PERGAMON

Automatica 38 (2002) 1539–1547

automatica

www.elsevier.com/locate/automatica

Brief Paper

Non-asymptotic confidence ellipsoids for the least-squares estimate☆

Erik Weyer[a,*], M.C. Campi[b]

[a] Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, VIC. 3010, Australia
[b] Department of Electrical Engineering and Automation, University of Brescia, Via Branze 38, 25123 Brescia, Italy

Received 24 November 2000; received in revised form 7 January 2002; accepted 16 January 2002

Abstract

In this paper, we consider the finite sample properties of least-squares system identification, and derive *non-asymptotic* confidence ellipsoids for the estimate. The shape of the confidence ellipsoids is similar to the shape of the ellipsoids derived using asymptotic theory, but unlike asymptotic theory, they are valid for a finite number of data points. The probability that the estimate belongs to a certain ellipsoid has a natural dependence on the volume of the ellipsoid, the data generating mechanism, the model order and the number of data points available. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: System identification; Least squares; Confidence ellipsoids; Finite sample properties

1. Introduction

In this paper, we consider the properties of least-squares system identification when only a finite number of data points are available. The asymptotic properties of least-squares identification are well understood, see e.g. Ljung (1999) or Söderström and Stoica (1989), but it is only recently that results addressing the finite sample properties have started appearing, e.g. Weyer, Williamson, and Mareels (1999), Weyer and Campi (1999), Campi and Weyer (2002), and Weyer (2000).

In applications such as evaluation of model uncertainty it is common to use the asymptotic confidence regions for the parameter estimate, even when only a finite number of data points are available. In this paper, we derive non-asymptotic confidence ellipsoids for the least-squares estimate. It is shown that the confidence ellipsoids depend in a natural way on factors such as the model and system order, the pole locations and the number of data points available.

The main tool we make use of is exponential inequalities in order to bound differences between expected values and empirical values. Earlier, using different techniques, Spall (1995) has considered uncertainty bounds for general M-estimators for a finite number of data points. His results are, however, difficult to use in the situation we consider here.

It should also be mentioned that finite sample properties have been studied in the deterministic set membership and worst case identification settings. In this context, the identification algorithms deliver all models which are in agreement with the observed data, so that finite-sample results are automatically included in the identification result. Differently from these settings, the present paper concentrates on the standard least-squares identification method in a stochastic framework.

The paper is organised as follows. In the next section we introduce the identification setting. In Section 3 we first bound the difference between the expected and empirical values of the matrix and vector which make up the normal equation (Theorem 1), and then we use this result to derive the non-asymptotic confidence ellipsoids (Theorem 2). Technical results are given in the appendices.

2. Identification setting

2.1. The data generation mechanism

We assume that the observed data are generated by a linear system

$$y(t) = G_0(q^{-1})u(t) + H_0(q^{-1})e(t), \tag{1}$$

☆ This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Antonio Vicino under the direction of Editor Torsten Söderström.

* Corresponding author. Tel.: +61-3-8344-9726; fax: +61-3-8344-6678.

E-mail address: e.weyer@ee.mu.oz.au (E. Weyer).

PII: S0005-1098(02)00064-X

where the input signal $u(t)$ is stochastic and generated by

$$u(t) = V_0(q^{-1})w(t), \quad (2)$$

where $G_0(q^{-1})$, $H_0(q^{-1})$ and $V_0(q^{-1})$ are transfer functions in the backward shift operator q^{-1}, i.e. $q^{-1}y(t) = y(t-1)$; however, for the sake of readability, we omit throughout the dependence on q^{-1}. $w(t)$ and $e(t)$ are sequences of independent Gaussian random variables, independent of each other, with zero mean and variance σ_w^2 and σ_e^2, respectively. The assumptions on $u(t)$ and $e(t)$ are not crucial. For example, the Gaussian assumption is only used to establish Cramer's condition (B.1) in one of the technical lemmas, and the results can easily be extended to other types of independent noise sequences, and we can also allow for a deterministic input signal.

We assume that G_0, H_0 and V_0 can be written as

$$G_0 = \frac{B_0}{A_0}, \quad H_0 = \frac{C_0}{D_0}, \quad V_0 = \frac{R_0}{S_0},$$

where

$$A_0 = 1 + a_{01}q^{-1} + \cdots + a_{0n_0}q^{-n_0},$$
$$B_0 = b_{01}q^{-1} + \cdots + b_{0n_0}q^{-n_0},$$
$$C_0 = 1 + c_{01}q^{-1} + \cdots + c_{0n_0}q^{-n_0},$$
$$D_0 = 1 + d_{01}q^{-1} + \cdots + d_{0n_0}q^{-n_0},$$
$$R_0 = 1 + r_{01}q^{-1} + \cdots + r_{0n_1}q^{-n_1},$$
$$S_0 = 1 + s_{01}q^{-1} + \cdots + s_{0n_1}q^{-n_1},$$

and n_0 and n_1 are upper bounds on the degrees. Moreover, we assume that the zeros of A_0, C_0, D_0, R_0 and S_0 are inside a circle of a known radius $\eta < 1$, i.e. we assume stability of the system with a known margin, and also that the transfer function between the noise sequence and the output has a stable inverse with the same stability margin. The zeros of B_0 are assumed to be inside a circle of known radius μ, where μ might be larger than 1, i.e. we allow for non-minimum phase zeros in the transfer function between $u(t)$ and $y(t)$, and finally we assume that $|b_{01}|$ is bounded by a known constant B. For simplicity we assume that $B \geqslant 1$.

2.2. Model class

The model class considered is

$$A(q^{-1})y(t) = B(q^{-1})u(t) + v(t), \quad (3)$$

where $v(t)$ is a disturbance and

$$A(\theta) = 1 + a_1q^{-1} + \cdots + a_nq^{-n},$$
$$B(\theta) = b_1q^{-1} + \cdots + b_nq^{-n}.$$

Eq. (3) can be written in linear regression form $y(t,\theta) = \phi^{\mathrm{T}}(t)\theta + v(t)$ by introducing

$$\phi(t) = [-y(t-1),\ldots,-y(t-n),u(t-1),\ldots,u(t-n)]^{\mathrm{T}},$$
$$\theta = [a_1,\ldots,a_n,b_1,\ldots,b_n]^{\mathrm{T}}.$$

Note that system (1) does not need to belong to the model class.

2.3. The identification criterion

From a system identification perspective, the most important feature of the above model is its associated one step ahead predictor which is given by $\hat{y}(t,\theta) = \phi^{\mathrm{T}}(t)\theta$, and the corresponding prediction error is $\varepsilon(t,\theta) = y(t) - \hat{y}(t,\theta)$. Ideally, one would like to choose θ such that the following theoretical identification cost:

$$V(\theta) = E\varepsilon^2(t,\theta) \quad (4)$$

is minimised where E is the expectation operator. The value of θ which minimises (4) is given by

$$\theta^* = R^{-1}f, \quad (5)$$

where

$$R = E\phi(t)\phi^{\mathrm{T}}(t), \quad f = E\phi(t)y(t). \quad (6)$$

Since the data generation mechanism is unknown, one cannot compute expected value (4) and estimate (5). Instead the empirical version

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\varepsilon^2(t,\theta) \quad (7)$$

is used, and the corresponding estimate is the well-known least-squares estimate

$$\hat{\theta}_N = R_N^{-1}f_N, \quad (8)$$

where

$$R_N = \frac{1}{N}\sum_{t=1}^{N}\phi(t)\phi^{\mathrm{T}}(t), \quad f_N = \frac{1}{N}\sum_{t=1}^{N}\phi(t)y(t). \quad (9)$$

Clearly, $\hat{\theta}_N$ can only be expected to be close to θ^* when the number of data points tends to infinity, and this is indeed the case under mild assumptions, see, e.g. Ljung (1999). However, we never have an infinite number of data points, and a question that arises naturally is to quantify the difference between $\hat{\theta}_N$ and θ^* for a finite N.

When the system does not belong to the model class, θ^* represents a model which minimises the expected value of the squared prediction error. The total prediction error can then be split in a bias error $y(t) - \hat{y}(t,\theta^*)$ which is due to the fact that the model class is not rich enough, and a variance error $\hat{y}(t,\theta^*) - \hat{y}(t,\hat{\theta})$ due to the variability of the estimate. In the case of undermodelling our results bounds how far our estimate $\hat{\theta}_N$ is from the best possible θ^*.

3. The main result

In this section, we present the ellipsoidal confidence regions for the least-squares estimate. First we bound the probability that the differences $R_N - R$ and $f_N - f$ exceed a certain value (Theorem 1), and then we use these results to bound $\hat{\theta}_N - \theta^*$ (Theorem 2).

It is important to put our results in the right perspective. We have not made any attempt of optimising the bounds, and in some places we have made the bounds more conservative in order to get relatively simple expressions. The bounds are therefore looser than they need to be. The significance of the bounds is that they illustrate how the confidence ellipsoids depend on important variables and they show that in principle we can derive confidence ellipsoids for a finite number of data points. Certainly, more work is expected in the direction of deriving tighter bounds and, in our conclusions, we provide some ideas on how to pursue this. At the present stage of knowledge, we recognise the results of this paper to be a first step in a new—technically very challenging but of great importance—direction of research in system identification: characterising the finite sample properties of system identification methods.

Theorem 1. *Let $[R_N - R]_{k,l}$ denote the (k,l) element of $R_N - R$, and $[f_N - f]_k$ the kth element of $f_N - f$. Assume that the data has been generated according to* (1) *and* (2). *Let R and f be as in* (6) *and R_N and f_N as in* (9). *Then*

$$Pr\left\{ \max_{k,l\in\{1,\dots,2n\}} |[R_N - R]_{k,l}| < \varepsilon \text{ and } \max_{k\in\{1,\dots,2n\}} |[f_N - f]_k| < \varepsilon \right\} \geqslant 1 - \delta,$$

where

$$\delta = \begin{cases} \delta_1 + \delta_2 + \delta_3 & \text{if } \delta_1 + \delta_2 + \delta_3 < 1, \\ 1 & \text{otherwise,} \end{cases} \tag{10}$$

$$\delta_1 = \frac{2n(n+3)\exp(-N\varepsilon_{ww}^2/4\sigma_w^2(4\sigma_w^2 + \varepsilon_{ww}))}{(1 - \exp(-N\varepsilon_{ww}^2/4\sigma_w^2(4\sigma_w^2 + \varepsilon_{ww})))^2}, \tag{11}$$

$$\delta_2 = \frac{2n(n+1)\exp(-N\varepsilon_{ee}^2/4\sigma_e^2(4\sigma_e^2 + \varepsilon_{ee}))}{(1 - \exp(-N\varepsilon_{ee}^2/4\sigma_e^2(4\sigma_e^2 + \varepsilon_{ee})))^2}, \tag{12}$$

$$\delta_3 = \frac{2n(n+2)\exp(-N\varepsilon_{we}^2/4\sigma_w\sigma_e(4\sigma_w\sigma_e + \varepsilon_{we}))}{(1 - \exp(-N\varepsilon_{we}^2/4\sigma_w\sigma_e(4\sigma_w\sigma_e + \varepsilon_{we})))^2}, \tag{13}$$

$$\varepsilon_{ww} \leqslant \frac{\varepsilon(1-\eta)^{2n_0+2n_1+1}}{3\cdot 2^{2n_1}B^2(2(n_0+n_1)\eta + 3(1-\eta))}\left(\frac{\eta}{\eta+\mu}\right)^{2n_0-2}, \tag{14}$$

$$\varepsilon_{ee} \leqslant \frac{\varepsilon(1-\eta)^{2n_0+1}}{3\cdot 2^{n_0}(2n_0\eta + (1-\eta))}, \tag{15}$$

$$\varepsilon_{we} \leqslant \frac{\varepsilon(1-\eta)^{2n_0+n_1+1}}{3\cdot 2^{n_0+n_1+1}B((2n_0+n_1)\eta+2(1-\eta))}\left(\frac{\eta}{\eta+\mu}\right)^{n_0-1}. \tag{16}$$

Proof. See Appendix A.

The functional dependencies of δ are quite natural. In particular, δ tends to 1 when the bound on the pole positions $\eta \to 1$, and/or the system and model order $n, n_0, n_1 \to \infty$. This can be easily understood since under these conditions there will be a strong correlation between prediction errors far apart, and the probability that there is a large difference between the expected and empirical value increases. Also, as expected δ tends to zero as $N \to \infty$, but note that for small values of N and ε, δ may be equal to 1, in which case Theorem 1 does not yield any useful information.

We are now in the position that we can derive non-asymptotic confidence ellipsoids for the parameter estimate.

Theorem 2. *Assume that the data has been generated according to* (1) *and* (2). *Let $\hat{\theta}_N = R_N^{-1} f_N$ and $\theta^* = R^{-1} f$. If $R_N - 2n\varepsilon I$ is positive definite, then*

$$Pr\left\{ (\hat{\theta}_N - \theta^*)^{\mathrm{T}}(R_N - 2n\varepsilon I)(\hat{\theta}_N - \theta^*) \leqslant \frac{(\varepsilon\sqrt{2n}\|\hat{\theta}_N\| + \varepsilon)^2 2n}{\lambda_{\min}(R_N) - 2n\varepsilon} \right\} \geqslant 1 - \delta,$$

where δ is given in Theorem 1. *$\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue, and n is the model order.*

The shape of the non-asymptotic confidence ellipsoids is similar to those obtained from asymptotic theory under the assumption that the true system belongs to the model class. In the asymptotic case, the ellipsoids are given by $(\hat{\theta}_N - \theta^*)^{\mathrm{T}} R(\hat{\theta}_N - \theta^*)$ where the volumes and probabilities depends on factors such as the number of data points and the noise variance. As R is unknown it is in practice replaced by its sample mean R_N. The only difference between the ellipsoids is therefore that we subtract the diagonal matrix $2n\varepsilon I$ in the non-asymptotic case. Note, however, that even though the shape of the ellipsoids is similar, the probabilities we assign to ellipsoids with the same volume may be quite different in the asymptotic and finite sample case. The finite sample results are on the conservative side.

Proof. From Theorem 1 it follows that

$$(R_N + \tilde{R})\theta^* = f_N + \tilde{f}$$

with probability at least $1-\delta$ for some $\tilde{R}$ and $\tilde{f}$ satisfying

$$|[\tilde{R}]_{k,l}| \leqslant \varepsilon, \quad |[\tilde{f}]_k| \leqslant \varepsilon, \ \forall k,l \in \{1,\dots,2n\}.$$

It follows that

$$\begin{aligned}(\hat{\theta}_N-\theta^*)^{\mathrm{T}}R_N(\hat{\theta}_N-\theta^*) &= (\hat{\theta}_N-\theta^*)^{\mathrm{T}}(R_N\hat{\theta}_N-R_N\theta^*)\\ &=(\hat{\theta}_N-\theta^*)^{\mathrm{T}}(f_N-f_N-\tilde{f}+\tilde{R}\theta^*)\\ &=-(\hat{\theta}_N-\theta^*)^{\mathrm{T}}\tilde{R}(\hat{\theta}_N-\theta^*)+(\hat{\theta}_N-\theta^*)^{\mathrm{T}}(\tilde{R}\hat{\theta}_N-\tilde{f})\end{aligned}$$

and hence

$$(\hat{\theta}_N-\theta^*)^{\mathrm{T}}(R_N+\tilde{R})(\hat{\theta}_N-\theta^*)=(\hat{\theta}_N-\theta^*)^{\mathrm{T}}(\tilde{R}\hat{\theta}_N-\tilde{f}). \tag{17}$$

Since $\tilde{R}+2n\varepsilon I$ is positive definite it follows that

$$(\hat{\theta}_N-\theta^*)^{\mathrm{T}}(R_N+\tilde{R})(\hat{\theta}_N-\theta^*) \geqslant (\hat{\theta}_N-\theta^*)^{\mathrm{T}}(R_N-2n\varepsilon I)(\hat{\theta}_N-\theta^*). \tag{18}$$

Next we observe that

$$(\hat{\theta}_N-\theta^*)^{\mathrm{T}}(\tilde{R}\hat{\theta}_N-\tilde{f}) \leqslant \left(\sum_{i=1}^{2n}|\hat{\theta}_i-\theta_i^*|\right)\left(\sum_{j=1}^{2n}|\hat{\theta}_j|\varepsilon+\varepsilon\right).$$

From Schwarz's inequality it follows that $(\sum_{j=1}^{2n}|\hat{\theta}_j|)^2 \leqslant 2n\|\hat{\theta}_N\|^2$. Hence

$$(\hat{\theta}_N-\theta^*)^{\mathrm{T}}(\tilde{R}\hat{\theta}_N-\tilde{f}) \leqslant \sum_{i=1}^{2n}|\hat{\theta}_i-\theta_i^*|(\varepsilon\sqrt{2n}\|\hat{\theta}_N\|+\varepsilon). \tag{19}$$

Thus, by combining (17)–(19) it follows that

$$\begin{aligned}(\hat{\theta}_N-\theta^*)^{\mathrm{T}}(R_N-2n\varepsilon I)(\hat{\theta}_N-\theta^*) &\leqslant \sum_{i=1}^{2n}|\hat{\theta}_i-\theta_i^*|(\varepsilon\sqrt{2n}\|\hat{\theta}_N\|+\varepsilon)\\ &\leqslant \sqrt{2n}\|\hat{\theta}_N-\theta^*\|(\varepsilon\sqrt{2n}\|\hat{\theta}_N\|+\varepsilon).\end{aligned}$$

This implies that

$$\|\hat{\theta}_N-\theta^*\| \leqslant \frac{\sqrt{2n}(\varepsilon\sqrt{2n}\|\hat{\theta}_N\|+\varepsilon)}{\lambda_{\min}(R_N-2n\varepsilon I)}$$

and hence

$$(\hat{\theta}_N-\theta^*)^{\mathrm{T}}(R_N-2n\varepsilon I)(\hat{\theta}_N-\theta^*) \leqslant \frac{(\varepsilon\sqrt{2n}\|\hat{\theta}_N\|+\varepsilon)^2 2n}{\lambda_{\min}(R_N-2n\varepsilon I)}.$$

4. Concluding remarks

In this paper, we have derived non-asymptotic confidence ellipsoids for the least-squares estimate. The shape of the ellipsoids is similar to that obtained using asymptotic theory, although the probabilities we assign to the ellipsoids can be quite different. The probability that the estimate belongs to a certain ellipsoid has a natural dependence on the volume of the ellipsoid, the data generating mechanism, the model order and the number of data points available.

Our results are worst case in the sense they are valid for all plants such that the assumed conditions on the plant order, pole position, etc. are satisfied. For this reason, they cannot be expected to be tight. Currently, we are working towards refining these results so as to incorporate a posteriori information (provided by data) in model quality assessment. Though very interesting and promising, this study entails technical problems that are very challenging and not fully solved yet.

Appendix A. Proof of Theorem 1

The proof of Theorem 1 is organised as follows. In the next subsection, we introduce some notation. Then in Section A.2 we express the differences $|[R_N-R]_{k,l}|$ and $|[f_N-f]_k|$ in terms of the underlying stochastic processes $w(t)$ and $e(t)$ (Lemma 3). Then in Section A.3 we bound $|[R_N-R]_{k,l}|$ and $|[f_N-f]_k|$ assuming that the observed realisations of $w(t)$ and $e(t)$ satisfy certain inequalities (Lemma A.2). Finally in Section A.4 we bound the probabilities that the observed realisations of $w(t)$ and $e(t)$ satisfy the inequalities used bounding $|[R_N-R]_{k,l}|$ and $|[f_N-f]_k|$ (Lemma A.3). Theorem 1 follows by combining Lemmas A.2 and A.3.

A.1. Notation

Let

$$\begin{aligned}G_0(q^{-1})V_0(q^{-1}) &= g_1q^{-1}+g_2q^{-2}+\cdots,\\ H_0(q^{-1}) &= 1+h_1q^{-1}+h_2q^{-2}+\cdots,\\ V_0(q^{-1}) &= 1+v_1q^{-1}+v_2q^{-2}+\cdots\end{aligned}$$

and let

$$S_N^{ww}(k,l,i,j) = \frac{1}{N}\sum_{t=1}^{N}w(t-k-i)w(t-l-j)-\delta_{k+i-l-j}\sigma_w^2, \tag{A.1}$$

$$S_N^{ee}(k,l,i,j) = \frac{1}{N}\sum_{t=1}^{N}e(t-k-i)e(t-l-j)-\delta_{k+i-l-j}\sigma_e^2, \tag{A.2}$$

$$S_N^{we}(k,l,i,j) = \frac{1}{N}\sum_{t=1}^{N} w(t-k-i)e(t-l-j), \quad \text{(A.3)}$$

$$S_N^{ew}(k,l,i,j) = \frac{1}{N}\sum_{t=1}^{N} e(t-k-i)w(t-l-j), \quad \text{(A.4)}$$

where $\delta_{k+i-l-j} = 1$ if $k+i = l+j$ and 0 otherwise.

A.2. Expressions for the elements in $R_N - R$ and $f_N - f$

The elements of $R_N - R$ and $f_N - f$ are of the form

$$S_N^{yy}(k,l) = \frac{1}{N}\sum_{t=1}^{N} y(t-k)y(t-l) - Ey(t-k)y(t-l),$$

$k = 0,\dots,n, \quad l = 1,\dots,n,$

$$S_N^{yu}(k,l) = \frac{1}{N}\sum_{t=1}^{N} y(t-k)u(t-l) - Ey(t-k)u(t-l),$$

$k = 0,\dots,n, \quad l = 1,\dots,n,$

$$S_N^{uu}(k,l) = \frac{1}{N}\sum_{t=1}^{N} u(t-k)u(t-l) - Eu(t-k)u(t-l),$$

$k = 1,\dots,n, \quad l = 1,\dots,n.$

The following lemma provides bounds on the absolute values of these elements

Lemma A.1.

$$\begin{aligned}|S_N^{yy}(k,l)| \leqslant & \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} |g_i|\cdot|g_j|\cdot|S_N^{ww}(k,l,i,j)| \\ & + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} |h_i|\cdot|h_j|\cdot|S_N^{ee}(k,l,i,j)| \\ & + \sum_{i=1}^{\infty}\sum_{j=0}^{\infty} |g_i|\cdot|h_j|\cdot|S_N^{we}(k,l,i,j)| \\ & + \sum_{i=0}^{\infty}\sum_{j=1}^{\infty} |h_i|\cdot|g_j|\cdot|S_N^{ew}(k,l,i,j)|, \end{aligned} \quad \text{(A.5)}$$

$$\begin{aligned}|S_N^{yu}(k,l)| \leqslant & \sum_{i=1}^{\infty}\sum_{j=0}^{\infty} |g_i|\cdot|v_j|\cdot|S_N^{ww}(k,l,i,j)| \\ & + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} |h_i|\cdot|v_j|\cdot|S_N^{ew}(k,l,i,j)|, \end{aligned} \quad \text{(A.6)}$$

$$|S_N^{uu}(k,l)| \leqslant \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} |v_i|\cdot|v_j|\cdot|S_N^{ww}(k,l,i,j)| \quad \text{(A.7)}$$

with

$$|g_i| \leqslant 2^{n_1} B\left(1+\frac{\mu}{\eta}\right)^{n_0-1} \frac{i\cdots(i+n_0+n_1-2)}{(n_0+n_1-1)!}\eta^{i-1}, \quad \text{(A.8)}$$

$$|h_i| \leqslant 2^{n_0}\frac{(i+1)\cdots(i+n_0-1)}{(n_0-1)!}\eta^i, \quad \text{(A.9)}$$

$$|v_i| \leqslant 2^{n_1}\frac{(i+1)\cdots(i+n_1-1)}{(n_1-1)!}\eta^i. \quad \text{(A.10)}$$

Proof. Using (1) and (2) we find that

$$\begin{aligned} & S_N^{yy}(k,l) \\ & = \frac{1}{N}\sum_{t=1}^{N} y(t-k)y(t-l) - Ey(t-k)y(t-l) \\ & = \frac{1}{N}\sum_{t=1}^{N} (G_0V_0w(t-k) + H_0e(t-k))(G_0V_0w(t-l) \\ & \quad + H_0e(t-l)) - E(G_0V_0w(t-k) + H_0e(t-k)) \\ & \quad (G_0V_0w(t-l) + H_0e(t-l)) \\ & = \frac{1}{N}\sum_{t=1}^{N}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} g_ig_j[w(t-k-i)w(t-l-j) \\ & \quad - \delta_{k+i-l-j}\sigma_w^2] \\ & \quad + \frac{1}{N}\sum_{t=1}^{N}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} h_ih_j[e(t-k-i)e(t-l-j) \\ & \quad - \delta_{k+i-l-j}\sigma_e^2] \\ & \quad + \frac{1}{N}\sum_{t=1}^{N}\sum_{i=1}^{\infty}\sum_{j=0}^{\infty} g_ih_jw(t-k-i)e(t-l-j) \\ & \quad + \frac{1}{N}\sum_{t=1}^{N}\sum_{i=0}^{\infty}\sum_{j=1}^{\infty} h_ig_je(t-k-i)w(t-l-j) \end{aligned} \quad \text{(A.11)}$$

and (A.5) follows by changing the order of summation, taking absolute values and using (A.1)–(A.4).

The bounds on $S_N^{yu}(k,l)$ and $S_N^{uu}(k,l)$ follow similarly. The bounds on $|g_i|$, $|h_i|$ and $|v_i|$ are given by Corollary C.2 in Appendix C. □

A.3. Bound on $R_N - R$ and $f_N - f$

Lemma A.2. *Assume that* $\forall i \geqslant 0$, $\forall j \geqslant 0$, $\forall k \in \{0,\dots,n\}$, $\forall l \in \{1,\dots,n\}$

$$S_N^{ww}(k,l,i,j) \leqslant \varepsilon_{ww}(i+j+1), \quad \text{(A.12)}$$

$$S_N^{ee}(k,l,i,j) \leqslant \varepsilon_{ee}(i+j+1), \tag{A.13}$$

$$S_N^{we}(k,l,i,j) \leqslant \varepsilon_{we}(i+j+1), \tag{A.14}$$

$$S_N^{ew}(k,l,i,j) \leqslant \varepsilon_{we}(i+j+1), \tag{A.15}$$

where ε_{ww}, ε_{ee} *and* ε_{we} *are given by* (14)–(16). *Then*

$$\max_{k,l\in\{1,\dots,2n\}} |[R_N - R]_{k,l}| < \varepsilon \quad \textit{and}$$

$$\max_{k\in\{1,\dots,2n\}} |[f_N - f]_k| < \varepsilon.$$

Proof. First we consider the elements of $R_N - R$ and $f_N - f$ of the form $S_N^{yy}(k,l)$. We bound the four terms of the right-hand side of (A.5) separately. For the first term we get

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} |g_i|\cdot|g_j|\cdot|S_N^{ee}(k,l,i,j)| \leqslant 2^{2n_1}B^2\left(1+\frac{\mu}{\eta}\right)^{2n_0-2}$$

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{i\cdots(i+n_0+n_1-2)}{(n_0+n_1-1)!}\eta^{i-1}$$

$$\frac{j\cdots(j+n_0+n_1-2)}{(n_0+n_1-1)!}\eta^{j-1}\varepsilon_{ww}(i+j+1). \tag{A.16}$$

Introducing the variables $\tilde{i} = i-1$ and $\tilde{j} = j-1$ the last expression can be written as

$$2^{2n_1}B^2\left(1+\frac{\mu}{\eta}\right)^{2n_0-2}\varepsilon_{ww}\left[\sum_{\tilde{i}=0}^{\infty}\frac{\tilde{i}\cdots(\tilde{i}+n_0+n_1-1)}{(n_0+n_1-1)!}\eta^{\tilde{i}}\right.$$

$$\sum_{\tilde{j}=0}^{\infty}\frac{(\tilde{j}+1)\cdots(\tilde{j}+n_0+n_1-1)}{(n_0+n_1-1)!}\eta^{\tilde{j}}$$

$$+\sum_{\tilde{i}=0}^{\infty}\frac{(\tilde{i}+1)\cdots(\tilde{i}+n_0+n_1-1)}{(n_0+n_1-1)!}\eta^{\tilde{i}}$$

$$\sum_{\tilde{j}=0}^{\infty}\frac{\tilde{j}\cdots(\tilde{j}+n_0+n_1-1)}{(n_0+n_1-1)!}\eta^{\tilde{j}}$$

$$+3\sum_{\tilde{i}=0}^{\infty}\frac{(\tilde{i}+1)\cdots(\tilde{i}+n_0+n_1-1)}{(n_0+n_1-1)!}\eta^{\tilde{i}}$$

$$\left.\sum_{\tilde{j}=0}^{\infty}\frac{(\tilde{j}+1)\cdots(\tilde{j}+n_0+n_1-1)}{(n_0+n_1-1)!}\eta^{\tilde{j}}\right]$$

$$\leqslant 2^{2n_1}B^2\left(1+\frac{\mu}{\eta}\right)^{2n_0-2}$$

$$\frac{2(n_0+n_1)\eta+3(1-\eta)}{(1-\eta)^{2n_0+2n_1+1}}\varepsilon_{ww}$$

where we have used

$$\sum_{k=0}^{\infty}\frac{(k+1)\cdots(k+n-1)}{(n-1)!}\eta^k = \frac{1}{(1-\eta)^n}$$

and

$$\sum_{k=0}^{\infty}\frac{k(k+1)\cdots(k+n-1)}{(n-1)!}\eta^k = \frac{n\eta}{(1-\eta)^{n+1}}.$$

For the second term we get

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}|h_i|\cdot|h_j|\cdot|S_N^{ee}(k,l,i,j)|$$

$$\leqslant 2^{2n_0}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{(i+1)\cdots(i+n_0-1)}{(n_0-1)!}\eta^k$$

$$\frac{(j+1)\cdots(j+n_0-1)}{(n_0-1)!}\eta^l\varepsilon_{ee}(i+j+1)$$

$$\leqslant 2^{2n_0}\frac{2n_0\eta+(1-\eta)}{(1-\eta)^{2n_0+1}}\varepsilon_{ee}.$$

The bounds for the third and fourth term are identical and we get

$$\sum_{i=1}^{\infty}\sum_{j=0}^{\infty}|g_i|\cdot|h_j|\cdot|S_N^{we}(k,l,i,j)|$$

$$\leqslant 2^{n_0+n_1}B\left(1+\frac{\mu}{\eta}\right)^{n_0-1}$$

$$\sum_{i=1}^{\infty}\sum_{j=0}^{\infty}\frac{i\cdots(i+n_0+n_1-2)}{(n_0+n_1-1)!}\eta^{i-1}$$

$$\frac{(j+1)\cdots(j+n_0-1)}{(n_0-1)!}\eta^j\varepsilon_{we}(i+j+1)$$

$$\leqslant 2^{n_0+n_1}\left(1+\frac{\mu}{\eta}\right)^{n_0-1}B\frac{(2n_0+n_1)\eta+2(1-\eta)}{(1-\eta)^{2n_0+n_1+1}}\varepsilon_{we}.$$

Summing the above expressions and taking into account the bounds on ε_{ww}, ε_{ee} and ε_{we} given by (14)–(16) we find that

$$|S_N^{yy}(k,l)| < \varepsilon, \quad k=0,\dots,n, \quad l=1,\dots,n.$$

Next we turn our attention to the elements of the form $S_N^{yu}(k,l)$. Bounding the two terms on the right-hand side of (A.6) separately we get

$$\sum_{i=1}^{\infty}\sum_{j=0}^{\infty}|g_i|\cdot|v_j|\cdot|S_N^{ww}(k,l,i,j)|$$

$$\leqslant 2^{2n_1}B\left(1+\frac{\mu}{\eta}\right)^{n_0-1}\sum_{i=1}^{\infty}\sum_{j=0}^{\infty}\frac{i\cdots(i+n_0+n_1-2)}{(n_0+n_1-1)!}$$

$$\eta^{i-1}\frac{(j+1)\cdots(j+n_1-1)}{(n_1-1)!}\eta^j\varepsilon_{ww}(i+j+1)$$

$$\leqslant 2^{2n_1}B\left(1+\frac{\mu}{\eta}\right)^{n_0-1}\frac{(n_0+2n_1)\eta+2(1-\eta)}{(1-\eta)^{n_0+2n_1+1}}\varepsilon_{ww}$$

and

$$\sum_{i=1}^{\infty}\sum_{j=0}^{\infty}|h_i|\cdot|v_j|\cdot|S_N^{ew}(k,l,i,j)|$$

$$\leqslant 2^{n_0+n_1}\sum_{i=1}^{\infty}\sum_{j=0}^{\infty}$$

$$\frac{(i+1)\cdots(i+n_0-1)}{(n_0-1)!}\eta^i\frac{(j+1)\cdots(j+n_1-1)}{(n_1-1)!}$$

$$\eta^j\varepsilon_{we}(i+j+1)$$

$$\leqslant 2^{n_0+n_1}\frac{(n_0+n_1)\eta+(1-\eta)}{(1-\eta)^{n_0+n_1+1}}\varepsilon_{we}.$$

Using the bounds on ε_{ww} and ε_{we} noting that $B \geqslant 1$ by assumption, we find that

$$|S_N^{yu}(k,l)| < \varepsilon, \quad k=0,\ldots,n, \quad l=1,\ldots,n.$$

Finally we consider the elements of the form $S_N^{uu}(k,l)$. From (A.7) we have

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}|v_i|\cdot|v_j|\cdot|S_N^{ww}(k,l,i,j)|$$

$$\leqslant 2^{2n_1}\sum_{i=1}^{\infty}\sum_{j=0}^{\infty}\frac{(i+1)\cdots(i+n_1-1)}{(n_1-1)!}\eta^i$$

$$\frac{(j+1)\cdots(j+n_1-1)}{(n_1-1)!}\eta^j\varepsilon_{ww}(i+j+1)$$

$$\leqslant 2^{2n_1}\frac{2n_1\eta+(1-\eta)}{(1-\eta)^{2n_1+1}}\varepsilon_{ww}.$$

Using the bound ε_{ww} we find that

$$|S_N^{uu}(k,l)| < \varepsilon, \quad k=1,\ldots,n, \quad l=1,\ldots,n$$

and the proof is completed. Note that $|S_N^{uu}(k,l)|$ and $|S_N^{yu}(k,l)|$ will usually be much smaller than ε since ε_{we} and ε_{ww} are chosen to make $|S_N^{yy}(k,l)| < \varepsilon$. □

A.4. Bounds on probabilities

In this section, we calculate lower bounds for the probabilities that

$$S_N^{ww}(k,l,i,j) \leqslant \varepsilon_{ww}(i+j+1), \tag{A.17}$$

$$S_N^{ee}(k,l,i,j) \leqslant \varepsilon_{ee}(i+j+1), \tag{A.18}$$

$$S_N^{we}(k,l,i,j) \leqslant \varepsilon_{we}(i+j+1), \tag{A.19}$$

$$S_N^{ew}(k,l,i,j) \leqslant \varepsilon_{we}(i+j+1) \tag{A.20}$$

uniformly $\forall i \geqslant 0,\ \forall j \geqslant 0,\ \forall k \in \{0,\ldots,n\},\ \forall l \in \{1,\ldots,n\}$.

Lemma A.3. *Let $S_0=\{0,\ldots,n\}$ and $S_1=\{1,\ldots,n\}$*

$$Pr\{|S_N^{ww}(k,l,i,j)| \leqslant \varepsilon_{ww}(i+j+1)\ \forall k\in S_0,\ \forall l \in S_1\ \forall i,j \geqslant 0\} \geqslant 1-\delta_1, \tag{A.21}$$

$$Pr\{|S_N^{ee}(k,l,i,j)| \leqslant \varepsilon_{ee}(i+j+1)\ \forall k,l\in S_1\ \forall i,j \geqslant 0\} \geqslant 1-\delta_2, \tag{A.22}$$

$$Pr\{|S_N^{we}(k,l,i,j)| \leqslant \varepsilon_{we}(i+j+1)\ \forall k\in S_0,\ \forall l \in S_1\ \forall i,j \geqslant 0,\ |S_N^{ew}(k,l,i,j)| \leqslant \varepsilon_{we}(i+j+1)\ \forall k\in S_0,\ \forall l \in S_1\ \forall i,j \geqslant 0\} \geqslant 1-\delta_3, \tag{A.23}$$

where

$$\delta_1 = \frac{2n(n+3)\exp(-N\varepsilon_{ww}^2/4\sigma_w^2(4\sigma_w^2+\varepsilon_{ww}))}{(1-\exp(-N\varepsilon_{ww}^2/4\sigma_w^2(4\sigma_w^2+\varepsilon_{ww})))^2},$$

$$\delta_2 = \frac{2n(n+1)\exp(-N\varepsilon_{ee}^2/4\sigma_e^2(4\sigma_e^2+\varepsilon_{ee}))}{(1-\exp(-N\varepsilon_{ee}^2/4\sigma_e^2(4\sigma_e^2+\varepsilon_{ee})))^2},$$

$$\delta_3 = \frac{2n(n+2)\exp(-N\varepsilon_{we}^2/4\sigma_w\sigma_e(4\sigma_w\sigma_e+\varepsilon_{we}))}{(1-\exp(-N\varepsilon_{we}^2/4\sigma_w\sigma_e(4\sigma_w\sigma_e+\varepsilon_{we})))^2}.$$

Proof. First we compute the probability that $|S_N^{ww}(k,l,i,j)| \leqslant \varepsilon_{ww}(i+j+1)$ for all $i \geqslant 0$ and $j \geqslant 0$ for fixed k and l. Inequalities (B.2) and (B.3) bound this probability for fixed, k, l, i and j. To get a uniform bound in i and j we sum over $i \geqslant 0$ and $j \geqslant 0$, and we find that uniform probability is at least $1-\delta$ where

$$\delta = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}4\exp\left(-\frac{N\varepsilon_{ww}^2(i+j+1)^2}{4\sigma_w^2(4\sigma_w^2+\varepsilon_{ww}(i+j+1))}\right)$$

$$= \sum_{m=0}^{\infty}(m+1)4\exp\left(-\frac{N\varepsilon_{ww}^2(m+1)^2}{4\sigma_w^2(4\sigma_w^2+\varepsilon_{ww}(m+1))}\right)$$

$$\leqslant 4\sum_{m=0}^{\infty}(m+1)\mathrm{e}^{-\gamma(m+1)},$$

where

$$\gamma = \frac{N\varepsilon_{ww}^2}{4\sigma_w^2(4\sigma_w^2+\varepsilon_{ww})}.$$

Using $\sum_{m=0}^{\infty}(m+c_1)a^{m+c_2} = (a^{c_2}/(1-a)^2)(1+(c_1-1)(1-a))$ it follows that

$$Pr\{|S_N^{ww}(k,l,i,j)| \leqslant \varepsilon_{ww}(i+j+1)\ \forall i \geqslant 0,\ \forall j \geqslant 0\}$$

$$\geqslant 1-\frac{4\mathrm{e}^{-\gamma}}{(1-\mathrm{e}^{-\gamma})^2}.$$

As $S_N^{ww}(k,l,i,j)=S_N^{ww}(l,k,i,j)$ there are only $n(n+3)/2$ different elements of this type for $k\in\{0,\ldots,n\}$, $l\in\{1,\ldots,n\}$, and hence we have that

$$Pr\{|S_N^{ww}(k,l,i,j)|\leqslant \varepsilon_{ww}(i+j+1)\ \forall k\in S_0,\ \forall l\in S_1,$$
$$\forall i\geqslant 0,\ \forall j\geqslant 0\}$$
$$\geqslant 1-\frac{2n(n+3)\mathrm{e}^{-\gamma}}{(1-\mathrm{e}^{-\gamma})^2}=1-\delta_1. \tag{A.24}$$

The proof for the bound on the uniform probability of $|S_N^{ee}(k,li,j)|\leqslant \varepsilon_{ee}(i+j+1)$ follows in exactly the same manner. The only difference being that we have only $n(n+1)/2$ different elements since $S_N^{ee}(k,l,i,j)$ do not occur in the expression for the elements of f_N-f.

Next we consider the probability that $|S_N^{we}(k,l,i,j)|\leqslant \varepsilon_{we}(i+j+1)$ and $|S_N^{ew}(k,l,i,j)|\leqslant \varepsilon_{we}(i+j+1)$. Following the same approach as above and using (B.4) we find that

$$Pr\{|S_N^{we}(k,l,i,j)|\leqslant \varepsilon_{we}(i+j+1)\ \forall i\geqslant 0,\forall j\geqslant 0\}$$
$$\geqslant 1-\frac{2\mathrm{e}^{-\gamma}}{(1-\mathrm{e}^{-\gamma})^2},$$

where

$$\gamma=\frac{N\varepsilon_{we}^2}{4\sigma_w\sigma_e(4\sigma_w\sigma_e+\varepsilon_{we})}.$$

We obtain the same bound for $|S_N^{ew}(k,l,i,j)|$. Next we observe that $S_N^{we}(k,l,i,j)=S_N^{ew}(l,k,j,i)$ and hence there are in total n^2+2n different elements of this type as k ranges over 0 to n and l ranges over 1 to n. □

Appendix B. Exponential inequalities

The main theorem we are going to make use of is the following one taken from Bosq (1998, Theorem 1.2).

Theorem B.1 (Bernstein's inequality). *Let $X_1,\ldots,X_N$ be independent zero mean real-valued random variables and let $S_N=\sum_{t=1}^N X_t$. Assume there exists a $c>0$ such that*

$$E|X_t|^p\leqslant c^{p-2}p!EX_t^2<\infty,\quad t=1,\ldots,N,\ p=3,4,\ldots . \tag{B.1}$$

Then,

$$Pr\{|S_N|\geqslant\varepsilon\}\leqslant 2\exp\left(-\frac{\varepsilon^2}{4\sum_{t=1}^N EX_t^2+2c\varepsilon}\right).$$

Based on Theorem B.1, we have the following corollary

Corollary B.2. *Let $w(t)$ and $e(t)$, $t\in\mathbb{Z}$ be zero mean iid Gaussian variables independent of each other with variance σ_w^2 and σ_e^2, respectively. Then*

$$Pr\left\{\frac{1}{N}\left|\sum_{t=1}^N e^2(t-k)-\sigma^2\right|\geqslant\varepsilon(k,k)\right\}$$
$$\leqslant 2\exp\left(-\frac{N\varepsilon^2(k,k)}{4\sigma_e^2(2\sigma_e^2+\varepsilon(k,k))}\right), \tag{B.2}$$

$$Pr\left\{\frac{1}{N}\left|\sum_{t=1}^N e(t-k)e(t-l)\right|\geqslant\varepsilon(k,l)\right\}$$
$$\leqslant 4\exp\left(-\frac{N\varepsilon^2(k,l)}{4\sigma_e^2(4\sigma_e^2+\varepsilon(k,l))}\right),\quad k\neq l, \tag{B.3}$$

$$Pr\left\{\frac{1}{N}\left|\sum_{t=1}^N e(t-k)w(t-l)\right|\geqslant\varepsilon(k,l)\right\}$$
$$\leqslant 2\exp\left(-\frac{N\varepsilon^2(k,l)}{4\sigma_e\sigma_w(4\sigma_e\sigma_w+\varepsilon(k,l))}\right). \tag{B.4}$$

Proof. See Campi and Weyer (2002). □

Appendix C. Bounds on coefficients

In this appendix, we first present a general result bounding the magnitude of the coefficients of certain polynomials in terms of the locations of the zeros and the order of the polynomials. Then we use this result to bound the magnitude of the coefficients of $G_0(q^{-1})V_0(q^{-1})$, $H_0(q^{-1})$ and $V_0(q^{-1})$.

Lemma C.1. *Let*

$$M(q^{-1})=1+m_1q^{-1}+\cdots m_{n_m}q^{-n_m},$$
$$P(q^{-1})=1+p_1q^{-1}+\cdots p_{n_p}q^{-n_p}$$

be polynomials with all zeros inside a circle of radius $\eta<1$. Furthermore, let

$$W(q^{-1})=w_1q^{-1}+\cdots+w_{n_w}q^{-n_w}$$

be a polynomial with all zeros inside a circle of radius μ and leading coefficient bounded by $|w_1|<B$. Then the coefficients of the polynomials

$$M^{-1}(q^{-1})=1+\bar{m}_1q^{-1}+\bar{m}_2q^{-2}+\cdots, \tag{C.1}$$

$$M^{-1}(q^{-1})P(q^{-1})=1+\bar{p}_1q^{-1}+\bar{p}_2q^{-2}+\cdots, \tag{C.2}$$

$$M^{-1}(q^{-1})P(q^{-1})W(q^{-1})=\bar{w}_1q^{-1}+\bar{w}_2q^{-2}+\cdots \tag{C.3}$$

are bounded by

$$|\bar{m}_k|\leqslant\frac{(k+1)\cdots(k+n_m-1)}{(n_m-1)!}\eta^k, \tag{C.4}$$

$$|\bar{p}_k|\leqslant 2^{n_p}\frac{(k+1)\cdots(k+n_m-1)}{(n_m-1)!}\eta^k, \tag{C.5}$$

$$|\bar{w}_k| \leqslant 2^{n_p} B \left(1+\frac{\mu}{\eta}\right)^{n_w-1} \frac{k\cdots(k+n_m-2)}{(n_m-1)!}\eta^{k-1}. \quad \text{(C.6)}$$

Proof. See Campi and Weyer (2002). □

Combining the bounds in the above Lemma we obtain

Corollary C.2. *The coefficients of*

$$G_0 V_0 = g_1 q^{-1} + g_2 q^{-2} + \cdots,$$

$$H_0 = 1 + h_1 q^{-1} + h_2 q^{-2} + \cdots,$$

$$V_0 = 1 + v_1 q^{-1} + v_2 q^{-2} + \cdots$$

are bounded by

$$|g_k| \leqslant 2^{n_1} B \left(1+\frac{\mu}{\eta}\right)^{n_0-1} \frac{k\cdots(k+n_0+n_1-2)}{(n_0+n_1-1)!}\eta^{k-1},$$

$$|h_k| \leqslant 2^{n_0} \frac{(k+1)\cdots(k+n_0-1)}{(n_0-1)!}\eta^k,$$

$$|v_k| \leqslant 2^{n_1} \frac{(k+1)\cdots(k+n_1-1)}{(n_1-1)!}\eta^k.$$

References

Bosq, D. (1998). *Nonparametric statistics for stochastic processes-estimation and prediction*, *Lecture Notes in Statistics* Vol. 110. Berlin: Springer.

Campi, M. C., & Weyer, E. (2002). Finite sample properties of system identification methods. IEEE Transactions on Automatic Control (to appear).

Ljung, L. (1999). *System identification-theory for the user* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Söderström, T., & Stoica, P. (1989). *System identification*. Englewood Cliffs, NJ: Prentice-Hall.

Spall, J. C. (1995). Uncertainty bounds for parameter identification with small sample sizes. *Proceedings of the 34th IEEE conference on decision and control, New Orleans, LA, US* (pp. 3504–3515).

Weyer, E. (2000). Finite sample properties of system identification of ARX models under mixing conditions. *Automatica*, *36*(9), 1291–1299.

Weyer, E., & Campi, M.C. (1999). Finite sample properties of system identification methods. *Proceedings of the 38th IEEE conference on decision and control*, Phoenix, AZ, USA, December 1999 (pp. 510–515).

Weyer, E., Williamson, R. C., & Mareels, I. M. Y. (1999). Finite sample properties of linear model identification. *IEEE Transactions on Automatic Control*, *44*(7), 1370–1383.

Erik Weyer received the Siv. Ing. degree in 1988 and Ph.D. in 1993, both from the Norwegian Institute of Technology, Trondheim Norway. During his Ph.D. studies he spent two years at the Australian National University.From 1994 to 1996 he was a Research Fellow at the University of Queensland, and since 1997 he has been with the Department of Electrical and Electronic Engineering, the University of Melbourne, where he is currently a Senior Research Fellow.His research interests are in the area of system identification and control.

Marco C. Campi was born in Tradate, Italy, in 1963. He received the Doctor degree in electronic engineering from the Politecnico di Milano, Milano, Italy, in 1988.From 1988 to 1989, he was a Research Assistant in the Department of Electrical Engineering, the Politecnico di Milano. From 1989 to 1992, he worked as a researcher at the Centro di Teoria dei Sistemi of the National Research Council in Milano. Since 1992, he has been with the University of Brescia, Italy, where he is currently Professor of Automatic Control.Dr. Campi is an Associate Editor of Automatica, Systems and Control Letters, and the European Journal of Control. Serves as Vice-Chair of the Technical Committee IFAC on Stochastic Systems (SS) and is a member of the Technical Committee IFAC on Modeling, Identification and Signal Processing (MISP). Moreover, he is a Distinguished Lecturer under the IEEE CSS Program.He has held visiting positions at many universities including the Australian National University, Canberra, Australia; the University of Illinois at Urbana-Champaign, USA; the Centre for Artificial Intelligence and Robotics, Bangalore, India; and the University of Melbourne, Australia.His current research interests include: system identification, control relevant system identification, adaptive and iterative control, learning theory, and stochastic systems.