ A \ \ B] \cdot \left[\begin{array}{c|c} \tilde{\mathbf{C}} & \tilde{\mathbf{B}}^\top \\ \hline \tilde{\mathbf{B}} & \tilde{\mathbf{A}} \end{array}\right] [\star]^\top \preceq 0\right\}$$

$$\left[\begin{array}{c|c} \tilde{\mathbf{C}} & \tilde{\mathbf{B}}^\top \\ \hline \tilde{\mathbf{B}} & \tilde{\mathbf{A}} \end{array}\right] := \left[\begin{array}{c|c} X_1X_1^\top - \Delta\Delta^\top & -X_1\begin{bmatrix} Z_0 \\ V_0 \end{bmatrix}^\top \\ \hline -\begin{bmatrix} Z_0 \\ V_0 \end{bmatrix}X_1^\top & \begin{bmatrix} Z_0 \\ V_0 \end{bmatrix}\begin{bmatrix} Z_0 \\ V_0 \end{bmatrix}^\top \end{array}\right].$$

The next assumption is analogous to Assumption 1.

**Assumption 2.** Matrix $\begin{bmatrix} Z_0 \\ V_0 \end{bmatrix}$ has full row rank.

$\tilde{\mathbf{A}} \succ 0$ by Assumption 2, and $\tilde{\mathcal{C}}$ can be rewritten as

$$\tilde{\mathcal{C}} = \{[A \ \ B] = Z^\top : (Z - \tilde{\zeta})^\top\tilde{\mathbf{A}}(Z - \tilde{\zeta}) \preceq \tilde{\mathbf{Q}}\} \tag{38}$$

$$\tilde{\zeta} := -\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}, \ \tilde{\mathbf{Q}} := \tilde{\mathbf{B}}^\top\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}} - \tilde{\mathbf{C}}.$$

The logical steps of Lemmas 1, 2 and Proposition 1 can be repeated in the same way after replacing $\begin{bmatrix} X_0 \\ U_0 \end{bmatrix}$ with $\begin{bmatrix} Z_0 \\ V_0 \end{bmatrix}$, so their results are summarized in the next lemma without proof.

**Lemma 3.** *Under Assumption 2, we have: $\tilde{\mathbf{A}} \succ 0$, $\tilde{\mathbf{Q}} \succeq 0$, $\tilde{\mathcal{C}}$ is bounded with respect to any matrix norm, and*

$$\tilde{\mathcal{C}} = \{(\tilde{\zeta} + \tilde{\mathbf{A}}^{-1/2}\Upsilon\tilde{\mathbf{Q}}^{1/2})^\top : \|\Upsilon\| \leq 1\}. \tag{39}$$

As in Section 3, the matrix-ellipsoid parametrization in (39) is key to apply Petersen's lemma, which allows us to obtain the next result for data-driven control of the polynomial system in (36).

**Proposition 2.** *Let Assumption 2 hold and $Z(0) = 0$. Given positive definite*[2] *polynomials $\ell_1$, $\ell_2$ with $\ell_1$ radially unbounded,*[3] *suppose there exist polynomials $V$, $k$, $\lambda$ with $V(0) = 0$ and $k(0) = 0$ such that for each $x$*

$$V(x) - \ell_1(x) \geq 0 \tag{40}$$

$$\begin{bmatrix} \ell_2(x) + \frac{\partial V}{\partial x}(x)\tilde{\zeta}^\top\begin{bmatrix} Z(x) \\ W(x)k(x) \end{bmatrix} & \star & \star \\ \tilde{\mathbf{A}}^{-1/2}\begin{bmatrix} Z(x) \\ W(x)k(x) \end{bmatrix} & -\lambda(x)I & \star \\ \lambda(x)\tilde{\mathbf{Q}}^{1/2}\frac{\partial V}{\partial x}(x)^\top & 0 & -4\lambda(x)I \end{bmatrix} \preceq 0 \tag{41}$$

$$\lambda(x) > 0. \tag{42}$$

[2] That is, zero at zero and positive elsewhere.

[3] That is, $\ell_1(x) \to +\infty$ as $|x| \to +\infty$.

*Then, the origin is globally asymptotically stable for*

$$\dot{x} = AZ(x) + BW(x)k(x) =: f_{A,B}(x)$$

*for all* $\begin{bmatrix} A & B \end{bmatrix} \in \tilde{\mathcal{C}}$, *and in particular for* $\begin{bmatrix} A_\star & B_\star \end{bmatrix} \in \tilde{\mathcal{C}}$, *i.e., for the closed loop* $\dot{x} = f_{A_\star,B_\star}(x)$.

Let us comment the conditions and the conclusion of Proposition 2. Condition (40) imposes positive definiteness and radial unboundedness of the Lyapunov function $V$; condition (42) is the positivity of the multiplier used in Petersen's lemma; condition (41) imposes decrease of the Lyapunov function for all $\begin{bmatrix} A & B \end{bmatrix} \in \tilde{\mathcal{C}}$. In particular, suppose $\tilde{\boldsymbol{\zeta}}^\top = \begin{bmatrix} A_\star & B_\star \end{bmatrix}$ in (41); then, the block (1, 1) alone of the matrix in (41) would express a model-based condition for global asymptotic stability for $\dot{x} = A_\star Z(x) + B_\star W(x)k(x)$. The conclusion is global asymptotic stability for the closed loop $\dot{x} = f_{A,B}(x)$ for all $\begin{bmatrix} A & B \end{bmatrix} \in \tilde{\mathcal{C}}$. Similarly to the linear case (see comment below (11)), this is relevant for the closed loop with disturbance $\dot{x} = f_{A_\star,B_\star}(x) + d$ obtained from (36) because global asymptotic stability guarantees input-to-state stability with "small disturbances" as shown in Sontag (1990, Thm. 2), to which we refer for precise characterizations.

**Proof of Proposition 2.** Note first that since $Z(0) = 0$ and $k(0) = 0$, the origin is an equilibrium of $f_{A,B}$ for all $\begin{bmatrix} A & B \end{bmatrix} \in \tilde{\mathcal{C}}$. Then, the proof consists in showing that $V$ is a Lyapunov function for all systems $\dot{x} = f_{A,B}(x)$, $\begin{bmatrix} A & B \end{bmatrix} \in \tilde{\mathcal{C}}$. Specifically, we show that *(i)* $V$ is positive definite and radially unbounded, and *(ii)* its derivative along solutions satisfies

$$\langle \nabla V(x), f_{A,B}(x) \rangle = \tfrac{\partial V}{\partial x}(x) \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} Z(x) \\ W(x)k(x) \end{bmatrix} \le -\ell_2(x) \ \ \forall x, \forall \begin{bmatrix} A & B \end{bmatrix} \in \tilde{\mathcal{C}}. \tag{43}$$

If the previous properties *(i)*-*(ii)* hold, classical Lyapunov theory (Khalil, 2002, Thm. 4.2) yields the conclusion of the proposition. Positive definiteness of $V$ follows from $V(0) = 0$, (40) and $\ell_1$ positive definite; radial unboundedness of $V$ follows from (40) and $\ell_1$ radially unbounded. We then address the derivative of $V$ along solutions. Set $\begin{bmatrix} A & B \end{bmatrix} = Z^\top \in \tilde{\mathcal{C}}$ in (43) and substitute the parametrization of $Z$ from (39); (43) holds if and only if, for each $x$,

$$\begin{aligned} &-\ell_2(x) \ge \langle \nabla V(x), f_{A,B}(x) \rangle = \tfrac{\partial V}{\partial x}(x) \tilde{\boldsymbol{\zeta}}^\top \begin{bmatrix} Z(x) \\ W(x)k(x) \end{bmatrix} \\ &+ \begin{bmatrix} Z(x) \\ W(x)k(x) \end{bmatrix}^\top \tilde{\mathbf{A}}^{-1/2} \Upsilon \tilde{\mathbf{Q}}^{1/2} \tfrac{1}{2} \tfrac{\partial V}{\partial x}(x)^\top \\ &+ \tfrac{1}{2} \tfrac{\partial V}{\partial x}(x) \tilde{\mathbf{Q}}^{1/2} \Upsilon^\top \tilde{\mathbf{A}}^{-1/2} \begin{bmatrix} Z(x) \\ W(x)k(x) \end{bmatrix} \ \ \forall \Upsilon \text{ with } \|\Upsilon\| \le 1. \end{aligned} \tag{44}$$

We now show that this is true thanks to (41) and (42). By Schur complement for nonstrict inequalities (Boyd et al., 1994, p. 28) and (42), (41) is equivalent to

$$\begin{aligned} &-\ell_2(x) \ge \tfrac{\partial V}{\partial x}(x) \tilde{\boldsymbol{\zeta}}^\top \begin{bmatrix} Z(x) \\ W(x)k(x) \end{bmatrix} \\ &+ \begin{bmatrix} Z(x) \\ W(x)k(x) \end{bmatrix}^\top \cdot \tfrac{\tilde{\mathbf{A}}^{-1}}{\lambda(x)} [\star]^\top + \tfrac{\partial V}{\partial x}(x) \cdot \tfrac{\lambda(x) \tilde{\mathbf{Q}}}{4} [\star]^\top. \end{aligned} \tag{45}$$

In other words, we have by (41) and (42) that for each $x$, there exists $1/\lambda(x) > 0$ such that (45) holds. Apply Fact 2 pointwise (i.e., for each $x$) to (45) with $\mathbf{E}$ and $\mathbf{G}^\top$ corresponding respectively to $\begin{bmatrix} Z(x) \\ W(x)k(x) \end{bmatrix}^\top \tilde{\mathbf{A}}^{-1/2}$ and $\frac{1}{2} \frac{\partial V}{\partial x}(x) \tilde{\mathbf{Q}}^{1/2}$; the fact that for each $x$, there exists $1/\lambda(x) > 0$ such that (45) holds implies that for each $x$, (44) holds or, equivalently, that for each $x$, (43) holds. All properties required of $V$ have been shown, and the conclusion of the proposition follows. □

When writing the Lyapunov derivative along solutions as in (43) and substituting the expression of $\tilde{\mathcal{C}}$ as in (44), the utility of Petersen's lemma beyond the case of linear systems becomes clear. We use the nonstrict version of it in Fact 2 (instead of the strict version in Fact 1) in view of the next sum-of-squares relaxation and the subsequent numerical implementation, where only *nonstrict* inequalities can effectively be implemented. Polynomial positivity in the conditions of Proposition 2 is impractical to verify, so we turn them into sum-of-squares conditions in the next theorem.

**Theorem 3.** *Let Assumption 2 hold and* $Z(0) = 0$. *Given positive definite polynomials* $\ell_1, \ell_2$ *with* $\ell_1$ *radially unbounded and a positive scalar* $\epsilon_\lambda$, *suppose there exist polynomials* $V$, $k$, $\lambda$ *with* $V(0) = 0$ *and* $k(0) = 0$ *such that*

$$V - \ell_1 \in \mathcal{S} \tag{46a}$$

$$-\begin{bmatrix} \ell_2 + \frac{\partial V}{\partial x} \tilde{\boldsymbol{\zeta}}^\top \begin{bmatrix} Z \\ Wk \end{bmatrix} & \star & \star \\ \tilde{\mathbf{A}}^{-1/2} \begin{bmatrix} Z \\ Wk \end{bmatrix} & -\lambda I & \star \\ \lambda \tilde{\mathbf{Q}}^{1/2} \frac{\partial V}{\partial x}^\top & 0 & -4\lambda I \end{bmatrix} \in \mathcal{S}_m \tag{46b}$$

$$\lambda - \epsilon_\lambda \in \mathcal{S}. \tag{46c}$$

*Then, (40)–(42) and the conclusion of Proposition 2 hold.*

**Proof.** (46a) and (46c) imply (40) and (42), respectively. Call $x \mapsto Q(x)$ the matrix polynomial in (46b), so that (46b) rewrites $-Q \in \mathcal{S}_{\mathrm{m}}$. By definition of $\mathcal{S}_{\mathrm{m}}$, see Chesi (2010, Eq. (9)), we have that for each $x$, $Q(x) \preceq 0$, i.e., (41). □

Let us comment Theorem 3. Quantities $Z$ and $W$ are the known regressors; $\tilde{\boldsymbol{\zeta}}$, $\tilde{\mathbf{A}}$, $\tilde{\mathbf{Q}}$ are obtained from data $X_1$, $Z_0$, $V_0$; $\ell_1$, $\ell_2$ and $\epsilon_\lambda$ are design parameters; finally, $V$, $k$ and $\lambda$ are decision variables. Then, the blocks (1, 1), (3, 1) and (1, 3) of the matrix in (46b) entail products between decision variables, which make condition (46b) bilinear and the feasibility program in (46) not convex. A suboptimal strategy that is widely adopted in the sum-of-squares literature, see Jarvis-Wloszek, Feeley, Tan, Sun, and Packard (2005), is to alternately solve for $V$ with $k$ and $\lambda$ fixed, and solve for $k$ and $\lambda$ with $V$ fixed. We illustrate this strategy in Section 6.

As in Section 3, when the set $\tilde{\mathcal{C}}$ is given directly in the form (38) as a matrix-ellipsoid over-approximation of a less tractable set (see the discussion in Section 4.4), one can use the next corollary instead of Theorem 3.

**Corollary 2.** *Let* $\tilde{\mathbf{A}} \succ 0$ *and* $\tilde{\mathbf{Q}} \succeq 0$ *hold for the set* $\tilde{\mathcal{C}} = \big\{ \begin{bmatrix} A & B \end{bmatrix} = Z^\top : (Z - \tilde{\zeta})^\top \tilde{\mathbf{A}} (Z - \tilde{\zeta}) \preceq \tilde{\mathbf{Q}} \big\}$ *as in (38) and* $Z(0) = 0$. *Given positive definite polynomials* $\ell_1, \ell_2$ *with* $\ell_1$ *radially unbounded and a positive scalar* $\epsilon_\lambda$, *suppose there exist polynomials* $V$, $k$, $\lambda$ *with* $V(0) = 0$ *and* $k(0) = 0$ *satisfying (46). Then, (40)–(42) and the conclusion of Proposition 2 hold.*

In the next section, we obtain $\tilde{\mathcal{C}}$ as described in Section 4.4 and, in particular, through the optimization problem in (34). This provides a set $\tilde{\mathcal{C}}$ directly in the form (38), so we will apply Corollary 2.

Finally, we follow up on the comparison with Guo et al. (2021) discussed in Section 1. As the proof of Proposition 2 shows, the data-based conditions (40)–(42) correspond naturally to enforcing model-based conditions (Khalil, 2002, Thm. 4.2) for all systems consistent with data. This makes this approach extendible to other cases such as *local* asymptotic stability. Indeed, if we consider (Khalil, 2002, Thm. 4.1), we obtain the next corollary.

**Corollary 3.** *Let Assumption 2 hold and* $Z(0) = 0$. *Given positive definite polynomials* $\ell_0, \ell_1, \ell_2$ *and a positive scalar* $c$ *yielding* $\mathcal{D}_c := \{x \in \mathbb{R}^n : \ell_0(x) \le c\}$, *suppose there exist polynomials* $s_1$, $s_2$, $V$, $k$, $\lambda$ *with* $V(0) = 0$ *and* $k(0) = 0$ *such that for each* $x$

$$s_1(x) \ge 0, \ \ s_2(x) \ge 0, \tag{47}$$

$$V(x) - \ell_1(x) + s_1(x)(\ell_0(x) - c) \ge 0 \tag{48}$$

$$\begin{bmatrix} \begin{Bmatrix} \ell_2(x) + \frac{\partial V}{\partial x}(x)\tilde{\boldsymbol{\zeta}}^\top \begin{bmatrix} Z(x) \\ W(x)k(x) \end{bmatrix} \\ - s_2(x)(\ell_0(x) - c) \end{Bmatrix} & \star & \star \\ \tilde{\mathbf{A}}^{-1/2} \begin{bmatrix} Z(x) \\ W(x)k(x) \end{bmatrix} & -\lambda(x)I & \star \\ \lambda(x)\tilde{\mathbf{Q}}^{1/2}\frac{\partial V}{\partial x}(x)^\top & 0 & -4\lambda(x)I \end{bmatrix} \preceq 0 \tag{49}$$

$$\lambda(x) > 0. \tag{50}$$

*Then, the origin is locally asymptotically stable for $\dot{x} = f_{A,B}(x)$ for all $\begin{bmatrix} A & B \end{bmatrix} \in \tilde{\mathcal{C}}$, and in particular for $\begin{bmatrix} A_\star & B_\star \end{bmatrix}$.*

The proof would follow the same rationale as the proof of Proposition 2, so we sketch only the key steps to highlight that the conditions (47)–(50) in Corollary 3 follow naturally from Khalil (2002, Thm. 4.1). (47) and (48) imply that $V(x) \geq \ell_1(x)$ for all $x \in \mathcal{D}_c$ and give (Khalil, 2002, Eq. (4.2)). To have (Khalil, 2002, Eq. (4.4)), we would like to impose for all $\begin{bmatrix} A & B \end{bmatrix} \in \tilde{\mathcal{C}}$ that $\langle \nabla V(x), f_{A,B}(x) \rangle \leq -\ell_2(x)$ for all $x \in \mathcal{D}_c$. This is implied by the fact that for all $x$, for all $\Upsilon$ with $\|\Upsilon\| \leq 1$, $\ell_2(x) + \frac{\partial V}{\partial x}(x)(\tilde{\boldsymbol{\zeta}} + \tilde{\mathbf{A}}^{-1/2}\Upsilon\tilde{\mathbf{Q}})^\top \begin{bmatrix} Z(x) \\ W(x)k(x) \end{bmatrix} - s_2(x)(\ell_0(x) - c) \leq 0$. This condition is indeed obtained from (47), (49)–(50) and Petersen's lemma. With Corollary 3, it is immediate to write its sum-of-squares relaxation for decision variables $s_1, s_2, V, k, \lambda$ in the same way we wrote Theorem 3 with Proposition 2. We note that local asymptotic stability can be verified from data for a polynomial system by specializing (Martin & Allgöwer, 2021), whereas Corollary 3 addresses control design.

## 6. Numerical illustration

In this section we consider as a running example the system in Khalil (2002, Example 14.9), i.e.,

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} x_1^2 - x_1^3 + x_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u + d. \tag{51}$$

This continuous-time polynomial system can be cast in the form in (36). We consider it as such in Section 6.3 to illustrate Theorem 3; we consider its linearization

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u + d =: A_\star^{\mathrm{ct}} x + B_\star^{\mathrm{ct}} u + d \tag{52}$$

in Section 6.2 to illustrate Theorem 2; with an Euler discretization and sampling time $\tau_s$, we consider

$$x^+ = A_\star^{\mathrm{dt}} x + B_\star^{\mathrm{dt}} u + d := (I + \tau_s A_\star^{\mathrm{ct}})x + (\tau_s B_\star^{\mathrm{ct}})u + d \tag{53}$$

in Section 6.1 to illustrate Theorem 1. We emphasize that these systems are used only for data generation, but the vector fields in (51)–(53) are not known to the data-based schemes. For $\delta > 0$, the disturbance $d$ in (51)–(53) is taken as $(\sqrt{\delta}\cos(2\pi 0.4t), \sqrt{\delta}\sin(2\pi 0.4t))$, where $t$ corresponds to integer multiples of $\tau_s$ in discrete time. Hence, $d$ satisfies $|d|^2 \leq \delta$. From this bound on $d$, the disturbance sequence $D$ in (8) satisfies then the bound $DD^\top \preceq T\delta I$. For the linear systems (52)–(53), we convert the bound on $d$ into $\Delta := \sqrt{T\delta}I$ in (8); for the polynomial system (51), we retain the bound on $d$ and consider $\tilde{\mathcal{C}}$ as an ellipsoidal over-approximation of the type described in Section 4.4. We solve all numericals programs using YALMIP (Löfberg, 2004) with its sum-of-squares functionality (Löfberg, 2009), MOSEK ApS and MATLAB® R2019b.

### 6.1. Linear system in discrete time

Consider (53) with $\tau_s = 0.5$, $T = 100$ and $\delta = 0.1$. The experiment generating data $U_0$, $X_0$ and $X_1$ in (6) is depicted in Fig. 1. A uniform random variable in $[-1, 1]$ is used as input $u$. Matrices $X_0$ and $U_0$ satisfy Assumption 1. Using the semidefinite program in Theorem 1, a controller $K = \begin{bmatrix} -0.1521 & -1.3475 \end{bmatrix}$ is designed, whose stabilization properties are certified by a Lyapunov function $x^\top P^{-1} x = x^\top \begin{bmatrix} 0.0043 & 0.0115 \\ 0.0115 & 0.1000 \end{bmatrix} x$. The resulting closed-loop solutions for $d = 0$ and the level sets of this Lyapunov function are depicted in Fig. 2.

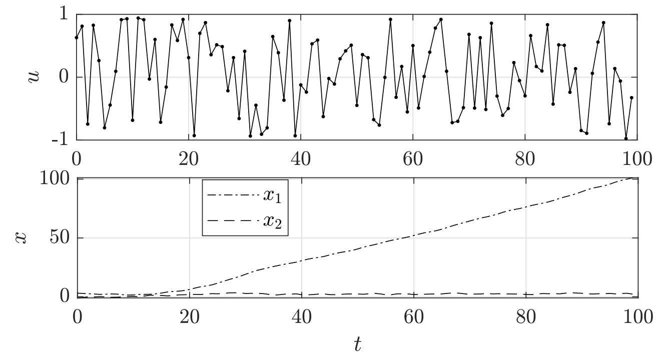

**Fig. 1.** Experiment yielding data from (53): input and state. Dots on the curve of $u$ indicate the discrete-time instants.

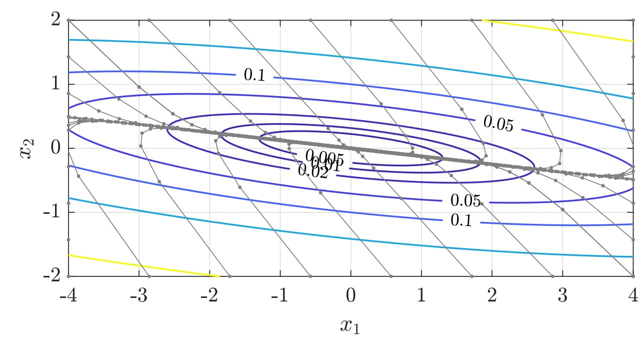

**Fig. 2.** Phase plot of (53) with $d = 0$ in closed loop with the data-based controller $K$ obtained in Section 6.1. Solutions are gray, where dots correspond to the different discrete-time instants. The level sets of the Lyapunov function are colored and their corresponding values are annotated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

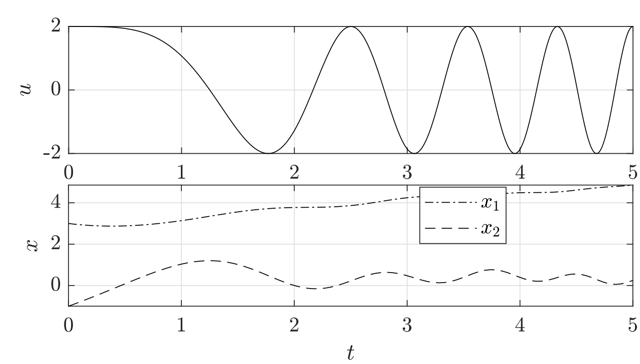

**Fig. 3.** Experiment yielding data from (52): input and state.

### 6.2. Linear system in continuous time

Consider (52) with $T = 100$ and $\delta = 0.1$. The experiment generating data $U_0$, $X_0$ and $X_1$ in (6) is depicted in Fig. 3. A sweeping sine with minimum & maximum frequencies 0 & 0.8 and amplitude 2 is used as input $u$. Matrices $X_0$ and $U_0$ satisfy Assumption 1. The times $t_0$, $t_1$, …, $t_{T-1}$ when state and state derivative are evaluated for $X_0$ and $X_1$ are uniformly spaced by

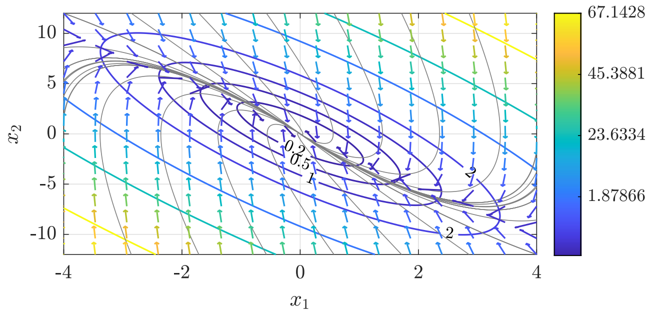


**Fig. 4.** Phase plot of (52) with $d = 0$ in closed loop with the data-based controller $K$ obtained in Section 6.2. Solutions are gray. The arrows represent the closed-loop vector field, and their color indicates their actual magnitude as in the right color bar. The level sets of the Lyapunov function are colored and their corresponding values are annotated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

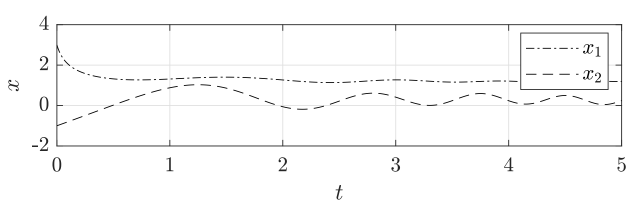


**Fig. 5.** Experiment yielding data from (51): state. The input is the same as in Fig. 3.

$5/T = 5/100$. Using the semidefinite program in Theorem 2, a controller $K = \begin{bmatrix} -21.4762 & -9.2835 \end{bmatrix}$ is designed, whose stabilization properties are certified by a Lyapunov function $x^\top P^{-1} x = x^\top \begin{bmatrix} 0.5214 & 0.1430 \\ 0.1430 & 0.0590 \end{bmatrix} x$. The resulting closed-loop solutions for $d = 0$ and the level sets of this Lyapunov function are depicted in Fig. 4.

### 6.3. Polynomial system

Consider (51) with $T = 1000$ and $\delta = 0.01$. Input and disturbance of the experiment generating data $V_0$, $Z_0$ and $X_1$ in (37) take the same form as in Section 6.2 with times $t_0$, $t_1$, …, $t_{T-1}$ uniformly spaced by $5/T = 5/1000$. Since (51) is now nonlinear unlike (52), these input and disturbance result in a different state evolution $x$, which is reported in Fig. 5.

Whereas the setup of the semidefinite programs in Theorems 1 and 2 is quite straightforward, the setup of the sum-of-squares program from Theorem 3 is less so, and we illustrate its most relevant aspects.

*(1)* The *selection of the regressors* $Z$ and $W$ in (36) is a key step. On one hand, the system is unknown and some of the monomials in the regressors may not appear in the "true" vector fields; on the other hand, the more the monomials and the associated unknown coefficients are, the larger the uncertainty typically is in such coefficients and a too large uncertainty affects feasibility of the sum-of-squares program in a critical way. Therefore, a parsimonious number of monomials is desirable; which monomials should be taken can be determined by trial and error by solving the program with different selections of regressors. We select here $Z(x) := (x_2, x_1^2, x_2^2, x_1^3, x_2^3)$ and $W(x) := 1$, which determine from (36) and (51)

$$\begin{bmatrix} A_\star \mid B_\star \end{bmatrix} := \left[ \begin{array}{ccccc|c} 1 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right].$$

*(2)* The *numerical experiment* is likewise important. Intuitively, the richer the data, the better; for this reason we selected as input a sweeping sine. Moreover, when the set $\tilde{\mathcal{C}}$ is obtained as an ellipsoidal over-approximation as described in Section 4.4, the size of $\tilde{\mathcal{C}}$, the associated uncertainty, and the degree of robustness required in the design of the controller all decrease with $T$, in general; hence, more data points $T$ enlarge the feasibility set of the sum-of-squares program in (46). We obtain the ellipsoidal over-approximation $\tilde{\mathcal{C}}$ by solving the optimization problem (34) with (35) and, for $i = 0, \dots, T-1$,

$$\varkappa_i^\circ := \dot{x}(t_i),\ \varkappa_i := Z(x(t_i)),\ \upsilon_i := W(x(t_i))u(t_i).$$

$\tilde{\mathcal{C}}$ is then defined by the matrices $\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}} := \tilde{\mathbf{B}}^\top \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{B}} - I$ returned by (34) or, alternatively, by $\tilde{\mathbf{Q}} = I$ and $\tilde{\zeta} := -\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}$. Matrix $\tilde{\zeta}$ is especially relevant as the center of the ellipsoid $\tilde{\mathcal{C}}$. For the experiment in Fig. 5, we obtain

$$\tilde{\zeta}^\top = \left[ \begin{array}{ccccc|c} 0.9569 & 1.0243 & 0.0000 & -1.0084 & -0.0627 & 0.0009 \\ -0.0160 & 0.0146 & -0.0336 & -0.0037 & 0.0334 & 1.0101 \end{array} \right]$$

which should be compared against $\begin{bmatrix} A_\star \mid B_\star \end{bmatrix}$.

*(3)* To solve the sum-of-squares program of Theorem 3, we commented after it that we adopt the common practice of *solving alternately two sum-of-squares programs*. Specifically, we first solve (46b) and (46c) with respect to the controller $k$ and multiplier $\lambda$ with fixed Lyapunov function $V$; with the returned controller and multiplier, we solve (46a) and (46b) with respect to $V$ with fixed $k$ and $\lambda$. To start up this procedure, we need an initial guess for $V$. In keeping with the data-based approach, we use the quadratic Lyapunov function that Theorem 2 returns for the linearized system with same disturbance level, in this case $x^\top \begin{bmatrix} 0.0278 & 0.0127 \\ 0.0127 & 0.0216 \end{bmatrix} x$. (This correspond to an experiment with small signals in a neighborhood of the origin, so that the linear approximation is trustworthy; the theoretical legitimacy of such an initial guess is based on De Persis and Tesi (2020, Thm. 6).) The initialization of $V$ is all the more important whenever the feasibility set is small. In the specific example, we run 15 iterations of this procedure (solving a total of 30 sum-of-squares programs).

*(4)* Finally, we mention two aspects regarding the solution of the two alternate programs above. An important aspect for feasibility of each of those is the *selection of the minimum & maximum degrees of polynomials*, as lucidly explained in Tan (2006, Appendix). In this example we select the minimum & maximum degrees for $V$, $k$, $\lambda$ as respectively 2 & 4, 1 & 3, 0 & 4. A minor aspect is that we can take in (46) the *parameter* $\ell_2$ *as decision variable* for greater flexibility since $\ell_2$ appears linearly anyway; when we solve for $k$ and $\lambda$, we also solve for $\ell_2$ and capture that it needs to be positive definite by imposing $\ell_2 \in \mathcal{S}$ (minimum & maximum degrees equal to 2 & 4).

With this procedure and design parameters $\ell_1(x) := 10^{-3}(x_1^2 + x_2^2)$ and $\epsilon_\lambda := 10^{-3}$, the obtained $V$, $k$, $\lambda$ are in the next table; the corresponding closed-loop solutions for $d = 0$ and the level sets of $V$ are depicted in Fig. 6.

| Qty | Expression |
|---|---|
| $V$ | $4.0698x_1^2 + 4.3023x_1x_2 + 3.5364x_2^2 + 0.003475x_1^3$ |
| | $+ 0.02465x_1^2x_2 - 0.01500x_1x_2^2 + 0.001575x_2^3$ |
| | $+ 0.008769x_1^4 + 0.003686x_1^3x_2 + 0.01263x_1^2x_2^2$ |
| | $+ 0.0006249x_1x_2^3 + 0.02279x_2^4$ |
| $k$ | $-1.0291x_1 - 1.7292x_2 - 0.8793x_1^2 + 0.2927x_1x_2 - 0.07565x_2^2$ |
| | $- 0.5511x_1^3 - 1.6307x_1^2x_2 + 0.08336x_1x_2^2 - 2.5235x_2^3$ |
| $\lambda$ | $0.04905 - 0.006151x_1 + 0.002003x_2 + 0.1106x_1^2$ |
| | $+ 0.004398x_1x_2 + 0.1123x_2^2$ |

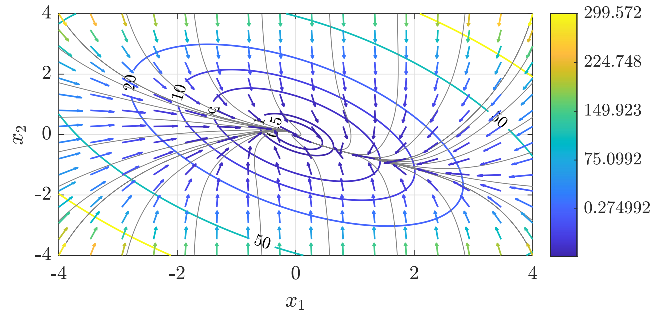


**Fig. 6.** Phase plot of (51) with $d = 0$ in closed loop with the data-based controller $k(\cdot)$ obtained in Section 6.3. See the caption of Fig. 4 for the interpretation of this figure.

### 6.4. Certainty equivalence and computation times

In Section 4.2 we discussed the interpretation of $\zeta^\top$ as the least-squares model $\begin{bmatrix} A_{ls} & B_{ls} \end{bmatrix}$ and the relation of our control law with certainty-equivalence control. We follow on from that section and analyze the performance of the two approaches for the discrete-time system of Section 6.1. Based on $\begin{bmatrix} A_{ls} & B_{ls} \end{bmatrix}$, we design a controller through the program

$$\begin{aligned} &\text{find} && K,\ P = P^\top \succ 0 \\ &\text{s. t.} && (A_{ls} + B_{ls}K)P(A_{ls} + B_{ls}K)^\top - P \prec 0, \end{aligned}$$

which is the counterpart of (10) for a known model and is thus a fair comparison due to the same underlying strategy to design the controller. This program is equivalent, by a Schur complement and a change of variables, to

$$\text{find} \quad Y,\ P = P^\top \succ 0 \tag{54a}$$

$$\text{s. t.} \quad \begin{bmatrix} -P & \begin{bmatrix} A_{ls} & B_{ls} \end{bmatrix} \begin{bmatrix} P \\ Y \end{bmatrix} \\ \begin{bmatrix} P \\ Y \end{bmatrix}^\top \begin{bmatrix} A_{ls} & B_{ls} \end{bmatrix}^\top & -P \end{bmatrix} \prec 0. \tag{54b}$$

If (54) is feasible, we call the returned solution $(Y_{ls}, P_{ls})$, and a controller gain is $K_{ls} := Y_{ls}P_{ls}^{-1}$. This least-squares controller stabilizes all systems corresponding to matrices $\begin{bmatrix} A & B \end{bmatrix} \in \mathcal{C}$ if and only if (10) is feasible for $K = K_{ls}$ or, under Assumption 1, if and only if (25) is feasible for $K = K_{ls}$ by Theorem 1, i.e., if and only if

$$\text{find} \quad P = P^\top \succ 0 \tag{55a}$$

$$\text{s. t.} \quad \begin{bmatrix} -P-\mathbf{C} & 0 & \mathbf{B}^\top \\ 0 & -P & \begin{bmatrix} P & Y_{ls}^\top \end{bmatrix} \\ \mathbf{B} & \begin{bmatrix} P \\ Y_{ls} \end{bmatrix} & -\mathbf{A} \end{bmatrix} \prec 0 \tag{55b}$$

is feasible. To compare feasibility of (25) and feasibility of (55) after a controller is designed from the least-squares estimate as in (54), we consider a grid of points $(\delta, T)$ where $|d|^2 \le \delta$ as in Section 6.1. For each value $(\delta, T)$ in this grid, each element of the input sequence is obtained from a uniform random variable in $[-1, 1]$ and each element of the unknown disturbance sequence is obtained from a uniform random variable in $\{d \in \mathbb{R}^n : |d|^2 \le \delta\}$, so that the disturbance sequence $D$ in (8) satisfies $DD^\top \preceq T\delta I$. For $x_0 = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$, the resulting state sequence determines $U_0$, $X_0$, $X_1$ and $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and we test feasibility of (25). $\mathbf{A}$, $\mathbf{B}$ determine also $\boldsymbol{\zeta}^\top = \begin{bmatrix} A_{ls} & B_{ls} \end{bmatrix}$ and we test feasibility of (55). For each value $(\delta, T)$, we consider a total of $n_{tot} = 25$ different sequences and count the times $n_{\text{Theorem 1}}$ and $n_{ls}$ when, respectively, (25) and (55) are feasible. The ratios $n_{\text{Theorem 1}}/n_{tot}$ and $n_{ls}/n_{tot}$ are in Fig. 7. The figure confirms the discussion in Section 4.2 and in particular that for small values of uncertainty (corresponding here to small $\delta$), certainty-equivalence control is able to stabilize all systems consistent with data but it fails to do so for larger levels of uncertainty, which confirms the need to explicitly take the uncertainty into account as (25) does.

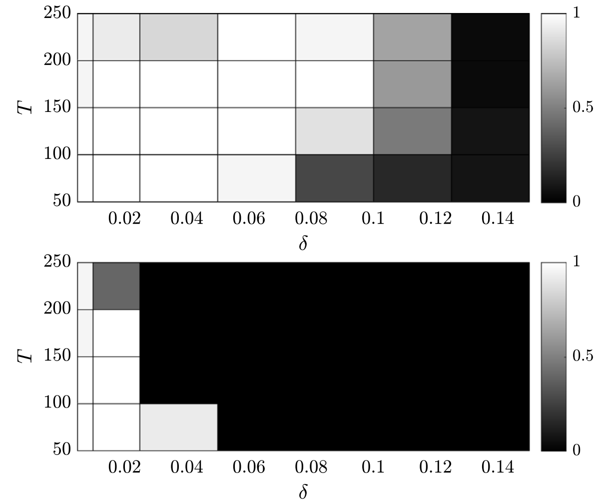


**Fig. 7.** For different values of $(\delta, T)$, the ratio $n_{\text{Theorem 1}}/n_{tot}$ of feasible instances of (25) (top) and the ratio $n_{ls}/n_{tot}$ of feasible instances of (55) (bottom) are depicted.

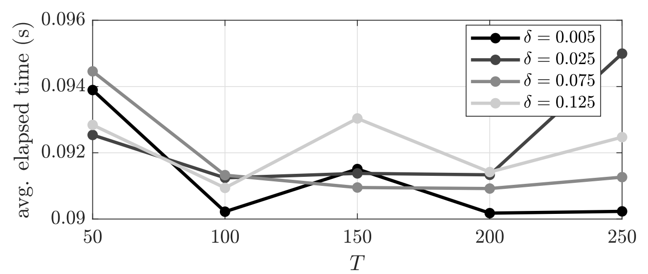


**Fig. 8.** For different values of $\delta$, the average elapsed time to certify (25) as feasible or infeasible is depicted as a function of $T$. These times are approximately constant across $T$.

For different values of $\delta$ and $T$, we measure the elapsed time[4] to certify (25) as feasible or infeasible and find the average over the $n_{tot}$ sequences. The resulting times are reported as a function of $T$ in Fig. 8, which shows that the average elapsed time is approximately constant across $T$. This follows from the fact that the number of neither decision variables nor constraints in (25) depends on $T$.

## Appendix. Proof of Fact 1

To give the proof, we need some auxiliary results. The first one is in the next fact.

**Fact 4** (*Petersen & Hollot, 1986, Lemma A.4*). *Consider matrices* $\mathbf{A}$, $\mathbf{C}$, $\mathbf{B}$ *in* $\mathbb{R}^{r\times r}$ *with* $\mathbf{A} = \mathbf{A}^\top \succeq 0$, $\mathbf{C} = \mathbf{C}^\top \succeq 0$ *and* $\mathbf{B} = \mathbf{B}^\top \prec 0$. *Suppose further that*

$$(w^\top \mathbf{B} w)^2 - 4w^\top \mathbf{A} w w^\top \mathbf{C} w > 0 \text{ for all } w \in \mathbb{R}^r \setminus \{0\}. \tag{A.1}$$

*Then* $\lambda^2\mathbf{A} + \lambda\mathbf{B} + \mathbf{C} \prec 0$ *for some* $\lambda > 0$.

[4] This elapsed time is obtained by the MATLAB® R2019b function `tic toc` on a machine with processor Intel® Core™ i7 with 4 cores and 1.80 GHz.

By Horn and Johnson (2013, Thm. 7.2.7), $\bar{\mathbf{F}} = \bar{\mathbf{F}}^\top$ is positive semidefinite if and only if there exists a $s$-by-$q$ matrix $\Phi$ such that $\bar{\mathbf{F}} = \Phi^\top\Phi$; hence, $\mathcal{F}$ in (1) rewrites as

$$\mathcal{F} := \{\mathbf{F} \in \mathbb{R}^{p\times q} : \mathbf{F}^\top\mathbf{F} \preceq \Phi^\top\Phi\}. \tag{A.2}$$

The second auxiliary result is in essence (Petersen, 1987, Lemma 3.1), of which however we need a slight extension to handle, in the set $\mathcal{F}$ in (A.2), the condition $\mathbf{F}^\top\mathbf{F} \preceq \Phi^\top\Phi$ (with positive *semidefinite* bound) instead of $\mathbf{F}^\top\mathbf{F} \preceq I$, which appears in Petersen (1987, Lemma 3.1). This is done in the next lemma, for which we present also a short proof to account for the required modification.

**Lemma 4.** *For vectors $x$ in $\mathbb{R}^p$, $y$ in $\mathbb{R}^q$ and set $\mathcal{F}$ in (A.2),* $\max_{\mathbf{F}\in\mathcal{F}}(x^\top\mathbf{F}y)^2 = |x|^2\ |\Phi y|^2$.

**Proof.** By Cauchy–Schwarz's inequality and (A.2),

$$|x^\top\mathbf{F}y| \le |x||\mathbf{F}y| = |x|\sqrt{y^\top\mathbf{F}^\top\mathbf{F}y} \le |x|\sqrt{y^\top\Phi^\top\Phi y},$$

that is, $|x^\top\mathbf{F}y| \le |x||\Phi y|$. From this relation we have that the statement is true if $x = 0$ or $\Phi y = 0$. The proof is complete if, for $x \neq 0$ and $\Phi y \neq 0$, we obtain $\mathbf{F} \in \mathcal{F}$ such that $|x^\top\mathbf{F}y| = |x||\Phi y|$, as we do in the rest of the proof. Since $x \neq 0$ and $\Phi y \neq 0$, take the specific selection

$$\mathbf{F} := xy^\top\Phi^\top\Phi/\big(|x||\Phi y|\big).$$

First, we show $\mathbf{F} \in \mathcal{F}$. Indeed,

$$\mathbf{F}^\top\mathbf{F} = \frac{\Phi^\top(\Phi y x^\top x y^\top\Phi^\top)\Phi}{|x|^2|\Phi y|^2} = \frac{\Phi^\top(\Phi y y^\top\Phi^\top)\Phi}{|\Phi y|^2}$$

and $\mathbf{F} \in \mathcal{F}$ because $\Phi y y^\top\Phi^\top \preceq |\Phi y|^2 I$ (for all $v \in \mathbb{R}^s$, $v^\top\big(\Phi y y^\top\Phi^\top\big)v \le |v^\top\Phi y||y^\top\Phi^\top v| \le |\Phi y|^2|v|^2$). Second,

$$(x^\top\mathbf{F}y)^2 = \Big(x^\top\frac{xy^\top\Phi^\top\Phi}{|x||\Phi y|}y\Big)^2 = |x|^2|\Phi y|^2$$

and we have also shown $|x^\top\mathbf{F}y| = |x||\Phi y|$. □

With Fact 4 and Lemma 4, we can prove Fact 1. The direction (2b) $\implies$ (2a) is easy since for all $\mathbf{F} \in \mathcal{F}$ in (A.2),

$$\begin{aligned}
0 &\succ \mathbf{C} + \lambda\mathbf{E}\mathbf{E}^\top + \lambda^{-1}\mathbf{G}^\top\Phi^\top\Phi\mathbf{G}\\
&\succeq \mathbf{C} + \lambda\mathbf{E}\mathbf{E}^\top + \lambda^{-1}\mathbf{G}^\top\mathbf{F}^\top\mathbf{F}\mathbf{G} \quad (\text{by } \lambda > 0)\\
&= \mathbf{C} + \mathbf{E}\mathbf{F}\mathbf{G} + \mathbf{G}^\top\mathbf{F}^\top\mathbf{E}^\top\\
&\qquad + (\sqrt{\lambda}\mathbf{E}^\top - \sqrt{\lambda}^{-1}\mathbf{F}\mathbf{G})^\top(\sqrt{\lambda}\mathbf{E}^\top - \sqrt{\lambda}^{-1}\mathbf{F}\mathbf{G})\\
&\succeq \mathbf{C} + \mathbf{E}\mathbf{F}\mathbf{G} + \mathbf{G}^\top\mathbf{F}^\top\mathbf{E}^\top.
\end{aligned}$$

We turn then to the direction (2a) $\implies$ (2b). (2a) is equivalent to the fact that for all $x \neq 0$ and for all $\mathbf{F} \in \mathcal{F}$, $x^\top\mathbf{C}x + 2x^\top\mathbf{E}\mathbf{F}\mathbf{G}x < 0$ and to the fact that for all $x \neq 0$, $0 > x^\top\mathbf{C}x + 2\max_{\mathbf{F}\in\mathcal{F}}(x^\top\mathbf{E}\mathbf{F}\mathbf{G}x) = x^\top\mathbf{C}x + 2\max_{\mathbf{F}\in\mathcal{F}}|x^\top\mathbf{E}\mathbf{F}\mathbf{G}x|$ because, for each $x$, $\max_{\mathbf{F}\in\mathcal{F}}(x^\top\mathbf{E}\mathbf{F}\mathbf{G}x) \ge \max_{\mathbf{F}\in\mathcal{F},x^\top\mathbf{E}\mathbf{F}\mathbf{G}x\ge 0}(x^\top\mathbf{E}\mathbf{F}\mathbf{G}x) = \max_{\mathbf{F}\in\mathcal{F},x^\top\mathbf{E}\mathbf{F}\mathbf{G}x\le 0}(-x^\top\mathbf{E}\mathbf{F}\mathbf{G}x)$. Apply Lemma 4 and obtain that for all $x \neq 0$, $x^\top\mathbf{C}x + 2|\mathbf{E}^\top x||\Phi\mathbf{G}x| < 0$. For this to hold, we necessarily have $x^\top\mathbf{C}x < 0$ for all $x \neq 0$, i.e., $\mathbf{C} \prec 0$. Under $\mathbf{C} \prec 0$, $x^\top\mathbf{C}x + 2|\mathbf{E}^\top x||\Phi\mathbf{G}x| < 0$ for all $x \neq 0$ is equivalent to the fact that for all $x \neq 0$, $(x^\top\mathbf{C}x)^2 > 4|\mathbf{E}^\top x|^2|\Phi\mathbf{G}x|^2$. This relation corresponds to (A.1) and the hypothesis of Fact 4 is verified. Hence, we conclude from Fact 4 that $\lambda^2\mathbf{E}\mathbf{E}^\top + \lambda\mathbf{C} + \mathbf{G}^\top\Phi^\top\Phi\mathbf{G} \prec 0$ for some $\lambda > 0$, which is equivalent to $\lambda^2\mathbf{E}\mathbf{E}^\top + \lambda\mathbf{C} + \mathbf{G}^\top\bar{\mathbf{F}}\mathbf{G} \prec 0$ and (2b).

## References


Anderson, J., Doyle, J. C., Low, S. H., & Matni, N. (2019). System level synthesis. *Annual Reviews in Control*, *47*, 364–393.

Antsaklis, P. J., & Michel, A. N. (2006). *Linear systems*. Birkhäuser.

Baggio, G., Katewa, V., & Pasqualetti, F. (2019). Data-driven minimum-energy controls for linear systems. *IEEE Control Systems Letters*, *3*(3), 589–594.

Berberich, J., Romer, A., Scherer, C. W., & Allgöwer, F. (2020). Robust data-driven state-feedback design. In *Proc. Amer. contr. conf*.

Berberich, J., Scherer, C. W., & Allgöwer, F. (2020). Combining prior knowledge and data for robust controller design. arXiv preprint arXiv:2009.05253.

Bisoffi, A., De Persis, C., & Tesi, P. (2020). Data-based stabilization of unknown bilinear systems with guaranteed basin of attraction. *Systems & Control Letters*, *145*, Article 104788.

Bisoffi, A., De Persis, C., & Tesi, P. (2021a). Data-driven control via Petersen's lemma. arXiv preprint arXiv:2109.12175.

Bisoffi, A., De Persis, C., & Tesi, P. (2021b). Learning controllers for performance through LMI regions. arXiv preprint arXiv:2110.10777.

Bisoffi, A., De Persis, C., & Tesi, P. (2021c). Trade-offs in learning controllers from noisy data. *Systems & Control Letters*, *154*, Article 104985.

Boyd, S., El Ghaoui, L., Feron, E., & Balakrishnan, V. (1994). *Linear matrix inequalities in system and control theory*. SIAM.

Chesi, G. (2010). LMI techniques for optimization over polynomials in control: A survey. *IEEE Transactions on Automatic Control*, *55*(11), 2500–2510.

Coulson, J., Lygeros, J., & Dörfler, F. (2019). Data-enabled predictive control: In the shallows of the DeePC. In *Proc. Eur. contr. conf*.

Dai, T., & Sznaier, M. (2018). A moments based approach to designing MIMO data driven controllers for switched systems. In *Proc. IEEE conf. dec. contr*.

Dai, T., & Sznaier, M. (2021). A semi-algebraic optimization approach to data-driven control of continuous-time nonlinear systems. *IEEE Control Systems Letters*, *5*(2), 487–492.

De Persis, C., & Tesi, P. (2020). Formulas for data-driven control: Stabilization, optimality and robustness. *IEEE Transactions on Automatic Control*, *65*(3), 909–924.

De Persis, C., & Tesi, P. (2021). Low-complexity learning of linear quadratic regulators from noisy data. *Automatica*, *128*.

Dean, S., Mania, H., Matni, N., Recht, B., & Tu, S. (2020). On the sample complexity of the linear quadratic regulator. *Foundations of Computational Mathematics*, *20*, 633–679.

Dörfler, F., Coulson, J., & Markovsky, I. (2022). Bridging direct & indirect data-driven control formulations via regularizations and relaxations. *IEEE Transactions on Automatic Control*, Early Access.

Dörfler, F., Tesi, P., & De Persis, C. (2021). On the certainty-equivalence approach to direct data-driven LQR design. arXiv preprint arXiv:2109.06643.

Ferizbegovic, M., Umenberger, J., Hjalmarsson, H., & Schön, T. B. (2019). Learning robust LQ-controllers using application oriented exploration. *IEEE Control Systems Letters*, *4*(1), 19–24.

Fogel, E. (1979). System identification via membership set constraints with energy constrained noise. *IEEE Transactions on Automatic Control*, *24*(5), 752–758.

Guo, M., De Persis, C., & Tesi, P. (2021). Data-driven stabilization of nonlinear polynomial systems with noisy data. *IEEE Transactions on Automatic Control*, Early Access.

Hjalmarsson, H., & Ljung, L. (1993). A discussion of "unknown-but-bounded" disturbances in system identification. In *Proc. IEEE conf. dec. contr.* (pp. 535–536).

Horn, R. A., & Johnson, C. R. (2013). *Matrix analysis* (2nd ed.). Cambridge Univ. Press.

Jarvis-Wloszek, Z., Feeley, R., Tan, W., Sun, K., & Packard, A. (2005). Control applications of sum of squares programming. In *Positive polynomials in control*. Springer.

Ji, X., & Su, H. (2016). An extension of Petersen's lemma on matrix uncertainty. *IEEE Transactions on Automatic Control*, *61*(6), 1655–1657.

Khalil, H. K. (2002). *Nonlinear systems* (3rd ed.). Prentice Hall.

Khlebnikov, M. V., & Shcherbakov, P. S. (2008). Petersen's lemma on matrix uncertainty and its generalizations. *Automation and Remote Control*, *69*(11), 1932–1945.

Löfberg, J. (2004). YALMIP: A toolbox for modeling and optimization in MATLAB. In *Proc. IEEE int. symp. comp. aid. contr. sys. des*.

Löfberg, J. (2009). Pre- and post-processing sum-of-squares programs in practice. *IEEE Transactions on Automatic Control*, *54*(5), 1007–1011.

Mania, H., Tu, S., & Recht, B. (2019). Certainty equivalence is efficient for linear quadratic control. In *Proc. 33rd int. conf. neural inf. proc. sys*.

Martin, T., & Allgöwer, F. (2021). Dissipativity verification with guarantees for polynomial systems from noisy input-state data. *IEEE Control Systems Letters*, *5*(4), 1399–1404.
Milanese, M., & Novara, C. (2004). Set membership identification of nonlinear systems. *Automatica*, *40*(6), 957–975.
Petersen, I. R. (1987). A stabilization algorithm for a class of uncertain linear systems. *Systems & Control Letters*, *8*(4), 351–357.
Petersen, I. R., & Hollot, C. V. (1986). A Riccati equation approach to the stabilization of uncertain linear systems. *Automatica*, *22*(4), 397–411.
Rantzer, A. (2001). A dual to Lyapunov's stability theorem. *Systems & Control Letters*, *42*(3), 161–168.
Recht, B. (2019). A tour of reinforcement learning: The view from continuous control. *Annual Review of Control, Robotics, and Autonomous Systems*, *3*, 253–279.
Shcherbakov, P. S., & Topunov, M. V. (2008). Extensions of Petersen's lemma on matrix uncertainty. *IFAC Proceedings Volumes*, *41*(2), 11385–11390.
Sontag, E. D. (1990). Further facts about input to state stabilization. *IEEE Transactions on Automatic Control*, *35*(4), 473–476.
Tan, W. (2006). *Nonlinear control analysis and synthesis using sum-of-squares programming* (Ph.D. thesis), Berkeley: Univ. of California.
Tanaskovic, M., Fagiano, L., Novara, C., & Morari, M. (2017). Data-driven control of nonlinear systems: An on-line direct approach. *Automatica*, *75*, 1–10.
Treven, L., Curi, S., Mutnỳ, M., & Krause, A. (2021). Learning stabilizing controllers for unstable linear quadratic regulators from a single trajectory. In *Proc. 3rd conf. learn. dynam. contr.* (pp. 664–676).
Verhaegen, M., & Verdult, V. (2007). *Filtering and system identification: A least squares approach*. Cambridge Univ. Press.
van Waarde, H. J., & Camlibel, M. K. (2021). A matrix Finsler's lemma with applications to data-driven control. In *Proc. IEEE conf. dec. contr*.
van Waarde, H. J., Camlibel, M. K., & Mesbahi, M. (2021). From noisy data to feedback controllers: Nonconservative design via a matrix S-lemma. *IEEE Transactions on Automatic Control*, *67*(1), 162–175.
Willems, J. C., Rapisarda, P., Markovsky, I., & De Moor, B. (2005). A note on persistency of excitation. *Systems & Control Letters*, *54*(4), 325–329.
Xue, A., & Matni, N. (2021). Data-driven system level synthesis. In *Proc. 3rd conf. learn. dynam. contr*.

**Andrea Bisoffi** received the M.Sc. degree in Automatic Control Engineering from Politecnico di Milano, Italy, in 2013 and the Ph.D. degree in Mechatronics from University of Trento, Italy, in 2017. He was a post-doctoral researcher at KTH Royal Institute of Technology, Sweden, from 2017 to 2019, and is currently one at University of Groningen, The Netherlands. His research interests include hybrid and nonlinear control systems, with applications to mechanical systems, and data-driven control.

**Claudio De Persis**, is a Professor with the Engineering and Technology Institute, Faculty of Science and Engineering, University of Groningen, The Netherlands. He received the Laurea and Ph.D. degrees in engineering in 1996 and 2000 both from the University of Rome "La Sapienza", Italy. His main research interest includes automatic control and its applications.

**Pietro Tesi** received the Ph.D. degree in computer and control engineering from the University of Florence, Italy, in 2010, where is currently an Associate Professor. Prior to that, he has been an Assistant Professor at the University of Florence, Italy, and the University Groningen, the Netherlands. His main research interests include adaptive and learning systems, data-driven control and network systems. Prof. Tesi serves in the Editorial Board for the International Journal of Robust and Nonlinear Control, and is an Associate Editor for the IEEE Control Systems Letters. He is also a member of the IFAC Technical Committee on Networked Systems. Dr. Tesi is a recipient of the 2021 IEEE Control Systems Letters Outstanding Paper Award.