Approximate Nonlinear Regulation via Identification-Based Adaptive Internal Models

Michelangelo Bin, Pauline Bernard, and Lorenzo Marconi

Abstract—This paper concerns the problem of adaptive output regulation for multivariable nonlinear systems in normal form. We present a regulator employing an adaptive internal model of the exogenous signals based on the theory of nonlinear Luenberger observers. Adaptation is performed by means of discrete-time system identification schemes where any algorithm fulfilling some optimality and stability conditions can be used. Practical and approximate regulation results are given relating the prediction capabilities of the identified model to the asymptotic bound on the regulated variable, which become asymptotic whenever a “right” internal model exists in the identifier’s model set. The proposed approach, moreover, does not require “high-gain” stabilization actions, thus qualifying as a suitable solution for practical implementation.

I. INTRODUCTION

In this paper we consider the problem of adaptive output regulation for multivariable nonlinear systems of the form

\[
\begin{align*}
\dot{z} &= f(w, z, x) \\
\dot{x} &= Ax + B(q(w, z, x) + b(w, z, x)u) \\
y &= Cx
\end{align*}
\]

(1)

where \((z, x) \in \mathbb{R}^n_z \times \mathbb{R}^n_x\) is the state of the plant, \(u\) and \(y\), both taking values in \(\mathbb{R}^n_u\), are the control input and the measured output, \(w \in \mathbb{R}^n_w\) is an exogenous input, \(f : \mathbb{R}^n_w \times \mathbb{R}^n_z \times \mathbb{R}^n_x \to \mathbb{R}^n_z\), \(q : \mathbb{R}^n_w \times \mathbb{R}^n_z \times \mathbb{R}^n_x \to \mathbb{R}^n_z\) and \(b : \mathbb{R}^n_w \times \mathbb{R}^n_z \times \mathbb{R}^n_x \to \mathbb{R}^{n_y \times n_y}\) are continuous functions and, for some \(r \in \mathbb{N}\), \((A, B, C)\) are defined as

\[
A := \begin{pmatrix} 0_{n_y \times n_y} & I_{(r-1)n_y} \\
0_{n_y \times (r-1)n_y} & 0_{n_y \times n_y} \end{pmatrix}, \quad B := \begin{pmatrix} 0_{(r-1)n_y \times n_y} \\
I_{n_y} \end{pmatrix}, \\
C := \begin{pmatrix} I_{n_y} \\
0_{n_y \times (r-1)n_y} \end{pmatrix},
\]

namely, \(n_x = rn_y\) and \(x\) is a chain of \(r\) integrators of dimension \(n_y\). The output regulation problem associated to system (1) consists in finding an output-feedback controller that ensures boundedness of the closed-loop trajectories whenever \(w\) is bounded, and that asymptotically removes the effect of \(w\) on the regulated output \(y\), thus ideally obtaining \(y(t) \to 0\) as \(t \to \infty\). Output regulation is representative of many problems of practical interest depending on the role played by the exogenous signal \(w\). For instance, simple stabilization is obtained when \(w\) is not present, disturbance rejection is achieved when \(w\) models disturbances acting on the plant, tracking is obtained when \(y\) represents the “error” between a given plant’s output and a reference trajectory dependent of \(w\), and some robust control problems are obtained whenever \(w\) represents uncertain parameters or unmodeled dynamics. As customary in the output regulation literature, we assume here that the exogenous signal \(w\) belongs to the set of solutions of an exosystem of the form

\[
\dot{w} = s(w),
\]

(2)

originating in a compact invariant subset \(W\) of \(\mathbb{R}^{n_w}\).

Output regulation is subject to the following taxonomy. Asymptotic regulation denotes the case in which the control objective is to ensure \(\lim_{t \to \infty} y(t) = 0\). Approximate regulation denotes the case in which the control objective is relaxed to \(\lim \sup_{t \to \infty} |y(t)| \leq \varepsilon^*\), with \(\varepsilon^*\) that represents some performance specification or optimality condition. Practical regulation, instead, refers to the case in which \(\lim \sup_{t \to \infty} |y(t)|\) can be reduced arbitrarily by opportune tuning the regulator. When one of the above control objectives is achieved in spite of uncertainties in the plant’s model, we call it robust regulation. When some learning mechanism is introduced to compensate for uncertainties in the exosystem, the problem is typically referred to as adaptive regulation. Asymptotic output regulation is a rich research area with a well-established theoretical foundation. For linear systems a complete formalization and solution to the problem has been given in the mid 70s in the seminal works by Francis, Wonham and Davison (see e.g. [1], [2]), where the well-known internal model principle was first stated. Asymptotic output regulation for (single-input-single-output) nonlinear systems has been under investigation since the early 90s, first in a local context [3], [4], [5], [6], and lately in a purely nonlinear framework [7], [8], [9] based on the “non-equilibrium” theory [10]. In more recent times, asymptotic regulators have been also extended to some classes of multivariable nonlinear systems (see e.g. [11], [12], [13]).

One of the major limitations of the existing asymptotic regulators is their complexity: the sufficient conditions under which asymptotic regulation is ensured are typically expressed by equations whose analytic solution becomes a hard (if not impossible) task even for “simple” problems, with the consequence that the construction of the regulation quickly becomes unfeasible. As conjectured in [14], moreover, even if a regulator can be constructed, asymptotic regulation remains a fragile property that is lost at front of the slightest plant’s perturbation. This, in turn, motivates the interest towards approximate, practical and adaptive solutions, sacrificing the asymptotic property to gain robustness and practical feasibility. Among the approaches to approximate regulation it is worth mentioning [15], [16], whereas practical regulators can be found in [17], [18], [13]. Adaptive designs of regulators can...
be found, e.g., in [19], [20], [21], where linearly parametrized internal models are constructed in the context of adaptive control, in [22] where discrete-time adaptation algorithms are used in the context of multivariable linear systems, and in [23], [24], [25] where adaptation of a nonlinear internal model is approached as a system identification problem.

A further limitation, present in most of the aforementioned designs and representing a major obstacle to practical implementation, is that the stabilization techniques used in the regulator employ control “gains” that need to be taken very large to ensure closed-loop stability, resulting in undesired “peaking” phenomena in the transitory, amplification of noise, and exaggerate strength and rigidity in the counteraction of disturbances. Moreover, the introduction of internal model units and adaptation mechanisms typically leads to a further increase of the gain, namely one has to “pay” in terms of stabilization for introducing additional complexity potentially leading to better asymptotic performance. This, in turn, makes the output regulation theory not much interesting in practical stabilization, has been extended in [12] to a class of multivariable systems, where the controller is augmented by an internal model which also allows one to deal with (possibly asymptotic) output regulation problems. Although theoretically appealing, the design of [12] is not constructive, in the sense that only an existence result of the internal model unit is given and no constructive design conditions are given even for simple problems.

In this paper we start from the idea of [18] and [12] to construct a regulator for multivariable nonlinear systems embedding an internal model unit that is adapted at run-time on the basis of the measured closed-loop signals. Compared to [18], we consider multivariable regulation problems rather than single-variable practical stabilization. Compared to [12], we confer on the internal model unit the ability to adapt online, thus proposing a control solution which is constructive and does not rely on fragile analytical conditions as typically required by non-adaptive designs. Besides, unlike in [12], we ensure that the parameters of the controller are fixed a priori independently from the added internal model. On the heels of [23], [24], and contrary to canonical adaptive control designs, adaptation is not carried by means of “ad hoc” algorithms developed under structural assumption on the internal model unit and by means of Lyapunov-like arguments; rather we approach the adaptation of the internal model as a system identification problem, where the best model matching with the measured data and performance needs to be identified. We thus allow for different identification schemes to be used, by individuating a set of sufficient stability conditions that they need to satisfy to be used within the framework. As in [22], we consider here identifiers that are discrete-time, which turn the closed-loop system into a hybrid system. Despite the additional complexity in the analysis, this choice is motivated by the fact that identification schemes are typically discrete-time, and that in this way we also structurally support adaptive mechanisms working on sampled data.

The paper is organized as follows. In Section II we describe the standing assumptions and we further discuss the previous results and the contribution of the paper. In Section III we present the proposed regulator and in Section IV we state the main result of the paper, proved later in Section V. In Section VI we construct some identifiers for linear and nonlinear parametrizations and, finally, in Section VII we present a numerical example.

Notation: We denote by \( \mathbb{R} \) and \( \mathbb{N} \) the sets of real and natural numbers, \( \mathbb{R}_+ := [0, \infty) \) and \( \mathbb{N}^* := \mathbb{N} \setminus \{0\} \). When the underlying metric space is clear, we denote by \( \mathbb{B}_r \) the open ball of radius \( r \) and, if \( B \) is a set, we denote by \( \mathbb{B}_r^B \) the open ball of radius \( r \) around \( B \). If \( S \) is a set, \( \overline{S} \) denotes its closure. If \( B \) is another set, \( S \subseteq B \)(resp. \( S \subset B \)) means \( S \) is contained (resp. strictly contained) in \( B \). Norms are denoted by \( \| \cdot \| \), if \( A \subset \mathbb{R}^n \), \( |x|_A := \inf_{a \in A} |x - a| \) denotes the usual distance of \( x \in \mathbb{R}^n \) to \( A \). For \( x : \mathbb{N} \to \mathbb{R}^n \) (resp. \( \mathbb{R}^n \)), we denote \( x_j := \sup_{i \leq j} |x(i)| \) (resp. \( |x|_t := \sup_{s \leq t} |x(s)| \)). If \( A \subset \mathbb{R}^n \), we let for convenience \( |x|_{A,j} := \|x|_A \|_j \) (resp. \( |x|_{A,t} := \|x|_A \|_t \)). If \( A_1, \ldots, A_m \) are matrices, we let \( \text{diag}(A_1, \ldots, A_m) \) and \( \text{col}(A_1, \ldots, A_m) \) their block-diagonal and column concatenation respectively. We denote by \( \text{SPD}_n \) the set of positive semi-definite symmetric matrices of dimension \( n \). A function \( \kappa : \mathbb{R}_+ \to \mathbb{R}_+ \) is said to be of class-\( K \) (\( \kappa \in K \)) if it is continuous, strictly increasing, and \( \kappa(0) = 0 \). A function \( \kappa \in K \) is said to be of class-\( K_\infty \) (\( \kappa \in K_\infty \)) if \( \lim_{s \to \infty} \kappa(s) = \infty \). A function \( \beta : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+ \) is said to be of class-\( K_\infty \) (\( \beta \in K_\infty \)) if there is a \( \kappa \in K_\infty \) such that \( \beta(s,t) \leq \kappa(s) + \kappa(t) \) for each \( t \in \mathbb{R}_+ \) and, for each \( s \in \mathbb{R}_+ \), \( \beta(s, \cdot) \) is continuous and strictly decreasing to zero as \( t \to \infty \). A set \( A \subset \mathbb{R}^n \) denotes a set-valued map. A function \( f : \mathbb{R}^n \to \mathbb{R}^m \) is \( C^k \) if \( k \) times continuously differentiable. If \( h : \mathbb{R}^n \to \mathbb{R} \) is \( C^1 \) and \( f : \mathbb{R}^n \to \mathbb{R}^m \), for each \( i \in \{1, \ldots, n\} \) we denote by \( L_{f(x)}h \) the map \( x \mapsto L_{f(x)}h(x) := \partial h/\partial x_i(x) \cdot f(x) \). When the superscript \( (x_i) \) is obvious, it is omitted.

In this paper we deal with hybrid systems, i.e., systems that combine discrete- and continuous-time dynamics. They are formally described by equations of the form [20]

\[
\Sigma : \begin{cases}
  \dot{x} &= F(x,u) \\
  x^+ &= G(x,u)
\end{cases}
\]

where \( F \) and \( G \) denote the flow and jump maps and \( C \) and \( D \) the sets in which flows and jumps are allowed. Solutions to (3) are defined over hybrid time domains. A compact hybrid time domain is a subset of \( \mathbb{R}_+ \times \mathbb{N} \) of the form \( T = \bigcup_{j=1}^J [t_j, t_{j+1}] \times \{j\} \) for some finite \( J \in \mathbb{N} \) and \( 0 = t_0 \leq t_1 \leq \cdots \leq t_J \in \mathbb{R}_+ \). A set \( T \subset \mathbb{R}_+ \times \mathbb{N} \) is called a hybrid time domain if for each \( (T,J) \in \mathbb{R}_+ \times \mathbb{N} \) \( T \cap [0,T] \times \{1, \ldots, J\} \) is a compact hybrid time domain. If \( (x_i, s, i) \in T \), we write \( (t, j) \leq (s, i) \) if \( t + j \leq s + i \). For any \( (t, j) \in T \), we let \( t^j = \sup \{ t \in \mathbb{R}_+ | (t, j) \in T \} \), \( t_j := \inf \{ t \in \mathbb{R}_+ | (t, j) \in T \} \) and \( j^t \) in the same way. A function \( x : T \to X \) defined on a hybrid time domain \( T \) is called a
For a hybrid input in an open set including \( u \), \( A2 \) could be relaxed to assume that \( A \) is “just” locally exponentially stable for \( \theta \), provided that the only component of \( x \) that affects \( f \) is \( y = Cx \). \( A3 \) is instead a stabilizability assumption taken from \( [18] \), \( [12] \) and asking the designer to have available an estimate \( b \) of \( w(z, x) \) which captures enough information on its behavior. \( A3 \) in particular, implies that \( b(w, z, x) \) is nonsingular at each \( (w, z, x) \). We also remark that \( A3 \) could be weakened to a “local” version, i.e. requiring that a pair \((b, \mu)\) fulfilling \( 5 \) exists for each compact subset of \( W \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_z} \).

**B. Previous Approaches**

There follows by the structure of \((1), (2)\) that, under \( \Lambda2 \) the problem of asymptotic regulation could be in principle solved by a control law of the kind

\[
 u = -b(w(z, x))^{-1}q(w(z, x)) + b(w, z, x)^{-1}k(x),
\]

(7)

where the term \( -b(w(z, x))^{-1}q(w(z, x)) \) represents a non-vanishing “feedforward” action compensating for the influence of the dynamics of \((w, z)\) on \( \hat{x} \), and \( k(x) \) is a stabilizing control action vanishing with \( x \). However, \( 7 \) cannot be directly implemented even if the whole state \((z, x)\) were accessible, as it anyway would require \( w \) to be measured and the functions \( q \) and \( b \) to be perfectly known. To overcome those issues, in \( [13] \) the authors proposed a dynamic regulator in a single-variable context (i.e. \( n_y = 1 \)) where \( b \) and \( q \) in \( 7 \) are approximated by functions \( x \rightarrow \hat{q}(x) \) and \( x \rightarrow b(x) \) of \( x \) only, and an extended observer is introduced to provide an estimate \( \hat{x} \) of \( x \) and to compensate for the mismatch of \( \hat{b} \) and \( \hat{q} \) with the actual quantities. The control action in particular was taken as

\[
 u := \text{sat} \left( \hat{b}(\hat{x})^{-1} \left( -\hat{q}(\hat{x}) + k(\hat{x}) - \hat{\sigma} \right) \right),
\]

(8)

where \( \text{sat} \) is a suitably chosen saturation function and \( \hat{\sigma} \) is the term of the extended observer compensating for the mismatch between \( (\hat{q}(\hat{x}), \hat{b}(\hat{x})) \) and \( (q(w(z, x)), b(w, z, x)) \). This regulator was proved to recover the performance of the ideal control law \( 4 \) theoretically as closely and quickly as desired, by increasing the observer gains accordingly. Nevertheless, the regulator of \( [13] \) does not embed any system able to generate the ideal feedforward term \( -b(w(z, x))^{-1}q(w(z, x)) \), which indeed can be only approximated by the extended observer. Therefore, the attained regulation result is only practical, with the observer gains that must be taken high enough to accommodate the desired asymptotic bound. This design thus has two main drawbacks. First, the ideal steady state in which \( y = 0 \) is not a trajectory of the system and, as such, it is not even stable, so that a considerable transitory is possible even if the system is initialized close to the desired operating point. Second, good performance are only obtained by increasing
the observer gains accordingly. As the observer gains grow, however, the peaking and the noise amplification grow, so that a compromise between regulation performance and noise amplification must be sought. A remarkable property of this approach is that the stabilizing action $k(x)$ is not forced to be “high-gain” and is fixed a priori in the “ideal” controller \( \eta \).

On the other hand, when $n_2 = n_q = 1$, it was shown in \[9\] that, under $\mathcal{A}_0$ and if $b(w, z, x)$ is lower bounded by a positive constant, the problem of asymptotic output regulation for \( \mathcal{A}_0 \) can always be solved by means of a controller of the form

\[
\begin{align*}
\dot{\eta} &= F\eta + Gu \\
\gamma &= \gamma(\eta) + \kappa(x),
\end{align*}
\]

with state $\eta \in \mathbb{R}^{n_\eta}$, $n_\eta = 2(n_z + n_w + 1)$, $(F, G)$ a controllable pair with $F$ a Hurwitz matrix, and with $\gamma : \mathbb{R}^{n_\eta} \to \mathbb{R}$ and $\kappa : \mathbb{R} \to \mathbb{R}$ suitably defined continuous functions. The term $\kappa(x)$ plays here the same role as $k(x)$ in \[6\], while the term $\gamma(\eta)$ is meant to reproduce the feedforward action $-b(w, z, x)^{-1}q^*(w, z, x)$ at the steady state. For this reason, the restriction of \[9\] to the set in which $x = 0$, namely

\[
\begin{align*}
\dot{\eta} &= F\eta + G\gamma(\eta), \\
u &= \gamma(\eta),
\end{align*}
\]

is called the internal model unit, as it is a copy of the process that generates the ideal feedforward action making the set where $y = 0$ invariant (property that the regulator of \[18\] does not have). This approach, however, has two main drawbacks: the stabilizing action $k(x)$ is necessarily high-gain to bring the system close to the steady state where $\gamma(\eta)$ behaves as desired, and even if $\gamma$ always exists, no analytical or numerical method exists to construct it even for simple problems.

In \[12\], the authors extended both the approaches of \[18\], \[9\] described above to the class of systems \[1\], \[2\]. The approach of \[12\], in particular, is based on an extension of the extended observer of \[18\] to multivariable systems where, however, $b$ is taken constant in \[8\] and equal to $b$ of \[A3\] and the term $b(\hat{x})^{-1}q^*(\hat{x})$ is substituted by the output $\gamma(\eta)$ of an internal model unit of the kind \[9\], appropriately extended to fit the multivariable setting. Then, $u$ is taken as

\[
\begin{align*}
u &= \gamma(\eta) + b^{-1}(\text{sat}(\hat{\sigma}) + k(x)).
\end{align*}
\]

Compared to \[18\], this design is potentially asymptotic (whenever \[9\] is chosen correctly), thus yielding $y \to 0$ without taking the observer gains inconveniently large. Compared to \[9\], apart from the extension to multivariable normal forms, the approach of \[12\] allows one to use stabilization control actions that are not high-gain, thus qualifying as an alternative more suitable for practical implementation. However, the problems related to the construction of $\gamma$ inherited from \[9\] persist, with the consequence that, although theoretically appealing, the approach of \[12\] is not constructive. Besides, the saturation level of the map $\text{sat}$ depends on the choice of internal model, and in particular of $\gamma$ itself, and on the initial error in the initialization of $\eta$ relative to its (unknown) ideal steady state. Some existing methods to approximate $\gamma$ have been proposed in \[13\], yet their implementation remains tedious and the computational complexity easily grows with the desired precision and the dimension of the problem. Otherwise, adaptive designs exist that tune $\gamma$ online (see \[21\], \[25\]), yet they are far from a definite answer and are all based on high-gain stabilization.

C. Contribution of the paper

In this paper we present a regulator embedding an adaptive internal model unit and non-high-gain stabilization actions, by thus merging all the desired properties mentioned before. Adaptation is cast as a discrete-time system identification problem \[27\] defined over samples of the closed-loop system trajectories. Instead of developing a single ad hoc adaptation algorithm, we give sufficient conditions under which arbitrary identification schemes can be used. We then specifically develop the relevant case of weighted least squares for linear parametrizations and mini-batch algorithms for nonlinear parametrizations, thus embracing many existing and frequently-used techniques performing white- and black-box identification. The proposed regulator is proved to achieve both practical and approximate regulation, with an asymptotic bound that is directly related to the prediction capabilities of the identifier. Hence, the result becomes asymptotic whenever the identified model is perfect. Compared to \[18\], the proposed regulator has the ability to learn online and employ an internal model unit reproducing the ideal feedforward action making the set in which $y = 0$ asymptotically stable. Compared to \[9\], the proposed approach does not rely on high-gain stabilization and, compared to \[9\] and \[12\], we introduce adaptation of internal model, identification and observer units.

III. The Regulator

The proposed regulator is a hybrid system described by

\[
\begin{align*}
\dot{\varsigma} &= 1 \\
\dot{\eta} &= F\eta + Gu \\
\dot{x} &= A\hat{x} + B(\hat{\sigma} + bu) + \Lambda(\ell)H(y - \hat{x}_1) \\
\dot{\hat{x}} &= -b\psi(\theta, \eta, u) + \ell^{\sigma+1}H_{\ell+1}(y - \hat{x}_1) \\
\dot{\xi} &= 0 \\
\dot{\theta} &= 0 \\
(\varsigma, \eta, \hat{x}, \hat{\sigma}, \xi, \theta, y) &\in C_\varsigma \times \mathbb{R}^n + \mathbb{R}^n + \mathbb{R}^n + \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \\
\varsigma^+ &= 0, \quad \eta^+ = \eta \\
\hat{x}^+ &= \hat{x}, \quad \hat{\sigma}^+ = \hat{\sigma} \\
\xi^+ &= \varphi(\xi, \eta, u) \\
\theta^+ &= \theta(\xi) \\
(\varsigma, \eta, \hat{x}, \hat{\sigma}, \xi, \theta, y) &\in D_\varsigma \times \mathbb{R}^n + \mathbb{R}^n + \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \\
\end{align*}
\]

and with output

\[
\begin{align*}
u &= b^{-1}(\text{sat}(\hat{\sigma} + \kappa(\hat{x}))),
\end{align*}
\]

where $b$ and $(A, B)$ are the same matrices of $\mathcal{A}_0$ and $\mathcal{A}$ respectively, $n_\eta \in \mathbb{N}$, $\Xi$ and $\Theta$ are finite-dimensional normed vector spaces, $(F, G) \in \mathbb{R}^{n_z \times n_\eta} \times \mathbb{R}^{n_\eta \times n_b}$ and $(\Lambda(\ell), H, H_{\ell+1}) \in \mathbb{R}^{n_b \times n_x} \times \mathbb{R}^{n_x \times n_y} \times \mathbb{R}^{n_y \times n_b}$ are matrices to be defined, $\ell \in \mathbb{R}_+$ is a control parameter, $\psi : \Theta \times \mathbb{R}^n \times
Figure 1. Block-diagram of the regulator.

\[ \mathbb{R}^{n_y} \rightarrow \mathbb{R}^{n_y}, \varphi : \Xi \times \mathbb{R}^{n_y} \times \mathbb{R}^{n_y} \rightarrow \Xi, \text{sat} : \mathbb{R}^{n_y} \rightarrow \mathbb{R}^{n_y}, \]
\[ \kappa : \mathbb{R}^{n_y} \rightarrow \mathbb{R}^{n_y}, \theta : \Xi \rightarrow \Theta \] functions to be defined and, with \( T, t \in \mathbb{R}_+ \) satisfying \( 0 < T \leq T \)
\[ C_\xi := [0, T], \quad D_\xi := [T, T]. \]

The subsystem, whose block-diagram is depicted in Figure 1, is composed of: a) a purely continuous-time subsystem \((\eta, \hat{x}, \sigma)\) whose dynamics depends on a parameter \( \theta \) that is constant during flow, b) a purely discrete-time subsystem \((\xi, \theta)\) updated at jump times, c) a hybrid clock \( \varsigma \) whose tick triggers those jumps, namely the updates of the parameter \( \theta \). The dynamics of the clock is decided by the user. The flow and jump sets \( C_\xi \) and \( D_\xi \), in turn, allow the usage of any, possibly aperiodic, strategy for the clock, by just forcing lower and upper bounds on the distance of two successive jump times. The subsystem \( \eta \), taking values in \( \mathbb{R}^{n_y} \), plays the role of an internal model unit, and is taken of the same form as \( \varphi \). The subsystem \((\dot{x}, \sigma)\), taking values in \( \mathbb{R}^{n_x+n_y} \), is an extended observer of the system (13), but with an additional “consistency term” \( -b\psi(\theta, \eta, u) \) which, as better clarified later, represents the output of the internal model unit. The subsystem \( \xi \), taking values in \( \Xi \), is the identifier, whose updates take place at jump times. The variable \( \theta \), taking values in \( \Theta \), is the identifier’s output, and it is included as a state in (11) to formalize the fact that it only changes at jump times. In the rest of the section we detail the construction of all these subsystems, along with all the degrees of freedom introduced in (11). In doing so, we make reference to a given arbitrary set of initial conditions for (11), (12) of the form \( W \times Z_0 \times X_0 \subset \mathbb{R}^{n_w} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \).

Remark 1: We underline that, contrary to (9), (12), the output \( \eta \) of the internal model unit does not enter directly in the definition of \( u \) (compare (12) with (9), (10)), but only in the dynamics of \( \sigma \) through the map \( \psi \). As it will be clarified in the next subsection, unlike in (12) this allows us to fix the saturation level of (12) independently from the extended observer, the internal model and the identifier.

A. The Stabilizing Action

In this section we fix the functions \( \kappa \) and \( \text{sat} \) in (12). The function \( \kappa \) is chosen as any \( C^1 \) function such that the system
\[ \dot{x} = Ax + B\kappa(x) + B\delta \] (13)
is ISS relative to the origin and with respect to \( \delta \) with locally linear asymptotic gains. Namely, we argue that there exist \( \beta_x \in KL \) and a locally Lipschitz \( \rho_x \in K \) such that
\[ |x(t)| \leq \max\{\beta_x(|x(0)|, t), \rho_x(|\delta|, t)\} \]
for all \( t \in \mathbb{R}_+ \). For instance, \( \kappa \) can be chosen as \( \kappa(x) = Kx \), with \( K \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \) such that \( A + BK \) is Hurwitz. There follows from A2 that the system (13) is ISS relative to the set
\[ B := A \times \{0\} = \{(w, z, x) \in A \times \mathbb{R}^{n_x} | x = 0\} \]
and with respect to the input \( \delta \). Let \((Z_0, X_0)\) be the sets of initial conditions for (11) and \( \varphi_0 \) a positive scalar such that \((W \times Z_0 \times X_0) \subset B_\varphi \). With \( \delta \) and \( \varphi_0 \) arbitrary positive scalars, there exists a compact set \( \Omega_0 \subset \mathbb{R}^{n_w} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \)
satisfying
\[ (W \times Z_0 \times X_0) \subset \mathbb{R}^{\varphi_0} \subset \mathbb{R}_+^{\varphi_0} \subset \Omega_0. \] (14)
and such that any trajectory of the system (4), (13) originating in \( \mathbb{R}^{\varphi_0} \) and with an input \( \delta \) satisfying \( |\delta| \leq \delta \), is completed, and fulfills \((w(t), z(t), x(t)) \in \Omega_0 \) for all \( t \in \mathbb{R}_+ \).

Let \( c : W \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_y} \) be defined as
\[ c(w, z, x) := -b(w, z, x)^{-1}g(w, z, x), \]
and, with \( \varphi_0 > 0 \) arbitrary, let \( M \) be any constant fulfilling
\[ M \geq \max_{(w, z, x) \in \Omega_0} |bc(w, z, x) + b(w, z, x)^{-1}g(w, z, x)| + \varphi_0. \] (15)

Then we define \( \text{sat}(\cdot) \) as any \( C^1 \) bounded function satisfying
\[ 0 \leq |(w, z, x) \in \mathbb{R}^{n_y} \]
\[ \text{sat}(s) = s \quad \forall s \in \mathbb{R}_+. \] (16)

Remark 2: The definition of \( M \) requires the knowledge of a bound on the maximum value that the functions \( bc(w, z, x) \)
and \( b(w, z, x)^{-1}g(w, z, x) \) attain in \( \Omega_0 \). While knowing a bound of \( c(w, z, x) \) is a quantitative information related to the plant, and in particular on the ideal feedback control action in a neighborhood of the set \( \mathbb{R}_+ \), the knowledge of a bound for \( b(w, z, x)^{-1}g(w, z, x) \) does not ask for any additional information. In fact, \( \kappa \) is known to the designer, while we have
\[ |b(w, z, x)^{-1}g(w, z, x)| \leq \mu^{-1} \]
for all \((w, z, x) \in W \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \) with \( \mu \) defined in \( A \). Indeed, \( b(w, z, x)^{-1} = (I + (b(w, z, x) - b)b^{-1})^{-1} \), so that by Proposition 10.3.2, \( |b(w, z, x)^{-1}| \leq (1 - |(b(w, z, x) - b)b^{-1}|) \leq \mu^{-1}. \)

B. The Internal Model Unit

The restriction of \( c \) on \( B \), which we denote by
\[ u^*(w) := c(w, \pi(w), 0), \] (17)
represents the steady-state value of the ideal feedback action \( c \) when \( y \) vanishes, i.e., \( u^*(w) \) is the control action that makes the set \( B \) invariant for (11). (2). The internal model unit \( \eta \) is a system constructed to be able to generate \( u^*(w) \) when \( y = 0 \), and its construction follows the approach of (9), where the dimension \( n_\eta \) of the state \( \eta \) is chosen as
\[ 1 \text{All the subsequent results can be proved even if sat is differentiable a.e.; the } C^1 \text{ requirement, in turn, is required to simplify the forthcoming analysis.} \]
Let $n_\eta = 2(n_w + n_z + 1)$, and the pair $(F, G)$ is taken as a real realization of any complex pair $(F_c, G_c)$ of dimension $n_w + n_z + 1$, with $G_c$ a matrix with non-zero entries and $F_c$ a diagonal matrix whose eigenvalues have sufficiently negative real part. More precisely, this choice is legitimated by the following lemma, which is a direct consequence of [9].

**Lemma 1:** Suppose that $A2$ holds and let $n_\eta = 2(n_w + n_z + 1)$. Then there exist a controllable pair $(F, G) \in \mathbb{R}^{n_\eta \times n_\eta} \times \mathbb{R}^{n_\eta \times n_w}$, with $F$ a Hurwitz matrix, and continuous maps $\tau : \mathbb{R}^{n_w} \to \mathbb{R}^{n_\eta}$ and $\gamma : \mathbb{R}^{n_\eta} \to \mathbb{R}^{n_\eta}$ such that

$$\gamma \circ \tau(w) = u^*(w) \quad \forall w \in W \quad (18)$$

and, for every input $\gamma(x, \delta_1, \delta_2) \in \mathbb{R}^{n_\eta} \times \mathbb{R}^{n_\eta} \times \mathbb{R}^{n_\eta}$ satisfying $|\delta_1| \leq \rho_0(|w, z, x|)$ for some $\rho_0 > 0$, the system

$$\begin{align*}
\dot{\eta} &= s(w) \\
\dot{z} &= f(w, z, x) \\
\dot{w} &= F \eta + G u^*(w) + \delta_1 + \delta_2
\end{align*} \quad (19)$$

is forward complete and it is ISS relative to the set $D : = \{(w, z, \eta) \in A \times \mathbb{R}^{n_\eta} \mid \eta = \tau(w)\}$, with respect to the input $(x, \delta_2)$ with locally Lipschitz asymptotic gains.

Lemma 1 implies the existence of $\beta_1 \in \mathbb{L}$ and a locally Lipschitz $\rho_1 \in \mathbb{K}$ such that every solution pair to (19) originating in $W \times \mathbb{R}^{n_\eta} \times \mathbb{R}^{n_\eta}$ satisfies

$$|(w(t), z(t), \eta(t))|_D \leq \max \{\beta_{11}((w(0), x(0), \eta(0))|_D, t), \rho_{11}((x, \delta_2)|_t)\},$$

for all $t \in \mathbb{R}_+$. System (19) is the zero dynamics, relative to the input-output pair $(u, y)$, of the plant augmented with the system $\eta$, and the result of Lemma 1 states that, in the zero dynamics set $D$, we have $\gamma(\eta) = u^*(w)$, i.e., the set

$$E : = \{(w, z, x, \eta) \in B \times \mathbb{R}^{n_\eta} \mid (w, z, \eta) \in D\} \quad (20)$$

is made invariant for the augmented system with $\delta_2 = 0$ by the input $\eta = \gamma(\eta)$. The role of the input $\delta_2$ will be clarified in the forthcoming sections. The map $\gamma$ in (18), which is the same as in $\Theta$, is introduced here to support the subsequent analysis and we stress that it is not used in the construction of the regulator, yet possibly only as a qualitative information guiding the designer in the choice of the model set for the identifier, as discussed below. The actual map $\psi$, through which $\eta$ affects the extended observer, will be instead defined in Section III-D.

### C. The Identifier

The identifier is a discrete-time system aimed to produce an estimate of the map $\gamma$ introduced in the previous section. The estimation of $\gamma$ is cast here as a system identification problem [27], and the particular design of the degrees of freedom $(\Xi, \psi, \Theta, \vartheta)$ corresponds to a choice of a given identification algorithm. What is the right identification algorithm to use, in turn, is a question whose answer strongly depends on the a priori information that the designer has on the plant, on the exosystem, and on the kind of uncertainties expected in the different models. In this paper we do not intend to limit to a single choice, which may be good in some settings and inappropriate in others, rather we give here a set of sufficient conditions, gathered in what we called the identifier requirement, representing the stability and optimality properties that any identification algorithm needs to possess to be used in the framework. We postpone examples of identifiers to Section V.

The identification problem underlying the design of the identifier is cast on the samples of the following core process

$$\begin{align*}
\zeta &= 1 \\
\dot{\zeta} &= s(\zeta) \quad (\zeta, w) \in C_c \times W \\
\dot{\zeta}^+ &= 0 \\
\dot{w}^+ &= w \quad (\zeta, w) \in D_c \times W,
\end{align*} \quad (21)$$

with outputs

$$\alpha_{in}(j) : = \tau(w(t^j)), \quad \alpha_{out}(j) : = u^*(w(t^j)), \quad (22)$$

where $u^*$ and $\tau$ are defined respectively in (17) and (18). In particular, the identifier is aimed at finding the “best” model $\hat{\gamma}$ relating the input-output data pairs $\{\alpha_{in}(j), \alpha_{out}(j)\}$, $j \in \mathbb{N}$, obtained by sampling $(\tau(w), u^*(w))$ during jumps. The first step in the construction of the identifier is the definition of a model set $M$, which is a space of functions where $\hat{\gamma}$ is supposed to range. As customary in the system identification literature, and due to obvious implementation constraints, we limit here to the case in which $M$ is finite-dimensional. This, in turn, allows us to parametrize $\hat{\gamma}$ by a parameter $\theta$ ranging in a finite-dimensional vector space $\Theta$, thus yielding

$$M : = \{\hat{\gamma}(\theta, \cdot) : \mathbb{R}^{n_\eta} \to \mathbb{R}^{n_\eta} \mid \theta \in \Theta\}. \quad (23)$$

The choice of the model set, and hence of $\Theta$, is guided by the available knowledge on the core process (21)-22 and, in particular, on the expected relation between $\alpha_{in}$ and $\alpha_{out}$, ideally given by the unknown map $\gamma$ (see Lemma 1). Depending on the amount of information available, $M$ may range from a very specific set of functions, such as linear regressions, to a space of universal approximators, including for instance Wavelet bases or Neural Networks.

Once $M$ and $\Theta$ are fixed, a cost function is defined on the input-output data set generated by (21)-(22), so as to assign to each model $\hat{\gamma}(\theta, \cdot)$ a quantitative value describing how well it fits. In particular, for each solution $(\zeta, w)$ to the core process (21) and for each $j \in \text{dom}(\zeta, w)$ we define the functional

$$J(\zeta, w)(j, \theta) : = \sum_{i=0}^{j-1} g(\varepsilon(\theta, w(t^i)), i, j) + \rho(\theta), \quad (24)$$

where

$$\varepsilon(\theta, w) : = u^*(w) - \hat{\gamma}(\theta, \tau(w)) \quad (25)$$

denotes the prediction error attained by the model $\hat{\gamma}(\theta, \cdot) \in M$ along the solution $(\zeta, w)$ of (21), $g : \mathbb{R}^{n_\eta} \to \mathbb{R}_+$ is a positive function representing the local weight assigned to the term $(\varepsilon(\theta, w(t^i), i, j)$ in the sum, and $\rho : \Theta \to \mathbb{R}_+$ is possibly a regularization function. The particular choice of $g$ and $\rho$, which is left as a degree of freedom to the designer, characterizes the selection criteria for the best model $\hat{\gamma}(\theta, \cdot)$ and, hence, individuates a set of possible algorithms that can be used.
With \( \text{(24)} \) we associate the set-valued map
\[
\text{Opt}(\varsigma, w)(j) := \arg\min_{\theta \in \Theta} J(\varsigma, w)(j, \theta),
\]
representing, at each \( j \), the set of optimal parameters according to \( \text{(24)} \). The choice of the remaining degrees of freedom \( (\Xi, \varphi, \Theta, \theta) \) is then made to satisfy the conditions contained in the forthcoming requirement, in which we make reference to the following cascade of the core process to the identifier:
\[
\begin{align*}
\dot{\varsigma} & = 1 \\
\dot{w} & = s(w) \\
\dot{\xi} & = 0, \quad \vartheta = 0 \\
(\varsigma, w, \xi, \vartheta, d_{\text{in}}, d_{\text{out}}) & \in C_\varsigma \times W \times \Xi \times \Theta \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_v}
\end{align*}
\]
\( (\varsigma, w, \xi, \vartheta, d_{\text{in}}, d_{\text{out}}) \) are compact sets satisfying
\[
\begin{align*}
\varsigma^+ & = 0 \\
w^+ & = w \\
\xi^+ & = \varphi(\xi, \alpha_{\text{in}}(w) + d_{\text{in}}, \alpha_{\text{out}}(w) + d_{\text{out}}) \\
\vartheta^+ & = \vartheta(\xi)
\end{align*}
\]for all \( \varsigma, w, \xi, \vartheta, d_{\text{in}}, d_{\text{out}} \in D_\varsigma \times W \times \Xi \times \Theta \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_v}, \)
\( (\varsigma, w, \xi, \vartheta, d_{\text{in}}, d_{\text{out}}) \) is a positive control parameter that has to be taken large enough to ensure closed-loop stability, and it will be fixed in the forthcoming Theorem \( \text{I} \). The matrix \( \Lambda(\ell) \) is chosen as \( \Lambda(\ell) := \text{diag}(H_{\text{in}}, \ell^2 I_{n_{\text{out}}}, \ldots, \ell^r I_{n_{\text{out}}}) \). For each \( i = 1, \ldots, r + 1 \) and \( j = 1, \ldots, n_y \), let \( h^i_j \in \mathbb{R} \) be such that, for each \( i = 1, \ldots, n_y \), the roots of the polynomials \( \lambda^{r+1} + h^i_1 \lambda^r + \cdots + h^i_r \lambda + h^i_{r+1} \) are all real and negative. Then, the matrices \( H \) and \( H_{r+1} \) are defined as follows
\[
H := \text{diag}(H_1, \ldots, H_r), \quad H_{r+1} := \text{diag}(h^1_1, \ldots, h^r_{n_y})
\]
Finally, with \( \Xi^* \) given by the identifier requirement, we define \( \Theta^* := \vartheta(\Xi^*) \) and we let \( \mathcal{H}_* \subset \mathbb{R}^{n_u} \) and \( \mathcal{U}_* \subset \mathbb{R}^{n_y} \) be any compact sets satisfying \( \tau(\mathcal{W}) \subset \mathcal{H}_* \) and \( u^*(\mathcal{W}) \subset \mathcal{U}_* \). Then, we let \( \psi \) be any function satisfying
\[
\psi(\theta, \eta, u) = \frac{\partial \hat{\gamma}(\theta, \eta)}{\partial \eta}(F \eta + Gu)
\]
for all \( (\theta, \eta, u) \in \Theta^* \times \mathcal{H}_* \times \mathcal{U}_* \) and, for some \( \bar{\psi} > 0 \),
\[
|\psi(\theta, \eta, u)| \leq \bar{\psi}
\]
for all \( (\theta, \eta, u) \in \Theta \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_y} \).

## IV. Main Result

The closed-loop system, obtained by interconnecting the plant \( (11) \) to the regulator \( (11) \)–\( (12) \), results in the following hybrid system
\[
\begin{align*}
\dot{\varsigma} & = 1 \\
\dot{w} & = s(w) \\
\dot{z} & = f(w, z, x) \\
\dot{x} & = Ax + B(q(w, z, x) + b(w, z, x)u) \\
\dot{\eta} & = F \eta + Gu \\
\dot{\bar{x}} & = A\bar{x} + B(\bar{\sigma} + bu) + \Lambda(\ell)HC(x - \bar{x}) \\
\dot{\bar{\sigma}} & = -b\psi(\theta, \eta, u) + \ell^{r+1}H_{r+1}C(x - \bar{x}) \\
\dot{\xi} & = 0, \quad \vartheta = 0 \\
(\varsigma, w, z, \eta, \bar{x}, \bar{\sigma}, \xi, \theta) & \in C
\end{align*}
\]
\[
\begin{align*}
\varsigma^+ & = 0 \\
w^+ & = w, \quad z^+ = z, \quad x^+ = x, \\
\eta^+ & = \eta, \quad \bar{x}^+ = \bar{x}, \quad \bar{\sigma}^+ = \bar{\sigma} \\
\xi^+ & = \varphi(\xi, \eta, u) \\
\vartheta^+ & = \vartheta(\xi)
\end{align*}
\]
\( (\varsigma, w, z, \eta, \bar{x}, \bar{\sigma}, \xi, \theta) \in D \)
with $u$ given by (12) and with flow and jump sets given by $C := C_\xi \times W \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_\nu} \times \mathbb{R}^{n_\nu} \times \mathbb{R}^{n_\nu} \times \mathbb{R}^{n_\nu} \times \Xi \times \Theta$ and $D := D_\xi \times W \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_\nu} \times \mathbb{R}^{n_\nu} \times \mathbb{R}^{n_\nu} \times \mathbb{R}^{n_\nu} \times \Xi \times \Theta$.

In the remainder of the paper we let $\mathcal{O} := \{(\xi, w, z, x, \eta, \hat{x}, \hat{\sigma}) \in (C_\xi \cup D_\xi) \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_\nu} | \hat{x} = x, \hat{\sigma} = -bu^*(w)\}$.

with $\mathcal{E}$ the set defined in (29). Furthermore, for any solution $x := (\xi, w, z, x, \eta, \hat{x}, \hat{\sigma}, \xi, \Theta)$ of (29), and with $(\xi^*, \Theta^*)$ the trajectory produced by the identifier requirement relative to the solution pair $(x, \eta, \xi, \Theta)$, we have a common asymptotic bound. The second claim stating that, no matter how wrong the internal model and/or capabilities evaluated along the ideal data $\xi = (\hat{x}, \hat{\sigma})$, they have a common asymptotic bound. The second claim extending that of [18] and practical regulation (34) implies that $\limsup_{j \to \infty} |\theta(j) - \theta^*(j)| \leq \alpha_\theta \left( \limsup_{t \to \infty} |\varepsilon^*(t, j)| \right)$, with $\alpha_\theta(\cdot) = \rho_\theta(\alpha_x(\cdot))$, that explicitly express the asymptotic bound of the regulated variable $y$ and the parameter estimation error $\theta - \theta^*$ in terms of the optimal prediction error. Hence, as a consequence of the third claim, we also conclude that, whenever $\varepsilon^* = 0$, i.e. when the actual internal model belongs to the identifier model set, then asymptotic regulation and asymptotic parameter estimation are achieved, thus extending the existence result of (12) to the adaptive case.

Remark 3: We underline that the choice of the map $\kappa$ and sat detailed in Section III-A are independent from the observer, the internal model and the identifier. Besides, the result is global in $(\eta, \xi)$, which differs from [12] where the result is semi-global with respect to $\eta$, with the saturation in the controller that must be adapted to the initialization compact set of the internal model. On the other hand, the result is semi-global with respect to the observer, since the gain $\ell$ must be adapted to the observer initialization set $S_0$.

Remark 4: Assumptions [A1] and the consequent claim of Lemma 1 and the identifier requirement in Definition 1 all ask or state some Lipschitz conditions on maps that play primary roles in the stability analysis. Nevertheless, we observe that all these regularity conditions may be relaxed to less restrictive Hölder continuity requirements by substituting the high-gain-based extended observer presented in Section III-D with an homogeneous observer of appropriate degree. The reader is referred to [31] for further details.

V. ON THE DESIGN OF IDENTIFIERS

A. Least-Squares Identifiers for Linear Parametrizations

In this section we present a construction of the identifier when the model set $\mathcal{M}$ consists of functions $\tilde{\gamma}$ that are linear in the parameters $\theta$ and the cost functional (24) is a (weighted) least-squares norm of the past prediction errors. For ease of notation we focus here on the single-variable case (i.e. with $n_y = 1$), as a multivariable identifier can be obtained by concatenating of $n_y$ single-variable identifiers. We consider a model set $\mathcal{M}$ containing functions of the form

$$\tilde{\gamma}(\theta, \cdot) := \sum_{i=1}^{n_\theta} \theta_i \sigma_i(\cdot) = \theta^T \sigma(\cdot),$$

with $n_\theta \in \mathbb{N}$ arbitrary, $\theta = \text{col}(\theta_1, \ldots, \theta_{n_\theta}) \in \mathbb{R}^{n_\theta}$, and $\sigma = \text{col}(\sigma_1, \ldots, \sigma_{n_\theta})$, with $\sigma_i : \mathbb{R}^{n_\nu} \to \mathbb{R}$ differentiable functions with locally Lipschitz derivative. The "least-squares" cost-functional is obtained by letting in (24) $g(s, i, j) := \mu^2 j^{-i-1} |s|^2$, with $\mu \in (0, 1)$ a design parameter playing the role of a forgetting factor, and $\rho(\cdot) := \theta^T \Omega \theta$, in which $\Omega \in \text{SPD}_{n_\theta}$. Thus, $\bar{\mathcal{J}}$ reads as

$$\bar{\mathcal{J}}(\xi, w)(\theta) := \sum_{i=0}^{j-1} \mu^2 j^{-i-1} |\varepsilon^*(\theta, w(t^i), i)|^2 + \theta^T \Omega \theta.$$  (35)

We design an identifier satisfying the identifier requirement relative to (35) as follows. First, we let $\Theta := \mathbb{R}^{n_\theta}$ and $\Xi := \text{SPD}_{n_\theta} \times \mathbb{R}^{n_\nu}$. For a $\xi \in \Xi$ we consider the partition $\xi = (\xi_1, \xi_2)$ with $\xi_1 \in \text{SPD}_{n_\theta}$ and $\xi_2 \in \mathbb{R}^{n_\nu}$, and we equip $\Xi$
with the norm \( |\xi| := |\xi_1| + |\xi_2| \). We consider the following persistence of excitation condition, in which we let \( \text{msv}(\cdot) \) denote the minimum non-zero singular value.

**A4** There exists \( \epsilon > 0 \) and, for each solution \((\varsigma, w)\) of the core process (21), a \( j^* \in \mathbb{N} \), such that

\[
\text{msv} \left( \Omega + \sum_{i=0}^{j-1} \mu^{-i-1} \sigma(\tau(w(t^i))) \sigma(\tau(w(t^i)))^\top \right) \geq \epsilon.
\]

for all \( j \geq j^* \).

With \( \mathcal{H}^* \subset \mathbb{R}^{n_0} \) and \( \mathcal{U}^* \subset \mathbb{R}^{n_0} \) compact subsets such that \( \tau(W) \subset \mathcal{H}^* \) and \( u^*(W) \subset \mathcal{U}^* \), let

\[
\rho_1 := (1 - \mu)^{-1} \sup_{\eta \in \mathcal{H}^*} |\sigma(\eta)\sigma(\eta)^\top|, \\
\rho_2 := (1 - \mu)^{-1} \sup_{(\eta, u) \in \mathcal{H}^* \times \mathcal{U}^*} |\sigma(\eta)u|
\]

and, with \( ^\top \) denoting the Penrose-Moore pseudoinverse, define

\[
\Xi := \{ \xi \in \Xi : \text{msv}(\Omega + \xi_1) \geq \epsilon, |\xi_1| \leq \rho_1, |\xi_2| \leq \rho_2 \}.
\]

Then, we let \( \Sigma : \mathbb{R}^{n_0} \rightarrow \text{SPD}_{n_0} \), \( \lambda : \mathbb{R}^{n_0} \times \mathbb{R}^{n_0} \rightarrow \mathbb{R}^{n_0} \) and \( \vartheta : \Xi \rightarrow \mathbb{R}^{n_0} \) be any uniformly continuous functions satisfying

\[
\Sigma(\eta) = \sigma(\eta)\sigma(\eta)^\top, \\
\lambda(\eta, u) = \sigma(\eta)u, \\
\vartheta(\xi) = (\xi_1 + \Omega)\xi_2
\]

respectively on the compact sets \( \mathcal{H}^* \), \( \mathcal{H}^* \times \mathcal{U}^* \) and \( \Xi^* \), and

\[
|\Sigma(\eta)| \leq \rho_2, \quad |\lambda(\eta, u)| \leq \rho_3, \quad |\vartheta(\xi)| \leq c_0
\]

everywhere else, for some \( \rho_2, \rho_3, c_0 > 0 \). Then the identifier is described by the following equations

\[
\xi_1^+ = \mu \xi_1 + \Sigma(\alpha_{in}) \\
\xi_2^+ = \mu \xi_2 + \lambda(\alpha_{in}, \alpha_{out}) \\
\theta^+ = \vartheta(\xi)
\]

and the following result holds.

**Proposition 1:** Assume A4. Then, the identifier \( \theta^+ \) satisfies the identifier requirement relative to (35).

The proof of Proposition 1 can be deduced by the same arguments of A3 and it is thus omitted. It is worth observing that, whenever the regularization matrix \( \Omega \) is positive definite, A4 always holds with \( j = 0 \) and \( \epsilon \) the smallest eigenvalue of \( \Omega \). The importance of regularization is well understood in system identification (see e.g. [33]), although it is also well-known that regularization introduces a bias on the parameter estimation, in the sense that in the case where the “true map” \( \gamma \) relating \( \alpha_{in} \) and \( \alpha_{out} \) should belong to the model set \( \mathcal{M} \), the “true parameter” \( \theta^* \) is a minimum of (35) only if \( \theta^* \in \ker \Omega \), so that having \( \Omega \) nonsingular makes the identifier (37) converge “only” to a neighborhood of \( \theta^* \) whose size is related to the eigenvalues of \( \Omega \) (and thus can be made arbitrarily small). Therefore, the regularization matrix \( \Omega \) is a degree of freedom that must be chosen to weight well-conditioning of the problem and asymptotic estimation performances. If \( \Omega \) is chosen singular (possibly the zero matrix), the identifier

\[
\text{denotes the minimum non-zero singular value.}
\]

requirement is still satisfied along the trajectories of \( \omega \) that are sufficiently informative according to A4. In this respect, A4 is a property of the ideal input signal \( \alpha_{in} = \tau(\omega) \), but we underline that the sampling time of the core process (21) depends on the chosen strategy for the clock which makes the PE property also a property of the controller.

**B. “Mini-Batch” Algorithms for Nonlinear Parametrizations**

In this section we present a construction of the identifier fulfilling the requirement of Definition 1 when the model set \( \mathcal{M} \) assumes the generic form (23), with \( \Theta = \mathbb{R}^{n_0} \) for some \( n_0 \in \mathbb{N} \) and \( \gamma(\theta, \cdot) \) that is given and that is regular enough to satisfy the regularity item of the identifier requirement. We start by assuming to have available a batch identification algorithm working on a data set of finite size \( N \), and we define an identifier fitting in our framework that repeatedly executes the algorithm on “moving window” of size \( N \).

More precisely, with \( \Sigma_n \) the space of functions \( \{1, \ldots, N\} \rightarrow \mathbb{R}^{n_0} \), and given two signals \( s_{in} \in \Sigma_n \) and \( s_{out} \in \Sigma_n \), we define the window cost

\[
I_{(s_{in}, s_{out})}(\theta) := \sum_{i=1}^N \omega(s_{out}(i) - \gamma(\theta, s_{in}(i)), i) + \rho(\theta),
\]

where \( \omega : \mathbb{R}^{n_0} \times \{1, \ldots, N\} \rightarrow \mathbb{R}_+ \) and \( \rho : \mathbb{R}^{n_0} \rightarrow \mathbb{R}_+ \) some integral cost and regularization terms. Then we assume the following.

**A5** There exists a uniformly continuous map \( \mathcal{G}_n : \Sigma_n \times \Sigma_n \rightarrow \mathbb{R}^{n_0} \) such that, for any solution \((\varsigma, w)\) to the core process (21), and with

\[
s_{in}(i) := \tau(w(t^{i-N+i})), \quad s_{out}(i) := u^*(w(t^{i-N+i})),
\]

for all \( i = 1, \ldots, N \), it holds that

\[
\mathcal{G}_n(s_{in}, s_{out}) \in \arg\min_{\theta \in \mathbb{R}^{n_0}} I_{(s_{in}, s_{out})}(\theta).
\]

The map \( \mathcal{G} \) represents any optimization algorithm that extracts the optimal model of \( \mathcal{M} \) from the finite data set represented by the “windowed samples” of \( (\tau(w), u^*(w)) \).

With \( \lambda : \mathbb{R}^{n_0} \rightarrow \Sigma_n \) the linear operator mapping the vector \( v = (v_1, \ldots, v_N) \), \( v_i \in \mathbb{R}^{n_0} \), to the signal \( s \in \Sigma_n \), \( s(i) := v_i \), we construct an identifier starting from \( \mathcal{G} \) by letting \( \Theta := \mathbb{R}^{n_0} \), \( \Xi := \mathbb{R}^{n_0} \times \mathbb{R}^{n_0} \), and \( (\vartheta, \theta) \) such that the state \( \xi := (\xi_1, \xi_2) \), with \( \xi_1 \in \mathbb{R}^{n_0} \) and \( \xi_2 \in \mathbb{R}^{n_0} \), and output \( \theta \) of the identifier satisfy

\[
\begin{align*}
\xi_1^+ &= H_1 \xi_1 + B_1 \alpha_{in} \\
\xi_2^+ &= H_2 \xi_2 + B_2 \alpha_{out} \\
\theta^+ &= \vartheta(\xi),
\end{align*}
\]

where for \( i = 1, 2 \), \( (H_i, B_i) \) have the “shift” form

\[
\begin{align*}
H_i := \begin{pmatrix} 0_{Nm_i \times m_i} & I_{(N-1)m_i} \\ 0_{m_i \times (N-1)n_i} & I_{m_i} \end{pmatrix}, & B_i := \begin{pmatrix} 0_{(N-1)m_i \times m_i} \\ I_{m_i} \end{pmatrix}
\end{align*}
\]

being \( m_1 = n_g \) and \( m_2 = n_g \), and

\[
\vartheta(\xi) := \mathcal{G}(\lambda(\xi_1), \lambda(\xi_2)).
\]
The identifier (38) consists of a pair of “shift registers” propagating and accumulating the new values of \( \alpha_{\text{in}} \) and \( \alpha_{\text{out}} \) and forming in this way a moving window. The output map (39) assigns to the parameter \( \theta \) the current value given by the algorithm \( G \) corresponding to the current data set stored in the state \( \xi \). This construction has the following property (proved in Appendix A).

**Proposition 2:** Assume \( A^5 \) then the identifier (38)-(39) satisfies the identifier requirement relative to the cost functional

\[
J(x,w)(j)(\theta) := \sum_{i=\max\{0,j-N\}}^{j-1} \omega(\varepsilon(\theta,w(i)),i) + \rho(\theta).
\]

The identifier (38) consists of a pair of “shift registers” propagating and accumulating the new values of \( \alpha_{\text{in}} \) and \( \alpha_{\text{out}} \) and forming in this way a moving window.

VI. EXAMPLE

We consider the problem of synchronizing the output of a Van der Pol oscillator with unknown parameter, with a triangular wave with unknown frequency. The plant, which consists in a forced Van der Pol oscillator, is described by the following equations

\[
\begin{align*}
\dot{p}_1 &= p_2, \\
\dot{p}_2 &= -p_1 + a(1 - p_2^2)p_2 + u,
\end{align*}
\]

with \( a \) an unknown parameter known to range in \([a_1,a_2]\) for some constants \( a_1 > a_2 > 0 \). According to (32), a triangular wave can be generated by an exosystem of the form

\[
\begin{align*}
\dot{w}_1 &= w_2, \\
\dot{w}_2 &= -\varrho w_1
\end{align*}
\]

and in which \( \varrho \) is the unknown frequency, assumed to lie in the interval \([\varrho_1,\varrho_2]\) with \( \varrho_1 > \varrho_2 > 0 \) known constants. The control goal thus consists in driving the output \( p_1 \) of (40) to the reference trajectory \( p_1^*(w) \). With \( s(w) := (w_2,-\varrho w_1) \), we define the error system \( x \) as

\[
x := \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} p_1 - p_1^*(w) \\ p_2 - L_{\alpha}(w)p_1^*(w) \end{pmatrix},
\]

which is of the form (1) without \( z \) and with

\[
\begin{align*}
y &= x_1 = p_1 - p_1^*(w), \\
b(w,x) &:= 1 \\
q(w,x) &:= -x_1 - p_1^*(w) - L_{\alpha}(w)p_1^*(w) + 2(1 - (x_1 + p_1^*(w))^2)(x_2 + L_{\alpha}(w)p_1^*(w)).
\end{align*}
\]

The ideal steady-state error-zeroing control action that the regulator should provide is given by

\[
u^*(w) = -q(w,0)/b(w,0) = p_1^*(w) + L_{\alpha}(w)p_1^*(w) - a(1 - p_1^*(w)^2)L_{\alpha}(w)p_1^*(w),
\]

and no analytic technique is known to compute the right function \( \gamma \) of the internal model of (9), (12) for which the regulator is able to generate \( u^*(w) \). Furthermore, the indicated \( u^* \) does not fulfill any of the immersion conditions known in the literature.

Regarding the exosystem (41), we observe that the quantity

\[
V_\phi(w_1, w_2) := \frac{1}{2}\varrho w_1 + \frac{1}{2}w_2
\]

remains constant along each solution. Hence, the set

\[
W := \bigcup_{\varrho \in [\varrho_1,\varrho_2]} V_\phi^{-1}([0,c])
\]

is invariant for (41). Furthermore, assumptions \( A^1 \) and \( A^3 \) hold by construction, with \( b = 1 \) and any \( \mu \in (0,1) \), and hence, the problem fits into the framework of this paper, and the proposed regulator is used with:

(i) \( \kappa(x) = -Kx \), with \( K \in \mathbb{R}^{2 \times 2} \) such that \( \sigma(A - BK) = \{-1,-2\} \), and sat implements the standard saturation function with level \( M = 100 \);

(ii) \( n_\eta = 2(n_w + 1) = 6 \) and

\[
F := \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}, \quad G := \begin{pmatrix} 0 \\ 0 \end{pmatrix};
\]

(iii) the identifier is chosen as a least-squares identifier of the kind presented in Section V-A, in which the regressor \( \sigma \) is defined to perform a polynomial expansion of \( \gamma \) with a polynomials of odd order. More precisely, with \( N \in \mathbb{N} \), for \( n \leq N \) we let

\[
I_n := \{(i_1,\ldots,i_n) \in \{1,\ldots,6\}^n \mid i_1 \leq \cdots \leq i_n \}
\]

be the set of non-repeating multi-indices of length \( n \), and with \( I \in I_n \), we let

\[
\sigma_I(\eta) := \eta_1 \cdots \eta_n.
\]

The regressor \( \sigma \) is then defined as

\[
\sigma = \text{col}(\sigma_I \mid I \in I_n, n \leq N, n \text{ odd}).
\]

In the forthcoming simulations we have taken \( N = 1,3,5 \). To ensure that the persistence of excitation condition of \( A^2 \) holds, we have taken a diagonal regularization matrix \( \Omega = 10^{-3}I \). The forgetting factor is instead chosen as \( \mu = 0.99 \).

(iv) the extended observer is implemented with \( \ell = 20, h_1 = 6, h_2 = 11, h_3 = 6 \) and with \( \phi \) that is obtained by saturating the function \( (\theta,\eta,u) \mapsto (\partial_1^2(\theta,\eta)/\partial \eta)(F\eta + G\bar{u}) \) with a saturation level of 100.

(v) finally, a periodic clock strategy is employed, obtained by letting \( \mathbf{\Upsilon} = \mathbf{T} = 0.1 \).

The following simulation show the regulator applied with \( a = \varrho = 2 \) in four cases: (1) without internal model, i.e. with \( \phi = 0 \), as in (13); (2) with the adaptive internal model obtained by setting \( N = 1 \) (i.e., with \( \sigma(\eta) = \eta \)); (3) with \( N = 3 \), i.e. with \( \sigma(\eta) = \text{col}(\eta_1,\ldots,\eta_6,\eta_3,\eta_1\eta_2,\eta_1\eta_3\eta_2\eta_3,\ldots,\eta_3^5) \) and (4) with \( N = 5 \), i.e. with \( \sigma(\eta) = \text{col}(\eta_1,\ldots,\eta_6,\eta_3^2,\eta_5,\eta_3^5,\ldots,\eta_5^5) \).

In particular, Figure 2 shows the steady-state evolution of the
Figure 2. The first plot shows the steady-state time evolution of the tracking error $y = p_1 - p_1^*(w)$ in the four cases obtained without adaptive internal model and with $N = 1, 3, 5$. The second and third plots provide a zoom to highlight the difference between the adaptive internal models obtained by the polynomials of order $N=1, 3, 5$ of its approximation given by $200$ reduced by more than $15$ time compared to the case in which the adaptive internal model is described above. The error obtained by introducing a linear internal model. Finally, with $N=3$ bounded (we refer in particular to (16) and (28)).

tracking error $y(t) = p_1(t) - p_1^*(w(t))$ in the four cases described above. The error obtained by introducing a linear adaptive internal model ($N = 1$) is reduced by more than $15$ time compared to the case in which the adaptive internal model is not present (i.e. $\phi = 0$). Adding to the model set the polynomials of order $3$ ($N = 3$), reduces the maximum error of more than $120$ times compared to the first case without internal model. Finally, with $N = 5$, the maximum error is reduced by more than $200$ times. Figure 3 shows instead the time evolution of the ideal steady-state control law $u^*(w)$ and of its approximation given by $\hat{\gamma}(\theta(j), \eta(t))$ in the three cases in which $N = 1, 3, 5$.

VII. PROOF OF THEOREM

We subdivide the proof in three parts, coherently with the three claims of the theorem. For compactness, in the following we will write $p := (w, z, x)$ in place of $(w, z, x)$, and we let

$$f(p, u) := \text{col}(s(w), f(p), Ax + B_q(p) + b(p)u).$$

For compactness, in the following we also use the symbol $\star$ in place of the arguments of functions that are uniformly bounded (we refer in particular to (16) and (28)).

A. Stability analysis

We first remark that since $T > 0$ and $T < +\infty$, then all the complete trajectories of (29) have an infinite number of jump and flow times. We then notice that $\eta$ is a Hurwitz linear system driven by a bounded input $u$. Hence its solutions are complete and bounded. Moreover, we can write

$$\eta = \tau(w) + d_{in}, \quad u = u^*(w) + d_{out}$$

with $u^*$ given by (17) and

$$d_{in} := \eta - \tau(w), \quad d_{out} := u - u^*(w)$$

that are bounded. Hence, the identifier requirement implies that also the identifier has complete and bounded solutions. We thus focus on the subsystem $(w, z, x, \dot{x}, \dot{\sigma})$. Let

$$\sigma := q(p) + (b(p) - b)u(x, \dot{x}, \dot{\sigma})$$

and change coordinates according to

$$(\dot{x}, \dot{\sigma}) \mapsto e := \text{col}(e_x, e_{\sigma}).$$

In view of (12), $u = u(\dot{x}, \dot{\sigma})$. Hence the dynamics of $x$ can be rewritten as

$$\dot{x} = Ax + B(b \sigma + bu + (b(p) - b)u(x, \dot{x}, \dot{\sigma}))$$

Following the standard high-gain paradigm, define

$$e_x := \ell \Lambda(\ell)^{-1}(\dot{x} - x), \quad e_{\sigma} := \ell^{-r} (\dot{\sigma} - \sigma)$$

and change coordinates according to

$$(\dot{x}, \dot{\sigma}) \mapsto e := \text{col}(e_x, e_{\sigma}).$$

In the new coordinates, (12) reads as

$$u = b^{-1} \text{sat}(-\ell' e_{\sigma} - \sigma + \kappa(\Lambda(\ell)\ell^{-1} e_x + x)),$$

and (43) gives rise to the implicit equation

$$T(\sigma, e_{\sigma}, \sigma) = 0,$$

where $T(\sigma, e_{\sigma}, \sigma) := \sigma - q(p) - (b(p) - b)\text{sat}(-\ell' e_{\sigma} - \sigma + \kappa(x))$. We observe that

$$\frac{\partial T}{\partial \sigma}(\sigma, e_{\sigma}, \sigma) = I + (b(p) - b)b^{-1} \text{sat}'(\sigma).$$

(13) and (16) give $|(b(p) - b)b^{-1} \text{sat}'(\sigma)| \leq 1 - \mu$, so that $\partial T(\sigma, e_{\sigma}, \sigma)$ is uniformly nonsingular. This, in turn, suffices to show that there exists a unique $C^1$ function $\phi_{\sigma}(p, e_{\sigma})$ satisfying $T_{\sigma}(p, e_{\sigma}, \phi_{\sigma}(p, e_{\sigma})) = 0$, and such that $\sigma$ can be written as

$$\sigma = \phi_{\sigma}(p, e_{\sigma}).$$

We further notice that, $T_{\sigma}(p, e_{\sigma}, \sigma) = 0$ also implies

$$\frac{\partial T_{\sigma}}{\partial p}(p, e_{\sigma}, \sigma)f(p, u) + \frac{\partial T_{\sigma}}{\partial e_{\sigma}}(p, e_{\sigma}, \sigma)e_{\sigma} + \frac{\partial T_{\sigma}}{\partial \sigma}(p, e_{\sigma}, \sigma)\dot{\sigma} = 0,$$

which in turn yields

$$\dot{\sigma} = \Delta_1 + \Delta_2 \ell' \dot{e}_{\sigma},$$

where, with $m(p, \star) := (b(p) - b)b^{-1} \text{sat}'(\star)$, we let

$$\Delta_1 = -(I + m(p, \star))^{-1} \frac{\partial T_{\sigma}}{\partial p}(p, e_{\sigma}, \sigma)f(p, u),$$

$$\Delta_2 = -(I + m(p, \star))^{-1} m(p, \star).$$
Due to $A^3$ and (10), $\|p(m,\cdot)\| \leq 1 - \mu$, so that $I + m(p,\cdot)$ is always invertible. Moreover, in view of Proposition 10.3.2], $\|(I + m(p,\cdot))^{-1}\| \leq \mu^{-1}$, so that we obtain

$$\Delta_2 \leq \mu^{-1} - 1 \quad (48)$$

for all $(p, e_p, \sigma) \in W \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \times \mathbb{R}^{n_y}$. The variable $e$ jumps according to $e^+ = e$ and flows according to

$$\dot{e}_x = \ell(A - HC) e_x + B(e, \sigma + \Delta_1, t)$$
$$\dot{e}_\sigma = -\ell H_{r+1} C e_x - \Delta_2 \dot{e}_\sigma + \ell^{-1} \Delta_4$$

where

$$\Delta_{3,\ell} := -\ell^{-1} \lambda (b(p) - b) (u(\hat{x}, \hat{\sigma}) - u(x, \sigma))$$
$$\Delta_4 := -b(\psi(*) - \Delta_1).$$

Since $I + \Delta_2 = I - (I + m(p,\cdot))^{-1} m(p,\cdot) = (I + m(p,\cdot))^{-1}$ and $\|(m(p,\cdot))^{-1}\| \leq 1 - \mu < 1$, then $I + \Delta_2$ is uniformly invertible, and solving (50) for $\dot{e}_\sigma$ yields

$$\dot{e}_\sigma = -\ell H_{r+1} C e_x - m(p,\cdot) \ell H_{r+1} C e_x + \ell^{-1} \Delta_4.$$ 

with $\Delta_5 := (I + m(p,\cdot)) \Delta_4$. Hence, by letting $e = \text{col}(e_x, e_\sigma)$ and

$$\overline{A} := \begin{pmatrix} A - HC & B \\ -H_{r+1} C & 0 \end{pmatrix}, \quad \overline{B}_x := \begin{pmatrix} B \\ 0_n \end{pmatrix}, \quad \overline{B}_\sigma := \begin{pmatrix} 0_n \otimes n_y \\ 1_n \end{pmatrix}$$

we obtain

$$\dot{e} = \ell \overline{A} e + \overline{B}_x \Delta_{3,\ell} + \overline{B}_\sigma \big(- m(p,\cdot) \ell H_{r+1} C e_x + \ell^{-1} \Delta_5\big).$$

Let $\Omega_0$ be the compact set introduced in Section III-A fulfilling (14). In view of $A^1$ and (10), there exists $a_0 > 0$ such that $|\Delta_4| \leq a_0$ holds for all $p \in \Omega_0$ and all $(e_x, \sigma) \in (\mathbb{R}^{n_y})^2$. In view of (28) and (43), $|\Delta_5| \leq a_2$ with $a_2 := |b| \psi_0 + a_0$. Moreover, since $\kappa$ and sat are $C^1$, and $u$ is bounded, then it is $C^1$ and Lipschitz. Thus, (28) and (44) imply the existence of $a_1 > 0$ such that $|\Delta_{3,\ell}| \leq a_1 |e_x|$ for all $\ell \geq 1$ and all $(p, e_x, \sigma)$. The stability properties of $e$ then follow by the Lemma below, proved in Appendix III-A.

**Lemma 2:** Consider a system of the form

$$\dot{\chi} = \ell \overline{A} - \ell \overline{B}_x \Delta_{3,\ell} + \overline{B}_\sigma (e, \sigma) \delta_1 + \delta_2 \quad (51)$$

With $\chi = \text{col}(\chi_1, \chi_2), \chi_1 \in \mathbb{R}^{n_x}, \chi_2 \in \mathbb{R}^{n_y}, \delta_1 : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_y}$ and $\delta_2 : \mathbb{R}^{n_y} \rightarrow \mathbb{R}^{n_y} + n_x$, locally integrable inputs such that, for some $a_3 > 0$, $|\delta_1| \leq a_3 |\chi_1|$, and $a : \mathbb{R}^{n_y} \rightarrow \mathbb{R}$ a continuous function satisfying $|a(t)| \leq \alpha < 1$ for all $t \in \mathbb{R}_+$. Then there exist $\ell_0, a_4, a_5, a_6 > 0$ such that, if $\ell \geq \ell_0$, it holds that

$$|\chi(t)| \leq \max \left\{ a_4 e^{-a_4 t} |\chi(0)|, a_6 e^{-a_6 |\delta_2| t} \right\}$$

for all $t \in \mathbb{R}_+$, for which the solution is defined.

In particular, since $|m(p,\cdot)| \leq 1 - \mu < 1$, Lemma 2 yields the existence of $\ell_0, a_4, a_5, a_6 > 0$ such that, if $\ell \geq \ell_0$, then

$$|e(t)| \leq \max \left\{ a_4 e^{-a_4 t} |e(0)|, a_6 e^{-\ell (r+1)} |\Delta_5| t \right\} \quad (52)$$

Moreover, the following Lemma holds.

**Lemma 3:** Suppose that, for some $T_0, a_0 > 0$, $|\Delta_5| \leq a_0$ for all $t \in [0, T_0)$. Then, for each $T \in (0, T_0)$ and each $\epsilon > 0$, there exists $\ell_1(T, \epsilon) > \ell_0$ such that, if $\ell \geq \ell_1(T, \epsilon)$, then for each solution of (29) originating in $X_0$, it holds that

$$\max \{|x(t) - \hat{x}(t)|, |\sigma(t) - \hat{\sigma}(t)|\} \leq \epsilon$$

for all $t \in [T, T_0)$.

Let $b_1 := \sup_{(p, x) \in \Omega_0 \times \mathbb{R}^{n_y}} |Ax + B(q(p) + b(p)b^{-1} \text{sat}(s))|$. Then, as long as $p \in \Omega_0$, we have $|\hat{x}| \leq b_1$. Thus, for all $t \in \mathbb{R}_+$, we have $|x(t)| \leq |x(0)| + b_1 t$.

**Lemma 4:** The unique solution (47) of (46) satisfies

$$\phi_\sigma(p, e_\sigma) := bb(p) - (1 - bb(p)) (r(x) - \ell e_\sigma)$$

for all $(p, e_\sigma) \in \Omega_0 \times \mathbb{R}^{n_y}$ such that $|e_\sigma| \leq \mu_2 q_2$ and $\mu_2$ the constants given respectively in $A^3$ and (14). Pick $r > 0$ and let

$$X_r = \{ x \in \mathbb{R}^{n_x} : |x - \hat{x}| \leq r, (w, z, x) \in \Omega_0 \}.$$

Then $X_0$ is compact and, by continuity of $\kappa$, there exists $\rho_1 \in K$ such that $|\kappa(\hat{x} - x)| \leq \rho_1 (|\hat{x} - x|)$ for all $(\hat{x}, x) \in (X^-)^2$.

With $q_2$ defined in (15), let $\ell_1^*$ be taken equal to the $\ell_1^*(T, \epsilon)$ produced by Lemma 3 for $T \in (0, \hat{t})$ and

$$\epsilon = \min \{ r, \mu_2/2 \} - \rho_1/2 \delta, \delta \in \{ 0, 2 \} \quad (54)$$

and pick $\ell > \ell_1^*$. Since $p(t) \in \Omega_0$ in $[\hat{t}, t]$, then Lemma 3 implies $|e^\epsilon_\sigma| \leq \mu_2 q_2$, so that from Lemma 4, $\sigma(t) = \phi_\sigma(p(t), e(t))$ given by (53) for all $t \in [T, \hat{t}]$. Furthermore, for all $t \in [T, \hat{t}], \hat{x}(t) \in X_r$, and the argument of sat(\cdot in (45) satisfies

$$\min \{|\sigma - \ell e_\sigma + \kappa(\Delta(\ell) \ell e_\sigma + x)| \leq \max \{|bc(p) + bb(p) - b(p)\kappa(x) + |b(p)\ell e_\sigma| + |\kappa(\Delta(\ell) \ell e_\sigma + x) - \kappa(x)|| \leq \rho(x) \} \quad (54)$$

in which we let $e(p) = -b(p)^{-1} q(p)$ and we used (54) and the fact that $|b(p)\ell_1^*| \leq \mu_2 - \mu_1$ everywhere (see Remark 2.4). Hence, for all $t \in [\hat{t}, t]$, the control $u = u(\hat{x}, \hat{\sigma})$ is out of the saturation, and similar arguments show that $u(x, \sigma)$ is too.

Thus, for all $t \in [T, \hat{t}], (54) \text{ and } (55) \text{ yield}

$$\dot{\delta} = -\ell_1^* e_\sigma + b(p)b^{-1} (\kappa(\Delta(\ell) \ell e_\sigma + x) - \kappa(x))$$

that, in view of (54), satisfies $|\delta(t)| \leq \delta \text{ in } [T, \hat{t}]$. Thus, since $p(T) \in \mathbb{E}_r$, we conclude by definition of $\Omega_0$ in Section III-A.
that all the trajectories of the closed-loop system (29) originating in $X_{\theta}$ are complete, eventually equibounded, and satisfy $(w(t), z(t), x(t)) \in \Omega_0$ for all $t \in \mathbb{R}_+$, which is the first claim.

\[ K_{3,\ell} := \overline{\mu}(p)\kappa'(x)(\kappa(x) + \delta) \]
with $\delta$ defined in (56). Thus, $\zeta_{\sigma}$ jumps according to

\[ \zeta_{\sigma}^+ = \zeta + \ell^{-r}b(p)(\varepsilon^* - \varepsilon^{\ast +}) \]
where we let $\varepsilon^{\ast +} := \varepsilon(\theta, \nu, u) - \varepsilon(\theta^*, \tau(w), u^*(w))$, and for all $(t, j) \geq (T, j^*)$, and it follows according to

\[ \dot{\zeta}_{\sigma} = \ell H_{r+1}C\zeta_{\sigma} + \overline{\mu}(p)\dot{\zeta}_{\sigma} + \ell^{-r}K \]

where

\[ K := -K_{1} - K_{2,\ell} - K_{3,\ell}b\bigl(\psi(\theta, \eta, u) - \psi(\theta^*, \tau(w), u^*(w))\bigr). \]

Notice that $I - \overline{\mu}(p) = bb^{-1}$ is uniformly invertible and bounded (see Remark 4). Then, solving (60) for $\dot{\zeta}_{\sigma}$ yields

\[ \dot{\zeta}_{\sigma} = \ell H_{r+1}C\zeta_{\sigma} + (b(p) - b)b^{-1}\ell H_{r+1}C\zeta_{\sigma} + \ell^{-r}b(p)b^{-1}K. \]

Hence, $\zeta$ satisfies

\[ \zeta^+ = \zeta + \mathcal{B}_x\ell^{-r}b(p)(\varepsilon^* - \varepsilon^{\ast +}) \]
during jumps and, in view of (57) and (61),

\[ \dot{\zeta} = \ell A \zeta + \mathcal{B}_x\bigl(-\alpha\ell H_{r+1}C\zeta_{\sigma} + \ell^{-r}b(p)b^{-1}K\bigr) + \mathcal{B}_x(\ell^{r-\tau}(p)\varepsilon^* + \Delta_{3,\ell}) \]
during flows, with $\alpha := -(b(p) - b)b^{-1}$ that due to A5 satisfies $|\alpha| \leq 1 - \mu$ everywhere. In view of A1 and since $p \in \Omega_0$ for all $t \in \mathbb{R}_+$, there exist $\nu_1, \nu_2, \nu_3 > 0$ such that

\[ \begin{align*}
|b(p)b^{-1}K_{1}| & \leq \nu_1(|p|_B + |u - u^*(w)|), \\
|b(p)b^{-1}K_{2,\ell}| & \leq \nu_2(|p|_B + \ell'|\zeta|), \\
|b(p)b^{-1}K_{3,\ell}| & \leq \nu_3(|p|_B + \ell'|\zeta| + |\varepsilon^*|)
\end{align*} \]
for $t \geq T$. Regarding the term $\psi(\theta, \eta, u) - \psi(\theta^*, \tau(w), u^*(w))$, we notice that for all $j \geq j^*$, $\theta^*(j) \in \Theta^* = \theta(\Xi^*)$, while $(\tau(w), u^*(w)) \in \mathcal{H}^* \times \mathcal{U}^*$ holds everywhere by construction. Thus, at each $(t, j) \geq (T, j^*)$, since in view of the identifier requirement $\partial^2/\partial \eta_1$ is locally Lipschitz and $|\theta - \theta^*| \leq \rho_0(\|\xi - \xi^*\|)$ with $\rho_0$ locally Lipschitz for all $(\theta, \theta^*, \xi, \xi^*) \in \Theta \times \Theta^* \times \Xi \times \Xi^*$, and since $\psi$ is globally bounded, there exists $\nu_4 > 0$ such that

\[ \begin{align*}
[\psi(\theta, \eta, u) - \psi(\theta^*, \tau(w), u^*(w))] & \\
& \leq \nu_4(|\xi - \xi^*| + |\eta - \tau(w)| + |u - u^*(w)|)
\end{align*} \]
We then observe that, in view of (45) and (58), and since for $t \geq T$ the control is out of saturation that

\[ |u - u^*(w)| \leq \nu_5(|p|_B + |\varepsilon^*| + \ell'|\zeta|) \]
for some $\nu_5 > 0$. Therefore, the last term of (61) satisfies

\[ |\ell^{-r}b(p)b^{-1}K| \leq \nu_6(\xi + \ell^{-r}\nu_7((|p|_B)e^* + |\xi - \xi^*| + |\varepsilon^*|)) \]
for all $(t, j) \geq (T, j^*)$, with $\nu_6 := \max\{\nu_2 + \nu_3, (|v_1 + \nu_4|)\}$ and $\nu_7 := \max\{\nu_1 + \nu_2 + \nu_3, (|v_1 + \nu_4|)\}$ and with $E$ given by (20). Hence, Lemma 3 and (59) yield the existence of
constants $\nu_8, \nu_9, \nu_{10}, \nu_{11} > 0$ such that the following bounds hold
\[
|\zeta(t, j)| \leq \max \left\{ \nu_8 e^{-\ell \omega_0 (t-j)}|\zeta(t_j, j)|, \nu_9 \ell^{-r} |\xi^*|_{(t, j)} \right\}, \\
|\zeta(t^j, j)| \leq \nu_{10} \ell^{-(r+1)} |\phi(\eta, j)|, \nu_{10} \ell^{-(r+1)} |\xi - \xi^*|_{j}
\]
for all $(t, j) \geq (T, j^*)$. We then observe that, since for each $(t, j) \in \text{dom } x$, $t^j - j \geq T$, then for each constant $\bar{\nu} > 0$, sufficiently large values of $\ell$ yield
\[
\lim_{t^j \to \infty} \bar{\nu}^{j+1} e^{-\ell \omega_0 t} \leq \lim_{t^j \to \infty} \bar{\nu}^{j+1} e^{-\ell \omega_0 t} = 0.
\]
Thus, there follows from (65) by induction and standard ISS arguments that there exist $\nu_{12} > 0$ and $\ell^*_s(\Gamma) > \ell_s$ such that $\ell > \ell^*_s(\Gamma)$ implies
\[
\lim_{t^j \to \infty} \nu_{12} \max \left\{ \ell^{-(r+1)} \limsup_{j \to \infty} |\phi(\eta, j)|, \ell^{-(r+1)} \limsup_{j \to \infty} |\xi - \xi^*|_{j} \right\},
\]
in which we used the fact that $\limsup_{t^j \to \infty} |\xi^*(t, j)| = \limsup_{t^j \to \infty} |\xi^*(t, j)|$. We now observe that (62) implies that the flow equation of $(w, z, \eta)$ can be written as (19), with $\delta_1 + \delta_2 = u - u^*(w)$ such that $|\delta_1| \leq \nu_5 |p|_B$ and $|\delta_2| \leq \nu_5 |\xi^* + \ell^* t|$. Moreover, substituting (45), (58) into the equation of $x$ yields (13) for all $t \geq T$, with $\delta$ defined in (56) that, for some $\nu_{13} > 0$, fulfills $|\delta| \leq \nu_{13} (|\xi^* + \ell^* t|)$. Hence, Lemma 1 and the ISS property of (13) yield the existence of $\nu_{13} > 0$ such that
\[
\lim_{t^j \to \infty} |p(t, \eta(t))| \leq \nu_{13} \max \left\{ \limsup_{t^j \to \infty} |\xi^*(t, j)|, \ell^* \limsup_{t^j \to \infty} |\zeta(t)| \right\},
\]
where we recall that $E$ is given by (20). On the other hand, the identifier requirement, the expression (42) and the bounds above yield the existence of a $\nu_{14} > 0$ such that
\[
\lim_{j \to \infty} |\xi(j) - \xi^*(j)| \leq \nu_{14} \max \left\{ \limsup_{t^j \to \infty} |\xi^*(t, j)|, \limsup_{t^j \to \infty} |p(t, \eta(t))|, \ell^* \limsup_{t^j \to \infty} |\zeta(t)| \right\}
\]
Denote for convenience $|(p, \eta, \xi)|_{\Gamma} := \max \{|(p, \eta, \xi)|_{\Gamma}, |\xi - \xi^*| \}$. Then, with $\nu_{15} := \max \{\nu_{13}, \nu_{14}, \nu_{13} \nu_{14}\}$, (65) and (66) yield
\[
\lim_{t^j \to \infty} |p(t, \eta(t), \xi(j))| \leq \nu_{15} \max \left\{ \ell^* \limsup_{t^j \to \infty} |\zeta(t)|, \limsup_{t^j \to \infty} |\xi^*(t, j)| \right\}.
\]
With reference to the set $\mathcal{O}$ and the function $|\cdot|_{\mathcal{O}}$ defined respectively in (30) and (31), we observe that $|x|_{\mathcal{O}} \leq \nu_{16} \max \{|(p, \eta, \xi)|_{\Gamma}, \ell^* |\xi| \}$ for some $\nu_{16} > 0$. Therefore, substituting (64) into (67) and using $|\xi| \leq |x|_{\mathcal{O}}$, yield
\[
\limsup_{t^j \to \infty} |x(t, j)|_{\mathcal{O}} \leq \nu_{17} \max \left\{ \ell^{-1} \limsup_{t^j \to \infty} |x(t, j)|_{\mathcal{O}}, \limsup_{t^j \to \infty} |\xi^*(t, j)| \right\},
\]
with $\nu_{17} := \nu_{16} \max \{\nu_{12}, \nu_{15}, \nu_{12} \nu_{15}\}$, and the claim follows with $\alpha_{\xi} = \nu_{17}$ by taking $\ell^*_s(\Gamma) = \max \{\ell^*_s(\Gamma), \nu_{17}\}$. ■

VIII. CONCLUSION

In this paper we proposed a regulator design for a class of multivariable nonlinear systems which employs an adaptive internal model unit and an extended high-gain observer to solve instances of practical, approximate and asymptotic output regulation problems. The proposed design employs system identification algorithms to carry out the estimation of an optimal internal model, and contrary to the majority of the existing designs does not need high-gain stabilization techniques. Future research directions will be aimed at exploiting the additional freedom on the stabilizer to deal with non minimum-phase systems, and at investigating further identification algorithms that fits in the framework, thus developing further the bridge with the system identification literature. We also aim to study the robustness of the proposed scheme in a formal framework of (19), by connecting the identifier’s validation to classical robustness concepts.

APPENDIX

A. Proof of Proposition 2

Consider the interconnection (26) with $(\Xi, \varphi, \Theta, \theta)$ given in Section V.B and $(a_{in}, a_{out}) = (a_{in}(w) + d_{in}, a_{out}(w) + d_{out})$. Define
\[
\xi^*_s(j) = \left( \frac{\alpha_{in}(w(j - N))}{\alpha_{out}(w(j - 1))} \right) \cdots \left( \frac{\alpha_{in}(w(j - N))}{\alpha_{out}(w(j - 1))} \right), \quad \xi^*_s(j) = \left( \frac{\alpha_{in}(w(j - N))}{\alpha_{out}(w(j - 1))} \right) \cdots \left( \frac{\alpha_{in}(w(j - N))}{\alpha_{out}(w(j - 1))} \right)
\]
and let $\theta^*(j) := G(\lambda(\xi^*_s(j)), \lambda(\xi^*_s(j)))$. Then for each solution pair $(\xi_s, \varphi, \Theta, \theta)$ to (26), $(\xi_s, \varphi, \Theta, \theta)$ is a solution pair to (26). We then observe that $J(\xi_s(\varphi))(\theta) = \mathcal{T}^N(\sin(\theta))(\theta)$ with $s_{in}(\theta) := \lambda(\xi^*_s(j))$ and $s_{out}(\theta) := \lambda(\xi^*_s(j))$. Hence the optimality and regularity items of the identifier requirements and the bound on $|\theta - \theta^*|$ in the stability item follow by (15). Finally, the rest of the stability item follows by the fact that the system $\xi := \xi - \xi^*$ is a linear system with input $(d_{in}, d_{out})$ and state matrix $H := \text{diag}(H_1, H_2)$ having all zero eigenvalues, and hence it is ISS relative to the origin and with respect to the input $(d_{in}, d_{out})$. ■

B. Proof of Lemma 2

The proof follows by the same arguments of (18). In particular, we first consider the system
\[
\dot{\chi} = \ell A\chi + \ell B_s H_{r+1} \delta_0 + B_{e} \delta_1 + \delta_2
\]
with $|\delta_1| \leq a_3 |\chi|$. Since $\mathcal{T}$ is Hurwitz there exists $P = P^T > 0$ fulfilling $\mathcal{A}^* P + P^* \mathcal{A} = -I$ and such that the Lyapunov
candidate $V(χ) := \sqrt{χ}Pχ$ satisfies $Δχ ≤ V(χ) ≤ \tilde{λ}χ$, with $\tilde{λ}$ and $\lambda$ respectively the smallest and largest eigenvalues of $P$. Then, there exist $b_1, b_2, b_3 > 0$ such that, for all $χ ≠ 0$,

$$V(χ) ≤ −b_1ℓV(χ) + b_2(ℓ|δ₀| + |δ₁| + |δ₂|)$$

$$≤ −b_1ℓ^2V(χ) + b_2(ℓ|δ₀| + |δ₂|)$$

Let $ℓ_0^* := 2b_3/b_1$. Then $b_3 − b_1ℓ ≤ −b_1ℓ$ for all $ℓ ≥ ℓ_0^*$, and this shows that $χ$ is ISS relative to the origin and with respect to the inputs $δ₀$ and $δ₂$. Moreover, the asymptotic gain between $δ₂$ and $|χ|$ is of the form $b_4/ℓ$, for some $b_4$ independent on $ℓ$.

Regarding the asymptotic gain between $δ₀$ and $|χ|$, we observe that (68) is a linear system, and there follows from the structure of $\bar{A}, \bar{B}$ and $C$, that when $δ₁ = 0$ and $δ₂ = 0,$ $C_χ^{(r+1)}H_1C_χ^{(r)} + \cdots + H_rC_χ + H_{r+1}C_χ = H_{r+1}δ₀$.

Thus, the transfer function from $δ₀$ to $C_χ$ has the form

$$(s^{r+1}I + H_1s^r + \cdots + H_r s + H_{r+1})^{-1}H_{r+1}.$$ 

Since all the poles are real and negative by construction, and since each $H_r$ is diagonal, it follows that the gain between $δ₀$ and $|C_χ|$ is unitary. Finally the proof follows by standard small-gain arguments by observing that system (51) is obtained as the interconnection of (68) and the algebraic system $δ₀ = −αC_χ$ and that, since $|α| ≤ \bar{α} < 1$ the overall gain is less than one. □

C. Proof of Lemma 3

In view of (44), if $ℓ ≥ ℓ_0^*$ (52) leads to the existence of $α_7, α_8 > 0$ such that

$$\max\{|x(t) − \hat{x}(t)|, |σ(t) − \hat{σ}(t)|\} ≤ \max\{ℓα_7e^{−α_7ℓt}, α_8e^{−ℓ}|\Delta δ₀|\}.$$ 

As $(x(0), σ(0), \hat{x}(0), \hat{σ}(0))$ lives in a compact set, there exists $b > 0$ such that $\max\{|x(t) − \hat{x}(t)|, |σ(t) − \hat{σ}(t)|\} ≤ \max\{ℓe^{−ℓb}\exp(−ℓα_5), α_8α_0e^{−ℓ}\}$ for all $t ∈ (0, T₀)$. Pick $T ∈ (0, T₀)$ and $ε > 0$ arbitrarily, and let

$$\bar{ℓ}(ℓ, ε) := \frac{T}{α_5} \log \left(\sqrt{b} / ε\right).$$

Then $\lim_{T → ∞} \bar{ℓ}(ℓ, ε) = 0$, so that there exists $\bar{ℓ}(ℓ, ε, T) > 0$ such that, for all $ℓ ≥ \tilde{ℓ}(ℓ, ε, T), \bar{ℓ}(ℓ, ε) ≤ T$ and $t ≥ \bar{ℓ}(ℓ, ε)$ yields

$$bℓε \exp(−α_5t) ≤ bℓε \exp(−\log(ℓεb)/ε) = ε.$$ 

Hence the claim holds with $ℓ_1^*(T, ε) := \max\{ℓ_0^*, \bar{ℓ}(ℓ, ε, T), α_8α_0/ε\}$. □

D. Proof of Lemma 2

We want to show that $φ_σ(p, ε) = φ$ with $φ$ the quantity defined by $φ := bb(p)^{-1}q(p) + (I − bb(p)^{-1})(κ(x) − ℓεεσ).$ First, notice that the quantity $s := −φ − ℓεεσ + κ(x) − ℓεεσ$ satisfies

$$s = −bb(p)^{-1}q(p) + bb(p)^{-1}(κ(x) − ℓεεσ).$$

In view of Remark 2 $|bb(p)^{-1}| ≤ μ^{-1}$ for all $p$. If $p ∈ Ω₀$ and $|ℓεεσ| ≤ μq₁$, in turn, we get

$$|s| ≤ \max\{|bc(p) + bb(p)^{-1}κ(x)| + μ^{-1}|ℓεεσ|$$

$$≤ \max\{|bc(p) + bb(p)^{-1}κ(x)| + q₁ ≤ M.$$ 

Hence, sat$(s) = s$, and it is easy to see that $T_+(p, ε, φ) = 0$. By uniqueness of solutions of (45), we conclude $φ = φ_σ(p, ε)$ which is the claim. □

REFERENCES

[1] B. A. Francis and W. M. Wonham, “The internal model principle of control theory,” *Automatica*, vol. 12, pp. 457–465, 1976.

[2] E. J. Davison, “The robust control of a servomechanism problem for linear time-invariant multivariable systems,” *IEEE Trans. Autom. Contr.*, vol. AC-21, no. 1, pp. 25–34, 1976.

[3] J. Huang and W. J. Rugh, “On a nonlinear multivariable servomechanism problem,” *Automatica*, vol. 26, no. 6, pp. 963–972, 1990.

[4] A. Isidori and C. I. Byrnes, “Output regulation of nonlinear systems,” *IEEE Trans. Autom. Contr.*, vol. 35, no. 2, pp. 131–140, 1990.

[5] J. Huang, “Asymptotic tracking and disturbance rejection in uncertain nonlinear systems,” *IEEE Trans. Autom. Contr.*, vol. 40, no. 6, pp. 1118–1122, 1995.

[6] C. Byrnes, F. Delli Priscoli, and A. Isidori, “Structurally stable output regulation for nonlinear systems,” *Automatica*, vol. 33, no. 3, pp. 369–385, 1997.

[7] C. I. Byrnes and A. Isidori, “Limit sets, zero dynamics and internal models in the problem of nonlinear output regulation,” *IEEE Trans. Autom. Contr.*, vol. 48, pp. 1712–1723, Oct. 2003.

[8] ——, “Nonlinear internal models for output regulation,” *IEEE Trans. Autom. Contr.*, vol. 49, pp. 2244–2247, Dec. 2004.

[9] L. Marconi, L. Praly, and A. Isidori, “Output stabilization via nonlinear Luenberger observers,” *SIAM J. Contr. Optim.*, vol. 45, pp. 2277–2298, 2007.

[10] C. I. Byrnes, A. Isidori, and A. Praly, “On the asymptotic properties of a system arising from the non-equilibrium theory of output regulation,” Mittag Leffler Institute, 2003.

[11] L. Wang, A. Isidori, H. Su, and L. Marconi, “Nonlinear output regulation for invertible nonlinear MIMO systems,” *Int. J. Robust Nonlinear Control*, vol. 26, pp. 2401–2417, 2016.

[12] L. Wang, A. Isidori, Z. Liu, and H. Su, “Robust output regulation for invertible nonlinear MIMO systems,” *Automatica*, vol. 82, pp. 278–286, 2017.

[13] M. Bin and L. Marconi, “Output regulation by post-processing internal models for a class of multivariable nonlinear systems.” Under review, 2019.

[14] M. Bin, D. Astolfi, L. Marconi, and L. Praly, “About robustness of internal model-based control for linear and nonlinear systems,” in *57th IEEE Conference on Decision and Control*, 2018.

[15] L. Marconi and L. Praly, “Uniform practical nonlinear output regulation,” *IEEE Trans. Autom. Contr.*, vol. 53, pp. 1184–1202, 2008.

[16] D. Astolfi, L. Praly, and L. Marconi, “Approximate regulation for nonlinear systems in presence of periodic disturbances,” in *54th IEEE Conference on Decision and Control*, 2015, pp. 7665–7670.

[17] A. Isidori, L. Marconi, and L. Praly, “Robust design of nonlinear internal models without adaptation,” *Automatica*, vol. 48, pp. 2049–2216, 2012.

[18] L. B. Freidovich and H. K. Khalil, “Preformance recovery of feedback-linearization-based designs,” *IEEE Trans. Autom. Contr.*, vol. 53, no. 10, pp. 2324–2334, 2008.

[19] A. Serrani, A. Isidori, and L. Marconi, “Semiglobal nonlinear output regulation with adaptive internal model,” *IEEE Trans. Autom. Contr.*, vol. 46, no. 8, pp. 1178–1194, 2001.

[20] F. D. Priscoli, L. Marconi, and A. Isidori, “A new approach to adaptive nonlinear regulation,” *SIAM J. Control Optim.*, vol. 45, no. 3, pp. 829–855, 2006.

[21] A. Pyrkin and A. Isidori, “Output regulation for robustly minimum-phase multivariable nonlinear systems,” in *IEEE 56th Conference on Decision and Control*, 2017, pp. 873–878.

[22] M. Bin, L. Marconi, and A. R. Teel, “Adaptive output regulation for linear systems via discrete-time identifiers,” *Automatica*, 2018, to appear.

[23] F. Forte, L. Marconi, and A. R. Teel, “Robust nonlinear regulation: Continuous-time internal models and hybrid identifiers,” *IEEE Trans. Autom. Contr.*, vol. 62, no. 7, pp. 3136–3151, 2017.
[24] M. Bin and L. Marconi, “class-type” identification-based internal models in multivariable nonlinear output regulation,” Under review, 2019.
[25] M. Bin, P. Bernard, and L. Marconi, “Adaptive output regulation via nonlinear luenberger observers,” Accepted to IFAC Symposium on Nonlinear Control Systems, 2019.
[26] R. Goebel, R. G. Sanfelice, and A. R. Teel, Hybrid Dynamical Systems. Princeton, N. J.: Princeton University Press, 2012.
[27] L. Ljung, System Identification: Theory for the user. Prentice Hall, 1999.
[28] H. K. Khalil, “On the design of robust servomechanisms for minimum phase nonlinear systems,” in 37th IEEE Conference on Decision and Control, 1998, pp. 3075–3080.
[29] S. Campbell and C. Meyer, Generalized Inverses of Linear Transformations. SIAM, 2009.
[30] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P. Glorennec, H. Hjalmarsson, and A. Juditsky, “Nonlinear black-box modeling in system identification: a unified overview,” Automatica, vol. 31, pp. 1691–1724, 1995.
[31] V. Andrieu, L. Praly, and A. Astolfi, “Homogeneous approximation, recursive observer design, and output feedback,” SIAM J. Contr. Opt., vol. 47, no. 4, pp. 1814–1850, 2008.
[32] M. Bin and L. Marconi, “Model identification and adaptive state estimation for a class of nonlinear systems;” under Review.
[33] J. Sjöberg, T. McKelvey, and L. Ljung, “On the use of regularization in system identification,” in IFAC 12th World Congress, 1993.
[34] R. Lee and M. Smith, “Nonlinear control for robust rejection of periodic disturbances,” Systems & Control Letters, vol. 39, pp. 97–107, 2000.