Pergamon

PII: S0005–1098(98)00064–8

Automatica, Vol. 34, No. 10, pp. 1261–1266, 1998

Printed in Great Britain
0005-1098/98 $—see front matter

Brief Paper

On Approximate Model-Reference Control of SISO Discrete-Time Nonlinear Systems*

HENK NIJMEIJER†‡ and SERGIO M. SAVARESI§

Key Words—Nonlinear systems; model matching control; discrete-time systems; non-linear predictor; distance.

Abstract—For discrete-time single-input single-output nonlinear plant and reference model, the model matching problem does not have an exact solution if the relative degree of the plant is larger than that of the reference model, but only approximate solutions can be considered. The notion of best approximate controller strictly depends on the definition of distance between the control system and the reference model. The goal of the paper is to analyze the performance limitations, which are inherently present because of the mismatch between the relative degrees of the plant and the model. Two different types of reference models (linear and nonlinear), and two different notions of distance (∞-norm-based and 2-norm-based distance) are considered. Upper and lower bounds for the maximum achievable performances are given, as a function of the parameters of the reference model. © 1998 Elsevier Science Ltd. All rights reserved.

1. *Introduction and problem statement*

The control problem we deal with is a model-matching problem (MMP), whose main elements and features are sketched in Fig. 1. This design problem can be roughly summarized as follows: given a nonlinear system Σ, a nonlinear controller Π, and a reference model Ψ (linear or nonlinear), all of them being discrete-time single-input single-output (SISO) systems, the problem is to select the controller Π such that the control system $\Sigma \circ \Pi$ is as close as possible to the reference model Ψ, according to some notion of distance $d[\Sigma \circ \Pi, \Psi]$. In particular, we restrict the input signal $v(t)$ to be subject to the constraint $|v(t)| \leq \bar{v}$, $\forall t$, and we consider this MMP in a deterministic setting, over a finite time horizon ($t \in \{0, 1, \dots, T\}$).

The above-sketched problem is, *per se*, quite general, and a large variety of control system design problems can be easily re-cast in order to fit with such a formulation.

The main goal of this paper is to investigate on a particular case of the MMP, namely *the case where the relative degree r_Σ of the system Σ is strictly larger than the relative degree r_Ψ of the model Ψ* (that is, roughly speaking, where the intrinsic input/output time-delay of Σ is strictly larger than that of Ψ). Such a "mismatch" between the relative degrees of Σ and Ψ raises many practical and theoretical issues, since, in this situation, the MMP does not admit, in general, an *exact* solution (namely a controller in correspondence of which the distance $d[\Sigma \circ \Pi, \Psi]$ is zero), whatever the controller Π is, but an *approximate* solution has to be searched for (Allgöwer, 1995; Savaresi, 1997).

The aim of the analysis we propose is to provide a better insight into the limit of performance which are *inherently* associated with any specific choice of a reference model with relative degree smaller than r_Σ. Such an analysis can also be helpful in providing some guidelines in the choice of *alternative* reference models, namely in the way the original reference model can be transformed into a model of relative degree r_Σ, with minimum loss of performance.

To formalize the MMP above sketched, each element which plays a role in such a design problem (see Fig. 1) is now briefly presented and discussed.

1. *The control system.* We assume that the system Σ is a discrete-time, time-invariant SISO system, of order $m \geq r_\Sigma$ and relative degree r_Σ, and we consider controllers $\Pi \in \wp$, $\wp$ being the set of all the time-invariant, discrete-time systems of relative degree larger or equal than 0 (causal systems), and of order smaller or equal than $m + n + r_\Sigma - 1$ (n being the order of the reference model). Moreover, we make the assumption that Σ is such that *for every time-invariant discrete-time SISO system Ξ of dimension smaller or equal than* $n + r_\Sigma - 1$ and of *relative degree larger or equal than* r_Σ there always exists a controller $\Pi \in \wp$ such that *an exact model matching* between Ξ and $\Sigma \circ \Pi$ can be achieved (over the considered time and input domains). This assumption on Σ is not, in general, particularly restrictive, since the model-matching between Σ and a system of relative degree larger or equal than r_Σ is inherently a "well-posed" problem (Kotta, 1995; Kotta and Nijmeijer, 1994; Njimeijer and van der Schaft, 1990). Since we are mainly interested in the analysis of the limit of performances imposed by the "mismatch" between the relative degrees of Σ and Ψ, the meaning of this assumption is simply to exclude from our analysis all other types (than the relative degree mismatch) of performance limitations.

2. *The reference model.* Two settings will be considered, namely the case when Ψ is linear, and the case when Ψ is nonlinear. Since, in our analysis, we adopt a purely input/output viewpoint, a natural choice for the reference model is an I/O difference-equation-based representation (see e.g. Narendra and Parthasarathy, 1990; Sjöberg, 1995). Moreover, for the sake of simplicity and without loss of generality, we always consider reference models of relative degree 0, and we always omit the index "Σ" to indicate the relative degree of Σ. Hence, we will consider time-invariant discrete-time SISO reference models of order n, having the form (subscripts "ℓ" and "$n\ell$" indicate linear and nonlinear models, respectively):

$$\Psi_\ell: \quad y(t) = a_1 y(t-1) + a_2 y(t-2) + \cdots + a_n y(t-n) + b_0 v(t) + b_1 v(t-1) + \cdots + b_n v(t-n), \tag{1}$$

$$\Psi_{n\ell}: \quad y(t) = f(y(t-1), y(t-2), \dots, y(t-n), v(t), v(t-1), \dots, v(t-n)). \tag{2}$$

* Received 23 May 1997; revised 29 January 1998; received in final form 25 March 1998. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Alberto Isidori under the direction of Editor Tarmer Başar. *Corresponding Author* Dr. Sergio M. Savaresi. Tel. + 39 2.23993556; Fax 39 2.23993412; E-mail savaresi@elet.polimi.it.

†Faculty of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

‡Faculty of Mechanical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. E-mail: h.nijmeijer@math.utwente.nl. Tel. + + 31.53. 4893442-Fax. + + 31.53.4340733.

§Dipartimento di Elettronica e Informazione, Politecnico di Milano, Piazza Leonardo da Vinci, 32, Milano, Italy.

For the regularity of f, we make the (conservative) assumption that f is of class $\mathscr{C}^{\infty}$ with respect to all of its arguments.

3. *The time horizon.* The MMP between the control system $\Sigma \circ \Pi$ and the reference model Ψ is considered over a finite time horizon $t \in \{0, 1, \ldots, T\}$, T being a *finite positive integer*, larger or equal than r_{Σ}.

4. *The input domain.* The input signal $v(t)$ is assumed to be subject only to an amplitude-bound, namely $|v(t)| \le \bar{v}$, $\bar{v}$ being a positive number; hence, each input sequence $\mathbf{v} := \{v(1), v(2), \ldots, v(T)\}$ belongs to a T-dimensional hyper-cube V centered in the origin, namely $\mathbf{v} \in V$, $V = [-\bar{v}, \bar{v}] \times [-\bar{v}, \bar{v}] \times \cdots \times [-\bar{v}, \bar{v}] \subset \Re^{T}$.

5. *The notion of distance.* Since the MMP we focus on can only be solved in an approximate fashion, the notion of distance $d[\Sigma \circ \Pi, \Psi]$, between the control system $\Sigma \circ \Pi$ and the reference model Ψ, plays a central role. As a matter of fact, the notion of the "best" controller strongly depends on the notion of distance we take.

In the rest of the paper, we consider the following two different definitions of (normalized) distance:

$$d_{\infty}[\Sigma \circ \Pi, \Psi] := \max_{\mathbf{v} \in V} \left\{ \max_{t \in \{1, 2, \ldots, T\}} |y_1(t; \Pi) - y_2(t))| / \bar{\mathbf{v}} \right\}, \quad (4)$$

$$d_2[\Sigma \circ \Pi, \Psi] := \max_{\mathbf{v} \in V} \left\{ \left(\frac{1}{T} \sum_{t=1}^{T} (y_1(t; \Pi) - y_2(t))^2 \right)^2 \Big/ \bar{v} \right\}, \quad (5)$$

where (see Fig. 1) $y_1(t; \Pi)$ and $y_2(t)$ are the output of $\Sigma \circ \Pi$ and Ψ, respectively. The subscripts "∞" and "2", used in equations (3) and (4) respectively, refer to the fact that equations (3) and (4) are based upon the "truncated" (at time T) version of the standard notion of ∞-norm and 2-norm of signals (Doyle *et al.*, 1992).

6. *The initial conditions.* We exclude from our analysis the situation where the distance between $\Sigma \circ \Pi$ and Ψ is biased by different initial conditions. To this end, we make the assumption that, at time $t = 0$, both $\Sigma \circ \Pi$ and Ψ are at rest, and that in such equilibrium conditions they are characterized by the same output. Without loss of generality, we assume that this equilibrium condition is the origin, namely that and $v(t) = y_1(t; \Pi) = y_1(t) = 0$.

Within the setting outlined in the issues 1–6 above, the rest of this paper is entirely devoted to the analysis of the following quantity (where d can be d_{∞} or d_2, and Ψ can be Ψ_{ℓ} or $\Psi_{n\ell}$):

$$\min_{\Pi \in \wp} \{d[\Sigma \circ \Pi, \Psi]\}. \quad (5)$$

In particular, our goal is to give lower an upper bounds for equation (5), as a function of the system characteristics of Ψ.

We will consider this problem in four different settings, given by the combination of different types of models (Ψ_{ℓ} and $\Psi_{n\ell}$,) and of different definitions of distances (d_{∞} and d_2). For each setting, results of different type and "strength" have been obtained. In particular, when using the ∞-norm-based notion of distance (3), it is possible to prove strong "equality" results, namely it is possible to provide the value of the lowest possible distance achievable between $\Sigma \circ \Pi$ and Ψ, as a function of the parameters of the reference model Ψ. Instead, when the 2-norm-based notion of distance (4) is resorted to, only "inequality" results (namely lower and upper bounds for equation (5)) can be given.

Apparently, in the case of *linear* reference models (Section 2), the bounds on the best achievable performance can be explicitly given as a function of the coefficients of Ψ_{ℓ}, whereas in the case of *nonlinear* reference models (Section 3), in general only "implicit" definitions of the performance bounds can be provided.

2. *Limit of performances in MMPs: the linear case*

The first case we consider is the case where the reference model Ψ_{ℓ} is linear. Before stating the main results concerning this specific situation, some definitions are due.

First remind that the linear model Ψ_{ℓ} defined in equation (1) with a difference equation, can be also written in the following form (z^{-1}, as usual, indicates the 1-step-delay operator):

$$\Psi_{\ell}: \quad y(t) = \frac{B(z^{-1})}{A(z^{-1})} v(t), \; A(z^{-1}) := 1 - a_1 z^{-1} - \cdots - a_n z^{-n},$$

$$B(z^{-1}) := b_0 + b_1 z^{-1} + \cdots + b_n z^{-n}, \; b_0 \neq 0. \quad (6)$$

By computing the r-steps polynomial division between $B(z^{-1})$ and $A(z^{-1})$, namely $B(z^{-1}) = E(z^{-1}) A(z^{-1}) + F(z^{-1}) z^{-r}$ (where $E(z^{-1}) = e_0 + e_1 z^{-1} + \cdots + e_{r-1} z^{-r+1}$ and $F(z^{-1}) z^{-r} = (f_0 + f_1 z^{-1} + \cdots + f_n z^{-n}) z^{-r}$ are the result and the residual, respectively, of such a division), it is easy to see that equation (6) can be also given the form

$$\Psi_{\ell}: \quad y(t) = E(z^{-1}) v(t) + \frac{F(z^{-1})}{A(z^{-1})} v(t - r). \quad (7)$$

The main advantage of using the form (7) of Ψ_{ℓ} is that it clearly separates the part of $y(t)$ which depends on $\{v(t), \ldots, v(t - r + 1)\}$, from that which depends on $\{v(t - r), \ldots, v(1)\}$. Using equation (7), some results can now be stated.

Theorem 1. Under the assumptions 1–6, $\min_{\Pi \in \wp} \{d_{\infty}[\Sigma \circ \Pi, \Psi_{\ell}]\} = |e_0| + |e_1| + |e_2| + \cdots + |e_{r-1}|$, where $e_0, e_1, \ldots, e_{r-1}$ follows from the representation (7) of Ψ_{ℓ}.

Proof. Consider the matching error between the outputs of $\Sigma \circ \Pi$ ($\Pi \in \wp$) and Ψ_{ℓ}, at a time instant t ($r \le t \le T$):

$$|y_1(t/t - r; \Pi) - y_2(t/t)| \quad (8)$$

(the time index after the "/" indicates that signal depends on inputs $v(\cdot)$ up to that time). Notice that, due to the assumptions on their relative degrees, the outputs of $\Sigma \circ \Pi$ and Ψ_{ℓ} depend, at time t, on the inputs up to time $t - r$ and t, respectively. Using

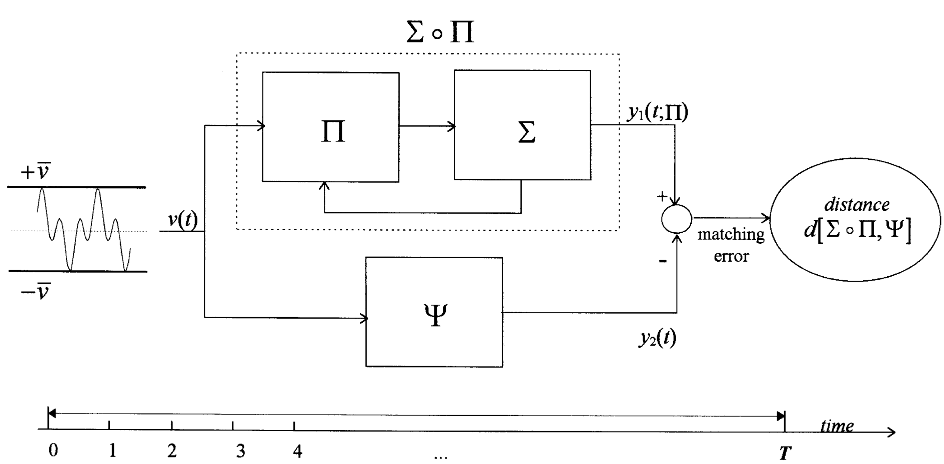

Fig. 1. Basic elements of the model-matching problem.

the expression (7) of Ψ_ℓ, the matching-error can be given the following form:

$$|y_1(t/t-r;\Pi)-y_2(t/t)| = \Bigg|[e_0 v(t)+e_1 v(t-1)+\cdots + e_{r-1}v(t-r+1)] + \left[\frac{F(z^{-1})}{A(z^{-1})}v(t-r)-y_1(t/t-r;\Pi)\right]\Bigg|.$$

Given the input sequence $\{v(t-r),\dots,v(1)\}$, select the inputs $\{v(t), v(t-1),\dots,v(t-r+1)\}$ as follows:

$$\begin{cases} v(t) = -\bar{v}\,\mathrm{sign}\langle e_0\rangle \times \mathrm{sign}\left\langle\left[\frac{F(z^{-1})}{A(z^{-1})}v(t-r)-y_1(t/t-r;\Pi)\right]\right\rangle,\\ v(t-1) = -\bar{v}\,\mathrm{sign}\langle e_1\rangle \times \mathrm{sign}\left[\frac{F(z^{-1})}{A(z^{-1})}v(t-r)-y_1(t/t-r;\Pi)\right]\Big\rangle,\\ \vdots\\ v(t-r+1) = -\bar{v}\,\mathrm{sign}\langle e_{r-1}\rangle \times \mathrm{sign}\left\langle\left[\frac{F(z^{-1})}{A(z^{-1})}v(t-r)-y_1(t/t-r;\Pi)\right]\right\rangle \end{cases} \quad (19)$$

(where $\mathrm{sign}\langle x\rangle = +1$ if $x \ge 0$, and $\mathrm{sign}\langle x\rangle = -1$ if $x<0$). Notice that equation (9) is an admissible choice for the inputs $\{v(t), v(t-1),\dots,v(t-r+1)\}$, and that, if equation (9) is used, the matching error at time t is such that

$$|y_1(t/t-r;\Pi)-y_2(t/t)| \ge (|e_0|+\cdots+|e_{r-1}|)\cdot\bar{v} + \left|\frac{F(z^{-1})}{A(z^{-1})}v(t-r)-y_1(t/t-r;\Pi)\right| \ge (|e_0|+\cdots+|e_{r-1}|)\cdot\bar{v}. \quad (10)$$

This means that it is always possible ($\forall\Pi\in\wp$) to "cause" a matching error at time t larger or equal than $(|e_0|+\cdots+|e_{r-1}|)\cdot\bar{v}$, by choosing an admissible input sequence as equation (9). Hence, by the definition of distance (3):

$$d_\infty[\Sigma\circ\Pi,\Psi_\ell] \ge |e_0|+|e_1|+\cdots+|e_{r-1}|. \quad (11)$$

Notice that, since the lower bound stated by equation (11) *does not depend on* Π, if there exists a controller Π such that $d_\infty[\Sigma\circ\Pi,\Psi_\ell] = |e_0|+|e_1|+\cdots+|e_{r-1}|$, this would imply that $|e_0|+\cdots+|e_{r-1}|$ is not only a lower bound for $d_\infty[\Sigma\circ\Pi,\Psi_\ell]$, but it is exactly the best achievable result. Therefore, consider the controller $\Pi^0\in\wp$ such that

$$\Sigma\circ\Pi^0:\quad y(t) = \frac{F(z^{-1})z^{-r}}{A(z^{-1})}v(t); \quad (12)$$

notice that such a controller exists (by virtue of the assumptions made in 1), since equation (12) is a system having relative degree r. Using such a parameter vector, the following holds:

$$\begin{aligned} d_\infty&[\Sigma\circ\Pi^0,\Psi_\ell]\\ &= \max_{v\in V}\left\{\max_{t\in\{1,2,\dots,T\}} |y_1(t;\Pi^0)-y_2(t)|/\bar{v}\right\}\\ &= \max_{v\in V}\left\{\max_{t\in\{1,2,\dots,T\}} \left|\frac{F(z^{-1})z^{-r}}{A(z^{-1})}v(t)-\frac{B(z^{-1})}{A(z^{-1})}v(t)\right| \Big/ \bar{v}\right\}\\ &= \max_{v\in V}\left\{\max_{t\in\{1,2,\dots,T\}} |e_0v(t)+e_1v(t-1)+\cdots + e_{r-1}v(t-r+1)|/\bar{v}\right\} = |e_0|+|e_1|+\cdots+|e_{r-1}|. \end{aligned}$$

Hence, since there exists at least one (admissible) controller, for which the lower bound in equation (11) is exactly attained, we can conclude that $\min_{\Pi\in\wp}\{d_\infty[\Sigma\circ\Pi,\Psi_\ell]\} = |e_0|+|e_1|+|e_2|+\cdots+|e_{r-1}|$. □

Theorem 2. Under the assumptions 1–6, $|e_0| \le \min_{\Pi\in\wp}\{d_\infty[\Sigma\circ\Pi,\Psi_\ell]\} \le |e_0|+|e_1|+|e_2|+\cdots+|e_{r-1}|$, where $e_0,e_1,\dots,e_{r-1}$ follows from the representation (7) of Ψ_ℓ.

Proof. The upper bound can be easily proved by reminding the result stated in Theorem 1, and the definition of 2-norm-based distance (4):

$$\begin{aligned} \min_{\Pi\in\wp}&\{d_2[\Sigma\circ\Pi,\Psi_\ell]\}\\ &= \min_{\Pi\in\wp}\left\{\max_{v\in V}\left\{\left(\frac{1}{T}\sum_{t=1}^{T}(y_1(t;\Pi)-y_2(t))^2\right)^{1/2}\Big/\bar{v}\right\}\right\}\\ &\le \min_{\Pi\in\wp}\left\{\max_{v\in V}\left\{\left(\frac{1}{T}\sum_{t=1}^{T}\left(\max_{t\in\{1,2,\dots,T\}}|y_1(t;\Pi)-y_2(t)|\right)^2\right)^{1/2}\Big/\bar{v}\right\}\right\}\\ &= \min_{\Pi\in\wp}\left\{\max_{v\in V}\left\{\max_{t\in\{1,2,\dots,T\}}|y_1(t;\Pi)-y_2(t)|\Big/\bar{v}\right\}\right\}\\ &= |e_0|+|e_1|+\cdots+|e_{r-1}| \end{aligned}$$

As for the lower bound, first notice that, for every $\Pi\in\wp$ and for every time $t\le T$, the following holds:

$$y_1(t/t-r;\Pi)-y_2(t/t) = e_0v(t)+\gamma(v(t-1), v(t-2),\dots,v(1);t;\Pi),$$

where γ is a function depending on the time instant t, on the controller $\Pi\in\wp$, and on the inputs up to $t-1$. Hence, if we choose the input $v(t)$ as $v(t) = -\bar{v}\,\mathrm{sign}\langle e_0\rangle\,\mathrm{sign}\langle\gamma(v(t-1), v(t-2),\dots,v(1);t;\Pi)\rangle$, we can always "cause", at every time instant $0\le t\le T$, a matching error such that $|y_1(t/t-r;\Pi)-y_2(t/t)| \ge |e_0|\cdot\bar{v}$, $\forall t\in\{0,1,\dots,T\}$.

In other words, we can always find an admissible input sequence such that

$$\max_{v\in V}\left\{\left(\frac{1}{T}\sum_{t=1}^{T}(y_1(t;\Pi)-y_2(t))^2\right)^{1/2}\Big/\bar{v}\right\} \ge |e_0|.$$

Since such lower bound does not depend on Π, $|e_0| \le \min_{\Pi\in\wp}\{d_2[\Sigma\circ\Pi,\Psi_\ell]\} \le |e_0|+|e_1|+|e_2|+\cdots+|e_{r-1}|$. □

A simple but interesting corollary of Theorems 1 and 2 is the following:

Corollary 1. Under the assumptions 1–6, *if the relative degree* r *of the system* Σ *is* 1, the following holds:

$$\max_{\Pi\in\wp}\{d_2[\Sigma\circ\Pi,\Psi_\ell]\} = \max_{\Pi\in\wp}\{d_\infty[\Sigma\circ\Pi,\Psi_\ell]\} = |e_0|.$$

Proof. The proof is trivial, and can be obtained by simply combining the results of Theorems 1 and 2, when $r=1$. □

Notice that Corollary 1 states that, even though in general the "best" controller designed according to the notion of ∞-norm and the "best" controller designed according to the notion of 2-norm are different, in the case when the relative degree of the system is 1, they coincide, and a strong "equality" result holds in both cases.

Remark 1 (*r-steps-ahead predictor*). It is interesting to point out that the Theorem 1 not only provides an explicit expression for the lower bound of $d_\infty[\Sigma\circ\Pi,\Psi_\ell]$, but also gives the explicit expression of the dynamic behavior of the control system at the optimum:

$$y(t) = (F(z^{-1})/A(z^{-1}))\,v(t-r). \quad (13)$$

It is interesting to remind that equation (13) plays a very special role in the context of the theory of linear stochastic and stationary processes (see e.g. Ljung, 1987), namely it is the best r-steps-ahead predictor for the system (1), when the input is a zero-mean stationary white noise, and the norm of the signals is given by their variance. In the rest of the paper we therefore refer to equation (13) as the "best r-steps-ahead predictor of Ψ_ℓ" (using the symbol $\Psi_\ell^{(r)}$).

Notice that, in our setting, the concept of "best r-steps-ahead predictor" has a purely deterministic meaning, and its optimality is related to the notion of norm (3), i.e. $\Psi_\ell^{(r)}$ has the special property of being the system with minimum distance $d_\infty[\Psi_\ell^{(r)}, \Psi_\ell]$ from Ψ_ℓ, among all the finite-dimensional systems of relative degree larger or equal than r. □

3. *Limit of performances in MMPs: the nonlinear case*

The second case we consider refers to the case when a nonlinear reference model $\Psi_{n\ell}$ is used.

In the linear case, a crucial role is played by the representation (7) of the linear model Ψ_ℓ; such expression of Ψ_ℓ is useful since it separates the part of $y(t)$ which depends on inputs $\{v(t), \dots, v(t-r+1)\}$, from that which depends on inputs $\{v(t-r), \dots, v(1)\}$; in particular, using such a representation, it is straightforward to obtain the system $\Psi_\ell^{(r)}$, namely the closest (in the sense of the ∞-norm-based notion of distance (3)) system to Ψ_ℓ, having relative degree r.

Analogous to the linear case, even in the nonlinear case the best r-steps-ahead deterministic predictor of the model $\Psi_{n\ell}$ plays a central role in determining the best achievable performance. Therefore, the first part of this section is devoted to the generalization of the "best r-steps-ahead deterministic predictor", to the nonlinear case.

The first step in the construction of $\Psi_\ell^{(r)}$ is to re-write the original expression

$$\Psi_{n\ell}:\quad y(t) = f(y(t-1), y(t-2), \dots, y(t-n), v(t), v(t-1), \dots, v(t-n))$$

of $\Psi_{n\ell}$ in a different form, by recursively substituting $r-1$ times the expression of $y(t-1), \dots, y(t-r+1)$ into equation (2):

$$\begin{aligned}
y(t) &= f(f(y(t-2), \dots, y(t-n-1), v(t-1), \dots, v(t-n-1)),\\
&\quad y(t-2), \dots, y(t-n), v(t), \dots, v(t-n)),\\
&:= f_{(2)}(y(t-2), y(t-3), \dots, y(t-n-1), v(t),\\
&\quad v(t-1), \dots, v(t-n-1)),\\
y(t) &= f_{(2)}(f(y(t-3), \dots, y(t-n-2), v(t-2), \dots,\\
&\quad v(t-n-2)), y(t-3), \dots, y(t-n-1), v(t), \dots,\\
&\quad v(t-n-1))\\
&:= f_{(3)}(y(t-3), y(t-4), \dots, y(t-n-2), v(t), v(t-1), \dots,\\
&\quad v(t-n-2)),\\
&\vdots\\
y(t) &= f_{(r-1)}(f(y(t-r), \dots, y(t-n-r+1), v(t-r+1), \dots,\\
&\quad v(t-n-r+1)), y(t-r), \dots, y(t-n-r+2), v(t), \dots,\\
&\quad v(t-n-r+2))\\
&:= f_{(r)}(y(t-r), y(t-r-1), \dots, y(t-n-r+1),\\
&\quad v(t), \dots, v(t-r+1), v(t-r), \dots, v(t-n-r+1)).
\end{aligned}$$

In the above expressions, with the symbol $f_{(i)}$ we indicate the function obtained after $i-1$ substitutions; notice that the index i in $f_{(i)}$ also indicates the number of delays between the output and the first appearance of y within $f_{(i)}$ (hence, using this notation, the original function f is named $f_{(1)} = f$).

The function we are mainly interested in is the function $f_{(r)}$, namely the function obtained after $r-1$ substitutions. In particular, by defining the n-dimensional vectors $\mathbf{y}_n(t)$ and $\mathbf{v}_n(t)$ as

$$\mathbf{y}_n(t) := \{y(t), y(t-1), \dots, y(t-n+1)\},$$

$$\mathbf{v}_n(t) := \{v(t), v(t-1), \dots, v(t-n+1)\},$$

and by using the function $f_{(r)}$ above, the nonlinear model $\Psi_{n\ell}$ can be given the following form:

$$\Psi_{n\ell}:\quad y(t) = f_{(r)}(\mathbf{y}_n(t-r), v(t), v(t-1), \dots, v(t-r+1), \mathbf{v}_n(t-r)). \tag{14}$$

Notice that the arguments of the above expression of $\Psi_{n\ell}$ are divided into three groups, namely the past outputs $\{y(t-r), \dots, y(t-n-r+1)\}$ (collected in the vector $\mathbf{y}_n(t-r)$), the past inputs up to time $t-r$ $\{v(t-r), \dots, v(t-n-r+1)\}$ (collected in the vector $\mathbf{v}_n(t-r)$), and the last r inputs $\{v(t), \dots, v(t-r+1)\}$.

Consider now the $2n$-dimensional vector $\{\mathbf{y}_n(t), \mathbf{v}_n(t)\}$, and consider the set of all the values such vector may take, for $t \le T-r$, when admissible inputs sequences $\mathbf{v} \in V$ (see 4) are used, namely $\Theta := \{\{\mathbf{y}_n(t), v_n(t)\}/t \le T-r \wedge \mathbf{v} \in V\}$.

Thanks to the hypothesis made on the regularity of the function f, it can be shown that Θ is a compact closed subset of $\Re^{2n}$, containing the origin. Corresponding to every $\{\mathbf{y}_n, \mathbf{v}_n\} \in \Theta$, consider now the following quantities:

$$\begin{cases}
y_{(r)}^{\max}(\mathbf{y}_n, \mathbf{v}_n) := \max\limits_{|v_1| \le \bar{v}, \dots, |v_r| \le \bar{v}} \{f_{(r)}(\mathbf{y}_n, v_1, v_2, \dots, v_r, \mathbf{v}_n)\} & \\
 & \{\mathbf{y}_n, \mathbf{v}_n\} \in \Theta\\
y_{(r)}^{\min}(\mathbf{y}_n, \mathbf{v}_n) := \max\limits_{|v_1| \le \bar{v}, \dots, |v_r| \le \bar{v}} \{f_{(r)}(\mathbf{y}_n, v_1, v_2, \dots, v_r, \mathbf{v}_n)\} &
\end{cases} \tag{15}$$

Notice that both $y_{(r)}^{\max}(\mathbf{y}_n, \mathbf{v}_n)$ and $y_{(r)}^{\min}(\mathbf{y}_n, \mathbf{v}_n)$ exist for every $\{\mathbf{y}_n, \mathbf{v}_n\} \in \Theta$, because of the regularity of f.

Using the definition (15), consider now the (nonlinear) mapping χ from $\Re^{2n}$ to $\Re^r$, defined as

$$\chi:\quad \Theta \to [-\bar{v}, \bar{v}] \times [-\bar{v}, \bar{v}] \times \cdots \times [-\bar{v}, \bar{v}]$$

$$\{\mathbf{y}_n, \mathbf{v}_n\} \to \{v_1, v_2, \dots, v_r\} \tag{16a}$$

such that

$$f_{(r)}(\mathbf{y}_n, \chi_1(\mathbf{y}_n, \mathbf{v}_n), \chi_2(\mathbf{y}_n, \mathbf{v}_n), \dots, \chi_r(\mathbf{y}_n, \mathbf{v}_n), \mathbf{v}_r)$$
$$= (y_{(r)}^{\max}(\mathbf{y}_n, \mathbf{v}_n) + y_{(r)}^{\min}(\mathbf{y}_n, \mathbf{v}_n))/2, \quad \forall \{\mathbf{y}_n \mathbf{v}_n\} \in \Theta, \tag{16b}$$

$$|\chi_i(\mathbf{y}_n, \mathbf{v}_n)| \le \bar{v}, \quad i = 1, 2, \dots, r, \quad \forall \{\mathbf{y}_n \mathbf{v}_n\} \in \Theta \tag{16c}$$

(where the symbols $\chi_1, \chi_2, \dots, \chi_r$ indicate the r components of the mapping χ). Notice that, using the definition (15) of $y_{(r)}^{\max}(\mathbf{y}_n, \mathbf{v}_n)$ and $y_{(r)}^{\min}(\mathbf{y}_n, \mathbf{v}_n)$, a mapping χ which fulfills the requirements (16b) and (16c) is *guaranteed to exist*, even though, in general, it is *not unique*. Using the mapping χ, the following nonlinear system $\Psi_{n\ell}^{(r)}$ can be defined:

$$\Psi_{n\ell}^{(r)}:\begin{cases}
\hat{y}(t) = f_{(r)}(\mathbf{y}_n(t-r), \chi_1(\mathbf{y}_n(t-r), \mathbf{v}_n(t-r)), \dots,\\
\qquad \chi_r(\mathbf{y}_n(t-r), \mathbf{v}_n(t-r)), \mathbf{v}_n(t-r)),\\
y(t-r) = f_{(r)}(y(t-r-1), \dots, y(t-r-n),\\
\qquad v(t-r), \dots, v(t-r-n)).
\end{cases} \tag{17}$$

The system (17) plays a central role in determining the best achievable performances, when the nonlinear reference model $\Psi_{n\ell}$ is used (see Theorems 3 and 4). About $\Psi_{n\ell}^{(r)}$, the following remarks can be made:

- As it will be shown in Theorem 3, equation (17) is the nonlinear counterpart of $\Psi_\ell^{(r)}$, and can be seen as the best r-steps-ahead nonlinear predictor of $\Psi_{n\ell}$, when the notion of distance (3) is used.
- It is important to remark that, even though the mapping χ defined in equation (16) is, in general, not unique, the I/O behavior of $\Psi_\ell^{(r)}$ defined in equation (17) is uniquely determined, when the reference model $\Psi_{n\ell}$ is fixed.
- The worst (normalized) r-steps-ahead "prediction-error", which $\Psi_\ell^{(r)}$ can make, is given by

$$\Delta_{(r)} := \max_{\{\mathbf{y}_n, \mathbf{v}_n\} \in \Theta} \{(y_{(r)}^{\max}(\mathbf{y}_n, \mathbf{v}_n) - y_{(r)}^{\min}(\mathbf{y}_n, \mathbf{v}_n))/2v\}. \tag{18}$$

$\Delta_{(r)}$ also represents the ∞-norm-based distance between $\Psi_\ell^{(r)}$ and $\Psi_{n\ell}$, namely $d_\infty[\Psi_{n\ell}^{(r)}, \Psi_{n\ell}] = \Delta_{(r)}$.

Using the above definitions, two results can now be proved.

Theorem 3. Under the assumptions 1–6, $\min_{\Pi \in \wp}\{d_\infty[\Sigma \circ \Pi, \Psi_{n\ell}]\} = \Delta_{(r)}$, with $\Delta_{(r)}$ defined in equation (18).

Proof. First consider a $2n$-dimensional vector $\{\tilde{\mathbf{y}}_n, \tilde{\mathbf{v}}_n\} \in \Theta$ (which, in general, is not unique) such that

$$(y_{(r)}^{\max}(\tilde{\mathbf{y}}_n, \tilde{\mathbf{v}}_n) - y_{(r)}^{\min}(\tilde{\mathbf{y}}_n, \tilde{\mathbf{v}}_n))/2\bar{v} = \Delta_{(r)}. \tag{19}$$

Notice that (due to the definition of Θ) it is always possible to find an admissible input sequence $\{v(t), v(t-1), \ldots, v(1)\}$, $t \leq T - r$, such that, when the nonlinear reference model $\Psi_{n\ell}$ is fed with such an input:

$$\{v(t), v(t-1), \ldots, v(t-n+1)\} = \tilde{\mathbf{v}}_n,$$
$$\{y(t), y(t-1), \ldots, y(t-n+1)\} = \tilde{\mathbf{y}}_n,$$

where $\{v(t), \ldots, v(t-n+1)\}$ are the last n samples of such an input sequence, and $\{y(t), \ldots, y(t-n+1)\}$ are the last n samples of the corresponding output of the reference model $\Psi_{n\ell}$. If we feed both the reference model $\Psi_{n\ell}$ and the control system $\Sigma \circ \Pi$ with such a "special" input sequence, at time $t + r$ the following holds:

$$|y_1(t+r/t; \Pi) - y_2(t+r/t+r)| = |y_1(t+r/t; \Pi)$$
$$- f_{(r)}(\tilde{\mathbf{y}}_n, v(t+r), v(t+r-1), \ldots, v(t+1), \tilde{\mathbf{v}}_n)|.$$

Notice that, at time $t + r$, the output of the control system $\Sigma \circ \Pi$ does not depend on $\{v(t+r), \ldots, v(t+1)\}$ (since the relative degree of $\Sigma \circ \Pi$ is larger or equal than r, $y_1(t+r/t; \Pi)$ it is uniquely determined by the input sequence up to time t). Hence, since we have assumed that $\{\tilde{\mathbf{y}}_n, \tilde{\mathbf{v}}_n\} \in \Theta$ is such that equation (19) holds, it is possible to select the inputs $\{v(t+r), \ldots, v(t+1)\}$ so that the corresponding matching error at time $t + r$ is larger or equal than $\Delta_{(r)}$. Hence, by virtue of the definition of distance (3), it is always true that

$$d_\infty[\Sigma \circ \Pi, \Psi_{n\ell}] \geq \Delta_{(r)}, \quad \forall \Pi \in \wp. \tag{20}$$

Notice that, since the lower bound stated by equation (20) *does not depend on* Π, if there exists a Π such that $d_\infty[\Sigma \circ \Pi, \Psi_{n\ell}] = \Delta_{(r)}$, this would imply that $\Delta_{(r)}$ is not only a lower bound for $d_\infty[\Sigma \circ \Pi, \Psi_{n\ell}]$, but it is exactly the best achievable result.

Now consider the controller $\Pi^0 \in \wp$, such that $\Sigma \circ \Pi^0 = \Psi^{(r)}_{n\ell}$, $\Psi^{(r)}_{n\ell}$ being the best nonlinear r-steps-ahead predictor for $\Psi_{n\ell}$, whose expression is given by equation (17). Such a controller exists (by virtue of the assumptions made in 1), since equation (17) is a system of order $n + r - 1$ having relative degree r. Using such a controller, and reminding the definition of $\Psi^{(r)}_{n\ell}$, $d_\infty[\Sigma \circ \Pi^0, \Psi_{n\ell}] = d_\infty[\Psi^{(r)}_{n\ell}, \Psi_{n\ell}] = \Delta_{(r)}$. Hence, since there exists at least one (admissible) controller, in correspondence of which the lower bound stated by equation (20) is reached, we conclude that $\min_{\Pi \in \wp}[d_\infty[\Sigma \circ \Pi, \Psi_{n\ell}]\} = \Delta_{(r)}$. □

Before presenting the main result for the case when the controller is designed according to the 2-norm-based notion of distance (4) (Theorem 4), notice that the nonlinear model (2) can be rewritten in compact form as

$$\Psi_{n\ell}: \quad y(t) = f_{(1)}(\mathbf{y}_n(t-1), v(t), \mathbf{v}_n(t-1)) \tag{21}$$

(where $\mathbf{y}_n(t-1)$ and $\mathbf{v}_n(t-1)$ are n-dimensional vectors, and $f_{(1)} = f$ is used instead of f). Then define

$$\left.\begin{aligned} y^{\max}_{(1)}(\mathbf{y}_n, \mathbf{v}_n) &:= \max_{|v_1| \leq \mathrm{v}} \{f_{(1)}(\mathbf{y}_n, v_1, \mathbf{v}_n)\} \\ y^{\min}_{(1)}(\mathbf{y}_n, \mathbf{v}_n) &:= \min_{|v_1| \leq \mathrm{v}} \{f_{(1)}(\mathbf{y}_n, v_1, \mathbf{v}_n)\} \end{aligned}\right\} \{\mathbf{y}_n, \mathbf{v}_n\} \in \Theta, \tag{22}$$

$$\delta_{(1)} := \min_{|\mathbf{y}_n, \mathbf{v}_n| \in \Theta} \{(y^{\max}_{(1)}(\mathbf{y}_n, \mathbf{v}_n) - y^{\min}_{(1)}(\mathbf{y}_n, \mathbf{v}_n))/2\bar{v}\}. \tag{23}$$

Theorem 4. Under the assumptions 1–6, $\delta_{(1)} \leq \min_{\Pi \in \wp}\{d_2[\Sigma \circ \Pi, \Psi_{n\ell}]\} \leq \Delta_{(r)}$, with $\Delta_{(r)}$ as in equation (18), and $\delta_{(1)}$ as in equation (30).

Proof. The upper bound can be proved by using Theorem 3, and the definition of distance (4):

$$\begin{aligned}
&\min_{\Pi \in \wp} \{d_2[\Sigma \circ \Pi, \Psi_{n\ell}]\} \\
&= \min_{\Pi \in \wp} \left\{ \max_{\mathbf{v} \in V} \left\{ \left(\frac{1}{T} \sum_{t=1}^{T} (y_1(t; \Pi) - y_2(t))^2 \right)^{1/2} \Big/ \bar{v} \right\} \right\} \\
&\leq \min_{\Pi \in \wp} \left\{ \max_{\mathbf{v} \in V} \left\{ \left(\frac{1}{T} \sum_{t=1}^{T} \left(\min_{t \in \{1,2,\ldots,T\}} |y_1(t; \Pi) - y_2(t)| \right)^2 \right)^{1/2} \Big/ \bar{v} \right\} \right\} \\
&= \min_{\Pi \in \wp} \left\{ \max_{\mathbf{v} \in V} \left\{ \min_{t \in \{1,2,\ldots,T\}} |y_1(t; \Pi) - y_2(t)| \Big/ \bar{v} \right\} \right\} \\
&= \min_{\Pi \in \wp} \{d_\infty[\Sigma \circ \Pi, \Psi_{n\ell}]\} = \Delta_{(r)}
\end{aligned}$$

As for the lower bound, notice that, by virtue of the definition (23) of $\delta_{(1)}$, we can always "cause" a matching error between $\Sigma \circ \Pi$ and $\Psi_{n\ell}$ such that $|y_1(t/t-r; \Pi) - y_2(t/t)| \geq \delta_{(1)} \cdot \bar{v}$, $\forall t \in \{1, 2, \ldots, T\}$. In other words, we can always find an admissible input sequence such that

$$\max_{\mathbf{v} \in V} \left\{ \left(\frac{1}{T} \sum_{t=1}^{T} (y_1(t; \Pi) - y_2(t))^2 \right)^{1/2} \Big/ \bar{v} \right\} \geq \delta_{(1)}.$$

Since such a lower bound does not depend on Π, $\delta_{(1)} \leq \min_{\Pi \in \wp}\{d_2[\Sigma \circ \Sigma, \Psi_{n\ell}]\} \leq \varDelta_{(r)}$. □

Similar to the linear case (see Remark 1) notice that Theorem 3 not only provides the lower bound of $d_\infty[\Sigma \circ \Pi, \Psi_{n\ell}]$, but also provides the expression of the dynamical behavior of the control system at the "optimum". As in the linear case, it is given by the best r-steps-ahead predictor $\Psi^{(r)}_{n\ell}$, defined as in equation (17).

It is interesting to notice that the procedure (outlined at the beginning of this subsection) used to build the *nonlinear* predictor (17) can also be used to compute the best predictor for a *linear* model (even though, in the linear case, the procedure described in Section 2 based on the polynomial division is, in general, more straightforward). In particular, a peculiar feature of the linear case is that the function χ defined in equation (16) is unique, and it is simply given by the null function $\chi_i(\mathbf{y}_n, \mathbf{v}_n) = 0$, $i = 1, 2, \ldots, r$, $\forall \{\mathbf{y}_n, \mathbf{v}_n\} \in \Theta$ (whereas, in general, in the nonlinear case, χ is a non-trivial function of its arguments, and the search for a function χ which fulfills the requirements (16b) and (16c) represents the most difficult step in the construction of the nonlinear predictor $\Psi^{(r)}_{n\ell}$).

To better illustrate the difference between the linear and the nonlinear case, a numerical example is given.

Example 1. Consider the following *nonlinear* system (of relative degree 0)

$$\Psi_{n\ell}: \quad y(t) = y(t-1) + \exp(v(t))v(t-1) + v(t-1)$$

over the time interval $t \in \{0, 1, \ldots, T\}$, where $|v(t)| \leq \bar{v}$. In order to find the closest system to $\Psi_{n\ell}$ of (for instance) relative degree 2, in the sense of the ∞-norm-based notion of distance (3), as a first step $\Psi_{n\ell}$ must be written as:

$$\begin{aligned} \Psi_{n\ell}: \quad y(t) &= [y(t-2) + \exp(v(t-1))v(t-2) + v(t-2)] \\ &\quad + \exp(v(t))v(t-1) + v(t-1), \end{aligned} \tag{24}$$

or, using the notation presented at the beginning of this subsection:

$$\Psi_{n\ell}: \quad y(t) = f_{(2)}(\mathbf{y}_1(t-2), v(t), v(t-1), \mathbf{v}_1(t-2)),$$
$$\mathbf{y}_1(t) := \{y(t)\}, \mathbf{v}_1(t) := \{v(t)\}.$$

A function χ which fulfills the conditions (16b) and (16c) for the system (24) is implicitly given by

$$\begin{aligned} &\exp(v(t))v(t-1) + v(t-1) + \exp(v(t-1))v(t-2) \\ &= v(t-2)\frac{\exp(\bar{v}) + \exp(-\bar{v})}{2}. \end{aligned}$$

By plugging the expression above in equation (24), the best 2-steps-ahead predictor is then obtained:

$$\Psi^{(2)}_{n\ell}: \begin{cases} \hat{y}(t) = y(t-2) + v(t-2) + v(t-2)\dfrac{\exp(\bar{v}) + \exp(-\bar{v})}{2} \\ y(t-2) = y(t-3) + \exp(v(t-2))v(t-3) + v(t-3) \end{cases} \tag{25}$$

$\hat{y}(t)$ being the output of such a system.

Consider now the *linear* system (of relative degree 0):

$$\Psi_{\ell}: \quad y(t) = y(t-1) + v(t) + v(t-1) \tag{26}$$

over the time interval $t \in \{0, 1, \ldots, T\}$, where $|v(t)| \leq \bar{v}$. Using the same procedure used in the nonlinear case, first rewrite equation

(26) in the following form:

$$\Psi_\ell\colon\ y(t) = [y(t-2) + v(t-1) + v(t-2)] + v(t) + v(t-1) = v(t) + 2v(t-1) + y(t-2) + v(t-2). \quad (27)$$

Since the part of the output which depends on $\{v(t), v(t-1)\}$ is separated by the rest, the function χ defined in equation (16) is implicitly given by

$$v(t) + 2v(t-1) = 0, \quad (28)$$

namely it is simply obtained by zeroing the part of the output which depends on $\{v(t), v(t-1)\}$, when the system is given the form equation (27). By plugging equation (28) into equation (27), the best 2-steps-ahead predictor of Ψ_ℓ can be obtained as:

$$\Psi_\ell^{(2)}\colon\ \begin{cases} \hat{y}(t) = y(t-2) + v(t-2), \\ y(t-2) = y(t-3) + v(t-2) + v(t-3). \end{cases} \quad (29)$$

Notice that equation (29) can be also rewritten as

$$\Psi_\ell^{(2)}\colon\ \hat{y}(t) = \frac{2z^{-2}}{1+z^{-1}} v(t),$$

which (obviously) exactly coincides with the 2-steps-ahead predictor one can obtain by using the polynomial-division-based procedure illustrated in Section 2. □

From the simple numerical example above, it should be apparent that the main difference between the linear and the nonlinear case, is that in the former case the predictor can be obtained by simply using $\chi \equiv 0$, whereas, in the latter, the search for a suitable function χ might be a non-trivial task.

Finally, it is interesting to notice that a result similar to that stated in the Corollary 1 does not hold in the nonlinear case. As a matter of fact, in the nonlinear case $\delta_{(1)} \leq \Delta_{(1)}$, since

$$\delta_{(1)} = \min_{\{\mathbf{y}_n, \mathbf{v}_n\} \in \Theta} \left\{ \frac{y_{(1)}^{\max}(\mathbf{y}_n, \mathbf{v}_n) - y_{(1)}^{\min}(\mathbf{y}_n, \mathbf{v}_n)}{2\bar{v}} \right\} \leq \min_{\{\mathbf{y}_n, \mathbf{v}_n\} \in \Theta} \left\{ \frac{y_{(1)}^{\max}(\mathbf{y}_n, \mathbf{v}_n) - y_{(1)}^{\min}(\mathbf{y}_n, \mathbf{v}_n)}{2\bar{v}} \right\} = \Delta_{(1)}.$$

Therefore, even when $r = 1$, the controllers designed according to the ∞-norm-based and to the 2-norm-based notions of distance do not, in general, coincide. Instead, in the linear case, the following relationships hold:

$$\begin{aligned} \Delta_{(r)} &= |e_0| + |e_1| + \cdots + |e_{r-1}|, \\ \delta_{(1)} &= |e_0|. \end{aligned} \quad (31)$$

From equation (31), it is apparent that, if $r = 1$, $\Delta_{(1)}$ and $\delta_{(1)}$ coincides (which is the result stated in Corollary 1). □

4. *Conclusions*

In this paper, the limits of performances, which are inherently imposed by the choice of a reference model having relative degree strictly smaller than the relative degree of the plant, have been analyzed and discussed.

It has been shown that the notion of best achievable performance strictly depends on the notion of distance between systems, since, in the case of "mismatch" between the relative degrees of the plant and the reference model, no exact model-matching can in general be achieved.

In particular, we have shown that a crucial role for the determination of the limits of performances is played by the notion of best r-steps-ahead predictor for the reference model. The notion of best (according to the ∞-norm-based distance) predictor have been developed both in the linear and in the nonlinear case in a deterministic setting.

Acknowledgements—These result were obtained during a visit of S. M. Savaresi at the University of Twente, supported by the HCM Network: "*Nonlinear and Adaptive Control*: *towards a Design Methodology for Physical Systems*".

References

Allgöwer, F. (1995). Ein Werkzeug zur Analyse und Synthese nichtlinearer Regelkreise. In *Entwurf nichtlinearer Regelungen*, S. Engell (Ed.), Oldenbourg, München.

Doyle J. C, B. A. Francis and A. R. Tannebaum (1992). *Feedback Control Theory*. Macmillan, New York.

Kotta, Ü. (1995). *Inversion Method in the Discrete-Time Nonlinear Control Systems Synthesis Problems*, Lecture notes in Control and Information Sciences, Vol. 205. Springer, Berlin.

Kotta, Ü. and H. Nijmeijer (1994). On dynamic input–output linearization of discrete-time nonlinear systems. *Int. J. Control*, **60**(6), 1319–1337.

Ljung L. (1987). *System Identification*: *Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ.

Narendra, K. S. and K. Parthasarathy (1990). Identification and control of dynamical systems using neural networks. *IEEE Trans. Neural Networks*, **1**(1), pp. 4–27.

Nijmeijer, H. and A. van der Schaft (1990). *Nonlinear Dynamical Control Systems*. Springer, New York.

Savaresi S. M. (1997). Parametric identification and control of nonlinear systems. Ph.D. Thesis, Politecnico di Milano, Italy. (in Italian).

Sjöberg, J. (1995). Non-linear system identification with neural networks, Linköping University, Sweden. Ph.D. Thesis.