Quantification of frequency domain error bounds with guaranteed confidence level in Prediction Error Identification †

X. Bombois[(1)], B.D.O. Anderson[(2)], M. Gevers[(3)]

[(1)] Delft Center for Systems and Control, Delft University of Technology

Mekelweg 2, 2628 CD Delft, The Netherlands

Tel: +31-15-2785150, Fax: +31-15-2786679, Email: X.J.A.Bombois@dcsc.tudelft.nl

[(2)] Research School of Information Sciences and Engineering

Australian National University, and National ICT Australia

[(3)] CESAME, Université Catholique de Louvain, Bâtiment Euler

B1348 Louvain la Neuve, Belgium

Abstract

This paper considers Prediction Error identification of linearly parametrized models in the situation where the system is in the model set. For such situation it is easy to construct a confidence ellipsoid in parameter space in which the true parameter lies with an a priori fixed probability level, α. Surprisingly perhaps, the construction of a corresponding uncertainty set in the frequency domain, to which the true system belongs with probability α, is still an open problem. We show in this paper how to construct such frequency domain uncertainty set with a probability level of at least α.

Keywords: Prediction Error Identification, error bounds, confidence region, identification for control.

1 Introduction

We consider Prediction Error (PE) identification of model structures $G(z, \theta)$ that are linear in the parameter vector $\theta \in \mathbf{R}^k$: $G(z, \theta) = \Lambda(z)\theta$, where $\Lambda(z)$ is a set of independent rational

†The European Commission is herewith acknowledged for its financial support in part to the research reported on in this paper. The support is provided via the Program Training and Mobility of Researchers (TMR) and Project System Identification (ERB FMRX CT98 0206) to the European Research Network System Identification (ERNSI). The authors also acknowledge the Belgian Programme on Inter-university Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. National ICT Australia is funded by the Australian Government's Department of Communications, Information Technology and the Arts and the Australian Research Council through *Backing Australia's Ability* and the ICT Centre of Excellence program. The scientific responsibility rests with its authors.

basis functions. The identification is performed using N input-output data collected on a "true" system $y(t) = G_0(z)u(t) + v(t)$, where $v(t)$ is additive (colored) Gaussian noise. We further assume that the true system can be parametrized exactly within the chosen model structure for some parameter vector θ_0: $G_0(z) = G(z, \theta_0)$. Our results also hold for more general model structures, modulo a linearization of these structures around the identified model.

Under the assumptions made above, a Prediction Error identification experiment based on N input-output data delivers an estimate $\hat{\theta}$ of the parameter vector θ_0, which is a Gaussian random variable[1] with mean θ_0 and a covariance matrix P_θ that can be estimated from the data: $\hat{\theta} \sim \mathcal{N}(\theta_0, P_\theta)$; see [6] for details. As a result, the random variable $(\hat{\theta} - \theta_0)^T P_\theta^{-1} (\hat{\theta} - \theta_0)$ has a $\chi^2(k)$ distribution, and one can therefore construct an ellipsoidal confidence region in parameter space to which the true θ_0 belongs with some a priori chosen probability level α. Specifically, this ellipsoidal region is defined as

$$U_\theta(\chi_k) = \{\theta \mid (\theta - \hat{\theta})^T P_\theta^{-1} (\theta - \hat{\theta}) \leq \chi_k\}, \tag{1}$$

where, for a desired probability level α, χ_k is obtained from the χ^2 distribution as the value for which $Pr(\chi^2(k) < \chi_k) \triangleq \alpha$.

In many applications, and in particular in *identification for control* applications, it is much more useful to characterize the uncertainty on the estimated model in the frequency domain (i.e. in the Nyquist plane) rather than in parametric domain. Thus, the object of this paper is to construct an uncertainty region $\mathcal{L}$ in the frequency domain, to which the true system $G_0(e^{j\omega})$ must belong with a probability level of at least α for all ω, using the parametric confidence region obtained by PE identification as a starting point.

This problem appears very simple. Indeed, by the linearity of the model structure, the estimated model $G(z, \hat{\theta})$ is also Gaussian with mean $G_0(z)$. If we represent the frequency response of the model $G(z, \theta)$ by a vector $g(e^{j\omega}, \theta)$ containing its real and imaginary parts

$$g(e^{j\omega}, \theta) \triangleq \begin{pmatrix} Re(G(e^{j\omega}, \theta)) \\ Im(G(e^{j\omega}, \theta)) \end{pmatrix} = \overbrace{\begin{pmatrix} Re(\Lambda(e^{j\omega})) \\ Im(\Lambda(e^{j\omega})) \end{pmatrix}}^{T(e^{j\omega})} \theta, \tag{2}$$

we then have

$$\hat{g}(e^{j\omega}) \triangleq g(e^{j\omega}, \hat{\theta}) \sim \mathcal{N}(g(e^{j\omega}, \theta_0), P(\omega)) \tag{3}$$

$$\text{with} \quad P(\omega) \triangleq T(e^{j\omega}) P_\theta T(e^{j\omega})^T \in \mathbf{R}^{2\times 2}. \tag{4}$$

Just as was done for the parameter estimate $\hat{\theta}$, the normal distribution of the 2-vector $g(e^{j\omega}, \hat{\theta})$ allows one to build elliptic confidence regions $U(\omega, \chi_2)$ at each frequency in the Nyquist plane, that are guaranteed to contain the true frequency response $g(e^{j\omega}, \theta_0)$ at a prescribed probability level α. The elliptic region at frequency ω is defined as

$$U(\omega, \chi_2) = \{g \in \mathbf{R}^2 \mid (g - \hat{g}(e^{j\omega}))^T P(\omega)^{-1} (g - \hat{g}(e^{j\omega})) \leq \chi_2\} \tag{5}$$

[1] The Gaussian assumption on $v(t)$ is not essential for the PE framework; it can be replaced by an assumption of quasistationarity on $v(t)$, in which case the estimated parameter vector $\hat{\theta}$ is asymptotically Gaussian.

where, for a desired probability level α, χ_2 is obtained from the χ^2 distribution as the value for which $Pr(\chi^2(2) < \chi_2) \stackrel{\Delta}{=} \alpha$.

Examples of the use of such frequency domain confidence regions, derived from the corresponding parametric ellipsoidal regions of the PE identification framework, abound in the literature (see e.g. [3, 1, 11, 4, **?**]). Such elliptic regions have also been used abundantly for model structures that are not linear in the parameters, using a first order approximation for the mapping from θ-space to $g(e^{j\omega})$-space (see e.g. [10, 9], or [5]-[7] for the special case where a Model Error Modeling approach is used). These frequency domain uncertainty regions can actually be computed by the System Identification Toolbox of Matlab[2] .

We show in this paper that such procedure may lead to misleading conclusions. Indeed, the construction of the elliptic uncertainty set $U(\omega, \chi_2)$, as described above, only guarantees that the true $g(e^{j\omega}, \theta_0)$ belongs to $U(\omega, \chi_2)$ **at frequency $\boldsymbol{\omega}$** with probability α. It does by no means imply that the whole frequency response $g(e^{j\omega}, \theta_0)$ belongs, with probability α, to the set $\mathcal{L}(\chi_2) \stackrel{\Delta}{=} \{g(e^{j\omega}, \theta) \mid g(e^{j\omega}, \theta) \in U(\omega, \chi_2)\ \forall\omega\ \}$ obtained by connecting the sets $U(\omega, \chi_2)$ for all ω. In fact, we shall illustrate later that $Pr(g(e^{j\omega}, \theta_0) \in U(\omega, \chi_2)\ \forall\omega)$ can be much smaller than α for model structures of reasonable complexity. Several authors who have used the elliptic frequency domain uncertainty sets $U(\omega, \chi_2)$ were aware of the fact that one could not claim to have an α-level confidence region for the whole frequency response by connecting together α-level confidence regions at each frequency. Some authors have also discussed the difficulty of constructing a frequency domain uncertainty set on the basis of the sets $U(\omega, \chi_2)$ to which the true frequency response $G_0(e^{j\omega})$ belongs with a prescribed confidence level α. Others have used the frequency domain ellipses $U(\omega, \chi_2)$ for robust control design without apparently fully realizing that the confidence level attached to each set, at a particular frequency, could not be extrapolated to the whole set obtained by 'gluing' these sets together.

The confidence ellipses $U(\omega, \chi_2)$ were apparently first used, in the context of 'linear in the parameter models', in [3]. In that paper, models of restricted complexity are handled, and the bias error is treated on an equal footing as the variance error using the stochastic embedding approach. Thus, the confidence ellipses take account of both bias and variance error on the estimated transfer function. The authors construct their $U(\omega, \chi_2)$ confidence ellipses using the χ^2 distribution with two degrees of freedom, and conclude that one can use these sets "to give confidence ellipses in the complex plane for the frequency response estimate $G(e^{j\omega}, \hat{\theta})$." Even though no specific statement is made about the confidence level obtained this way for the whole frequency response, that statement could certainly induce erroneous interpretations. In [1, 11] the uncertainty bounds obtained by the stochastic embedding approach of [3] were used for the design of a robust controller in an application of water pressure control in a pump. The design was based on an uncertainty set around the frequency function $G(e^{j\omega}, \hat{\theta})$ obtained from the confidence ellipses $U(\omega, \chi_2)$. A similar usage is made of these confidence ellipses in [**?**], where an identification for robust control procedure is proposed based on a mixed H_2/H_∞ approach. Such approach leads to a controller

[2]However, in the Matlab Identification Toolbox these uncertainty regions are represented in a Bode plot as uncertainty bands around the magnitude and the phase.

which could be thought to achieve the required performance with $G_0(z)$ with probability α. However, as shown in this paper, this probability is always smaller than α.

In [2, 4] uncertainty sets are estimated for linear-in-the-parameter models that take account of variance error, bias error, and initial condition effects. Confidence levels are given for these uncertainty sets at a particular frequency, but no statements are made about the confidence level for the set obtained by connecting all frequencies.

The difficulty of extrapolating the confidence level α obtained for the individual uncertainty sets $U(\omega, \chi_2)$ to a confidence level for the uncertainty region $\mathcal{L}(\chi_2)$ obtained by connecting these sets together was fully recognized by Tjärnström [9]. We quote from his paper: "It should be noted that the different regions are dependent through the estimate of a parametric model and it is thus not possible to say much about the simultaneous confidence degree.[3] This follows from the Bonferroni inequality [8]." This inequality gives a lower bound for the simultaneous confidence level as $\max(0, 1 - d(1 - \beta))$, where d is the number of individual confidence regions and β the probability level of one region. This lower bound thus decreases rapidly with the number of different frequency regions $U(\omega, \chi_2)$, and it is equal to zero if we construct an uncertainty band for the whole frequency range ($d = \infty$). To obtain an overall confidence region for the true $G_0(z)$ with a predefined confidence level, Tjärnström proposes a method based on the bootstrap technique developed in statistics in the seventies. In a nutshell, it consists of estimating the probability distribution of the prediction errors, and then generating a large number of simulated input-output data sets from the known inputs and residuals drawn from the estimated distribution function. For each of these data sets, a model is identified. An uncertainty set with prespecified confidence level can then be computed experimentally from the large number of estimated models. The procedure is interesting, but very heavy on computer time.

The problem addressed in the present paper is the same as the one addressed by Tjärnström, namely how to construct an uncertainty region for the whole transfer function $G_0(e^{j\omega})$ with a prespecified confidence level, when this transfer function is obtained by Prediction Error Identification. We propose a much simpler procedure based on the commonly used elliptic uncertainty sets $U(\omega, \chi)$. More precisely, we solve the following problem: "*How should one choose the size χ of the ellipses $U(\omega, \chi)$ in such a way that the probability that $G_0(z) \in \mathcal{L}(\chi)$ is at least equal to some prespecified level α?*". We shall show that a practical solution is simply to choose frequency domain uncertainty ellipses $U(\omega, \chi_k)$, rather than $U(\omega, \chi_2)$ as is commonly done, where $k = \dim(\theta)$ and χ_k is obtained from the χ^2 distribution as the value for which $Pr(\chi^2(k) < \chi_k) \triangleq \alpha$. By doing so, the set $\mathcal{L}(\chi_k)$ obtained by gluing together the ellipses $U(\omega, \chi_k)$ will have a confidence level of at least α. Examples will illustrate the difference between the confidence level for the sets $\mathcal{L}(\chi_2)$ and $\mathcal{L}(\chi_k)$, respectively, for different values of $k = \dim(\theta)$.

An additional contribution of our paper will be to clarify the relationship between the ellipsoidal parameter uncertainty set and the transfer function uncertainty set, and their

[3] By *simultaneous confidence degree* Tjärnström means the probability that the true G_0 belongs to the set $\mathcal{L}(\chi_2)$ obtained by connecting together all individual sets $U(\omega, \chi_2)$.

corresponding probability levels. In passing, we will demonstrate a rather intriguing property of this mapping: the set $\mathcal{L}(\chi_k)$ obtained by gluing together the ellipses $U(\omega, \chi_k)$ may contain models $G(z, \theta)$ for some θ that are not in the original ellipsoid $U_\theta(\chi_k)$ of which the ellipses $U(\omega, \chi_k)$ are the images by the linear mapping (4).

The material in the paper develops as follows. In Section 2 we define the relevant sets and state the problem of constructing a frequency domain uncertainty set with a guaranteed probability level. Section 3 presents some algebraic results on mappings between ellipsoidal sets of different dimensions. In Section 4, we show that the traditional procedure of constructing a frequency domain uncertainty region by gluing together the uncertainty ellipses $U(\omega, \chi_2)$ to which $G_0(\omega)$ belongs with probability α leads to a set $\mathcal{L}(\chi_2)$ with a confidence level smaller than α. In Section 5 we present our main result: we show that by constructing the uncertainty region $\mathcal{L}$ on the basis of the ellipses $U(\omega, \chi_k)$ rather than $U(\omega, \chi_2)$ we can guarantee for $\mathcal{L}(\chi_k)$ a prescribed probability of containing the true system. In Section 7 we compare the probability levels obtained for the whole uncertainty region $\mathcal{L}$ by these two approaches for some standard low order model structures. Some brief conclusions are offered in Section 8.

2 Definitions and problem statement

We consider that the true system has an input-output relation given by

$$y(t) = G_0(z)u(t) + v(t), \tag{6}$$

where $v(t)$ is a zero-mean Gaussian process, and where $G_0(z)$ can be parametrized, for some parameter vector $\theta_0 \in \mathbf{R}^k$, in the following linearly parametrized model structure $\mathcal{M}$:

$$\mathcal{M} = \{ \ G(z, \theta) = \Lambda(z)\theta \text{ with } \ \theta \in \mathbf{R}^k \ \} \tag{7}$$

$$\Lambda(z) = \begin{pmatrix} \Lambda_1(z) & \Lambda_2(z) & \ldots & \Lambda_k(z) \end{pmatrix} \text{ with } \Lambda_i(z) = b(z)A^i(z). \tag{8}$$

The $\Lambda_i(z)$ $(i = 1...k)$ are rational basis functions [?, ?, ?] of the form $b(z)A^i(z)$, where $b(z)$ is a given transfer function and $A(z)$ is an all-pass filter, i.e. $|A(e^{j\omega})| = 1 \ \forall\omega$.[4] Thus $G_0(z) = \Lambda(z)\theta_0$.

As stated in the introduction, the parameter estimate $\hat{\theta}$ obtained from N input-output data by PE identification is then a Gaussian random variable with mean θ_0 and with a covariance matrix $P_\theta \in \mathbf{R}^{k \times k}$ that can be estimated from the data [6]. Thus,

$$\hat{\theta} \sim \mathcal{N}(\theta_0, P_\theta). \tag{9}$$

It follows that

$$(\hat{\theta} - \theta_0)^T P_\theta^{-1}(\hat{\theta} - \theta_0) \sim \chi^2(k). \tag{10}$$

For any given level α, one can therefore construct the uncertainty set $U_\theta(\chi_k)$ defined in (1), to which the true θ_0 belongs with probability α if χ_k is chosen as the value for which

[4] $\Lambda_i(z) = z^{-i}$ is a particular case of (8).

$Pr(\chi^2(k) < \chi_k) \triangleq \alpha$.

By linearity of the model structure, the frequency response of the identified model, represented by the 2-vector $g(e^{j\omega}, \hat{\theta})$ (see (2)), is then also Gaussian with mean $g(e^{j\omega}, \theta_0)$ and variance $P(\omega)$: see (3)-(4). It follows that

$$(\hat{g}(e^{j\omega}) - g(e^{j\omega}, \theta_0))^T P(\omega)^{-1}(\hat{g}(e^{j\omega}) - g(e^{j\omega}, \theta_0)) \sim \chi^2(2). \tag{11}$$

Thus, at frequency ω, the true frequency response $g(e^{j\omega}, \theta_0)$ belongs with probability α to the uncertainty ellipse $U(\omega, \chi_2)$ defined in (5) if χ_2 is chosen as the value for which $Pr(\chi^2(2) < \chi_2) \triangleq \alpha$.

We now consider the frequency domain uncertainty region defined as follows:

$$\mathcal{L}(\chi) = \{G(z, \theta) \mid g(e^{j\omega}, \theta) \in U(\omega, \chi)\ \forall \omega\} \quad \text{where} \tag{12}$$

$$U(\omega, \chi) = \{g \in \mathbf{R}^2 \mid (g - \hat{g}(e^{j\omega}))^T P(\omega)^{-1}(g - \hat{g}(e^{j\omega})) \leq \chi\}. \tag{13}$$

The uncertainty region $\mathcal{L}(\chi)$ is thus the set of systems $G(z, \theta)$ whose frequency response is constrained to lie at each frequency ω in the ellipse $U(\omega, \chi)$ which is centered at the frequency response $\hat{g}(e^{j\omega})$ of the identified model.

A special case of $\mathcal{L}(\chi)$ is $\mathcal{L}(\chi_2)$, constructed from the sets $U(\omega, \chi_2)$ defined in (5). The set $\mathcal{L}(\chi_2)$ has been used in previous work, as noted in the introduction. The main contribution of this paper will be twofold:

- show that, if χ_2 is chosen such that $Pr(\chi^2(2) < \chi_2) \triangleq \alpha$, then $Pr(G(z, \theta_0) \in \mathcal{L}(\chi_2)) \neq \alpha$, and this probability is actually always smaller than α; this will be shown in Section 4.
- show which value of χ must be selected in the construction of the uncertainty ellipses $U(\omega, \chi)$ in such a way that $Pr(G(z, \theta_0) \in \mathcal{L}(\chi))$ is at least equal to α. This will be the object of Section 5.

In order to understand the first problem, let it suffice at this point to observe that there is a strong difference between the following two problems:

1. choose χ such that, at each ω, $Pr(g(e^{j\omega}, \theta_0) \in U(\omega, \chi)) = \alpha$;
2. choose χ such that $Pr(G_0(z) \in \mathcal{L}(\chi)) = Pr(g(e^{j\omega}, \theta_0) \in U(\omega, \chi)\ \ \forall \omega) = \alpha$.

The solution to the first problem, as we have shown above, is $\chi = \chi_2$ where χ_2 is such that $Pr(\chi^2(2) < \chi_2) = \alpha$ (e.g. $\chi_2 = 5.99$ when $\alpha = 0.95$). A generic example in Section 6 will illustrate that the confidence level of the corresponding uncertainty region $\mathcal{L}(\chi_2)$ can then be much smaller than α indeed.

Our approach to solve the above two problems is to first analyze the mapping (2) that connects the ellipsoidal parametric uncertainty set $U_\theta(\chi)$ to the elliptic transfer function uncertainty set $U(\omega, \chi)$, as well as its inverse mapping. This will enable us, in a second step, to analyze the connection between the parameter uncertainty set $U_\theta(\chi)$ and the transfer function uncertainty set $\mathcal{L}(\chi)$. To analyze these mappings, we introduce two new parameter sets.

The parameter set $C_\theta(U(\omega,\chi))$ is defined, at each frequency ω, as the set of parameters θ such that the frequency response $g(e^{j\omega},\theta)$ lies in $U(\omega,\chi)$ at frequency ω:

$$C_\theta(U(\omega,\chi)) = \{\theta \mid g(e^{j\omega},\theta) \in U(\omega,\chi)\}. \tag{14}$$

We show that $C_\theta(U(\omega,\chi))$ **is a different parameter set at each frequency**, and that each of these sets has infinite size. To establish these results, we shall present in Section 3 some linear algebra theorems about the properties of mappings of the form of the mapping $T(e^{j\omega})$ of (2).

We also define the parameter set $C_\theta(\mathcal{L}(\chi))$ as the set of parameters θ such that $G(z,\theta)$ lies in $\mathcal{L}(\chi)$:

$$C_\theta(\mathcal{L}(\chi)) = \{\theta \mid G(z,\theta) \in \mathcal{L}(\chi)\}, \tag{15}$$

where $\mathcal{L}(\chi)$ has been defined through (12)-(13). This set will play an important role in the computation of the confidence level for the uncertainty set $\mathcal{L}(\chi)$ since, by its very definition, $Pr(G_0(z) \in \mathcal{L}(\chi)) = Pr(\theta_0 \in C_\theta(\mathcal{L}(\chi)))$. It follows further from the last two definitions and the definition (12)-(13) of $\mathcal{L}(\chi)$ that

$$C_\theta(\mathcal{L}(\chi)) = \{\theta \mid g(e^{j\omega},\theta) \in U(\omega,\chi) \;\; \forall\omega\} = \bigcap_{\omega\in[0\;\;\pi]} C_\theta(U(\omega,\chi)). \tag{16}$$

The set $C_\theta(\mathcal{L}(\chi))$ is thus the intersection over all frequencies of the sets $C_\theta(U(\omega,\chi))$ which are the inverse images in θ-space of the ellipses $U(\omega,\chi)$, via the mapping (2). We now analyze some key properties of this mapping.

3 Linear algebra preliminaries

Let us consider two real linear spaces of size n and k, $(n < k)$, linked by a linear transformation T. This mapping has the following expression

$$x = Ty \tag{17}$$

where $y \in \mathbf{R}^k$, $x \in \mathbf{R}^n$ $(n < k)$ are real vectors, and $T \in \mathbf{R}^{n\times k}$ is a real matrix of rank n. Our interest lies in the situation where x and/or y are constrained to lie in an ellipsoid.

Theorem 3.1 *Consider the mapping T defined in (17) and an ellipsoid U_y with size parameter χ in the y-space:*

$$U_y = \{y \mid y^T P_y^{-1} y \le \chi\}, \tag{18}$$

with $P_y \in \mathbf{R}^{k\times k}$ a symmetric positive definite matrix. The image U_x of U_y under the mapping T, i.e. $U_x \triangleq \{x \mid x = Ty \text{ with } y \in U_y\}$, is an ellipsoid given by

$$U_x = \{x \mid x^T P_x^{-1} x \le \chi\}, \tag{19}$$

with $P_x = TP_yT^T \in \mathbf{R}^{n\times n}$.

Proof. See [**?**] or Appendix A. ■

Theorem 3.2 *Consider the mapping T defined in (17) and an ellipsoid U_x with size parameter χ in the x-space: $U_x = \{x \mid x^T P_x^{-1} x \leq \chi\}$, with $P_x \in \mathbf{R}^{n \times n}$ a symmetric positive definite matrix. Define the inverse image C_y of U_x using the mapping T as*

$$C_y \triangleq \{y \mid x = Ty \in U_x\}, \tag{20}$$

Then C_y is a volume given by

$$C_y = \{y \mid y^T R y \leq \chi\}, \tag{21}$$

with $R = T^T P_x^{-1} T$, a singular matrix $\in \mathbf{R}^{k \times k}$. Moreover, the matrix R defining C_y has rank n, i.e. it has $k - n$ zero eigenvalues. The volume C_y therefore has $k - n$ infinite principal axes. The directions y_i $(i = 1, ..., k - n)$ of these infinite principal axes are the eigenvectors corresponding to the null eigenvalues of R. Moreover, these eigenvectors y_i belong to the null space of T, i.e. $Ty_i = 0$.

Proof. See [**?**] or Appendix B. ∎

4 The uncertainty region based on ellipses of size χ_2

In this section we show that if χ_2 is chosen such that $Pr(\chi^2(2) < \chi_2) \triangleq \alpha$, then $Pr(G(z, \theta_0) \in \mathcal{L}(\chi_2)) < \alpha$. In other words, if the elliptic uncertainty sets $U(\omega, \chi_2)$ are scaled in such a way that the true transfer $g(e^{j\omega}, \theta_0)$ belongs at frequency ω to $U(\omega, \chi_2)$ with probability 95%, say, then the probability that the whole transfer function $G(z, \theta_0)$ belongs to the set $\mathcal{L}(\chi_2)$ obtained by "gluing together" these ellipses, will be strictly smaller than 95%.

Thus, assume that χ_2 is chosen such that $Pr(\chi^2(2) < \chi_2) \triangleq \alpha$. Then, by the definitions of $U(\omega, \chi)$ and of $C_\theta(U(\omega, \chi))$ (see (13) and (14) respectively), it follows that $Pr(\theta_0 \in C_\theta(U(\omega, \chi_2))) = Pr(g(\omega, \theta_0) \in U(\omega, \chi_2)) \triangleq \alpha$. We analyze the structure of these sets $C_\theta(U(\omega, \chi_2))$. It follows from Theorem 3.2 and from the definition (13) of $U(\omega, \chi)$ that, at any fixed but arbitrary frequency ω, the set $C_\theta(U(\omega, \chi_2))$ is a volume in θ-space centered at $\hat{\theta}$ with $k-2$ infinite axes in the direction of the vectors belonging to the null-space of $T(e^{j\omega})$:

$$C_\theta(U(\omega, \chi_2)) = \{\theta \in \mathbf{R}^k \mid (\theta - \hat{\theta})^T R(\omega)(\theta - \hat{\theta}) \leq \chi_2\}. \tag{22}$$

where $R(\omega) = T(e^{j\omega})^T P(\omega)^{-1} T(e^{j\omega})$. Key points here are that

- the sets $C_\theta(U(\omega, \chi_2))$ are different at each frequency since the matrices $R(\omega)$ are different at each frequency
- due to the independence of the basis functions defining $\Lambda(z)$ and hence $T(e^{j\omega})$, there cannot exist any particular direction that is an infinite axis of $C_\theta(U(\omega, \chi_2))$ at every frequency.

Now remember that $C_\theta(\mathcal{L}(\chi_2)) = \cap_{\omega\in[0\ \pi]} C_\theta(U(\omega,\chi_2))$: see (16). The set $C_\theta(\mathcal{L}(\chi_2)))$ is therefore a bounded set (i.e. with no infinite axis) and is thus clearly smaller than any one of the sets $C_\theta(U(\omega,\chi_2))$ for any frequency. Thus at each ω we have a strict inclusion:

$$C_\theta(\mathcal{L}(\chi_2)) \subset C_\theta(U(\omega,\chi_2)). \tag{23}$$

This has the following consequence on the probability level $Pr(G(z,\theta_0) \in \mathcal{L}(\chi_2))$.

Theorem 4.1 *Consider the identification of a true system $G_0(z) = G(z,\theta_0)$ ($\theta_0 \in \mathbf{R}^k$, $k > 2$) in a full-order model structure as presented in Section 2, and the uncertainty region $\mathcal{L}(\chi_2)$ defined in (12)-(13), where χ_2 is such that $Pr(\chi^2(2) < \chi_2) = \alpha$. Then*

$$Pr(G(z,\theta_0) \in \mathcal{L}(\chi_2)) < \alpha. \tag{24}$$

Proof. The result follows directly from the strict inclusion (23) and the fact that $Pr(G_0(z) \in \mathcal{L}(\chi_2)) = Pr(\theta_0 \in C_\theta(\mathcal{L}(\chi_2)))$ and that $Pr(\theta_0 \in C_\theta(U(\omega,\chi_2))) = \alpha$. ■

In Section 6, a generic example will show that $Pr(G_0(z) \in \mathcal{L}(\chi_2))$ can in fact be much smaller than the desired confidence level α when k is moderately large. This is to be expected since $C_\theta(U(\omega,\chi_2)) = \{\theta \mid g(e^{j\omega},\theta) \in U(\omega,\chi_2)\}$ is an infinite set (in $k-2$ directions) while $C_\theta(\mathcal{L}(\chi_2)) = \{\theta \mid G(z,\theta) \in \mathcal{L}(\chi_2)\}$ is a bounded set included in $C_\theta(U(\omega,\chi_2))$ for every ω.

5 Constructing an uncertainty region with guaranteed confidence level

In this section, we present the main result of our paper. We show that we can guarantee that $Pr(G(z,\theta_0) \in \mathcal{L}(\chi))$ is at least as large as some predefined level α if we use $\mathcal{L}(\chi_k)$ with χ_k defined as the value for which $Pr(\chi^2(k) < \chi_k) \triangleq \alpha$. In other words, we construct $\mathcal{L}(\chi_k)$ on the basis of the ellipses $U(\omega,\chi_k)$, rather than the smaller ellipses $U(\omega,\chi_2)$ as is commonly done.

To show this, we first observe that $Pr(G(z,\theta_0) \in \mathcal{L}(\chi)) = Pr(\theta_0 \in C_\theta(\mathcal{L}(\chi))$, and we analyze the connection between the parameter sets $U_\theta(\chi_k)$ and $C_\theta(\mathcal{L}(\chi_k))$: see (1) and (16), respectively. By definition of $U_\theta(\chi_k)$, we have $Pr(\theta_0 \in U_\theta(\chi_k)) = \alpha$. Now, using Theorem 3.1 and (4) we have the following alternative characterization of the sets $U(\omega,\chi_k)$:

$$U(\omega,\chi_k) = \{\ g(e^{j\omega},\theta) \mid \theta \in U_\theta(\chi_k)\ \}. \tag{25}$$

It then follows from the definition (14) of $C_\theta(U(\omega,\chi_k))$ that $U_\theta(\chi_k) \subseteq C_\theta(U(\omega,\chi_k))$ for each ω. In addition, since $C_\theta(U(\omega,\chi_k))$ has been shown to contain infinite axes (see Section 4) while $U_\theta(\chi_k)$ is a bounded ellipsoid, it follows necessarily that, at each ω,

$$U_\theta(\chi_k) \subset C_\theta(U(\omega,\chi_k)). \tag{26}$$

Combining (26) and (16), we have that $U_\theta(\chi_k) \subseteq C_\theta(\mathcal{L}(\chi_k))$. The following proposition, proved in Appendix C, actually shows that $C_\theta(\mathcal{L}(\chi_k)) \setminus U_\theta(\chi_k) \neq \emptyset$, i.e. the inclusion is strict:

$$U_\theta(\chi_k) \subset C_\theta(\mathcal{L}(\chi_k)). \tag{27}$$

Proposition 5.1 *Consider the identification of a true system $G_0(z) = G(z, \theta_0)$ ($\theta_0 \in \mathbf{R}^k$, $k > 2$) in a full-order model structure as presented in Section 2, the frequency domain uncertainty region $\mathcal{L}(\chi)$ of size χ defined in (12)-(13), and the sets $C_\theta(\mathcal{L}(\chi))$ and $U_\theta(\chi)$ defined in (15) and (1), respectively. Then $C_\theta(\mathcal{L}(\chi)) \setminus U_\theta(\chi) \neq \emptyset$.*

Proof: see Appendix C.

It follows from (27) and the property $Pr(\theta_0 \in U_\theta(\chi_k)) = \alpha$ that

$$Pr(\theta_0 \in C_\theta(\mathcal{L}(\chi_k)) > \alpha. \tag{28}$$

Collecting all these results, we have the following theorem concerning $Pr(G_0 \in \mathcal{L}(\chi_k))$.

Theorem 5.1 *Consider the identification of a true system $G_0(z) = G(z, \theta_0)$ ($\theta_0 \in \mathbf{R}^k$, $k > 2$) in a full-order model structure, as presented in Section 2, and the frequency domain uncertainty region $\mathcal{L}(\chi_k)$ of size χ_k defined in (12)-(13). Then we have the following relations between the uncertainty sets defined earlier:*

$$U_\theta(\chi_k) \subset C_\theta(\mathcal{L}(\chi_k)) \subset C_\theta(U(\omega, \chi_k)). \tag{29}$$

If α is the desired probability level and χ_k is such that $Pr(\chi^2(k) < \chi_k) = \alpha$ with $k = dim(\theta_0)$, then we have in particular

$$Pr(\chi^2(k) < \chi_k) = \alpha < Pr(G(z, \theta_0) \in \mathcal{L}(\chi_k)) < Pr(\chi^2(2) < \chi_k). \tag{30}$$

Proof. The lower bound in (29) has been shown above, while the upper bound follows by substituting χ_k for χ_2 in (23)[5]. The inequalities in (30) follow from $Pr(\theta_0 \in U_\theta(\chi_k)) = Pr(\chi^2(k) < \chi_k) = \alpha$, $Pr(G(z, \theta_0) \in \mathcal{L}(\chi_k)) = Pr(\theta_0 \in C_\theta(\mathcal{L}(\chi_k))$, and $Pr(\theta_0 \in C_\theta(U(\omega, \chi_k)) = Pr(\chi^2(2) < \chi_k)$, respectively. The last part follows from the definitions of $C_\theta(U(\omega, \chi_k))$ and $U(\omega, \chi_k)$ (see (14) and (5), respectively) and the property (11). ∎

The inequalities (30) summarize the main technical results of this paper. They hold of course for any value χ in lieu of χ_k. They state that the probability $Pr(G_0(z) \in \mathcal{L}(\chi))$ is larger than the probability that the true parameter vector θ_0 lies in $U_\theta(\chi)$, but smaller than the probability that $g(e^{j\omega}, \theta_0)$ lies in the ellipse $U(\omega, \chi)$ at any single frequency ω. Observe that, for $k > 2$, we have $\chi_k > \chi_2$ and hence $Pr(\chi^2(2) < \chi_k) > Pr(\chi^2(2) < \chi_2)$.

The theorem offers a simple procedure for the construction of a frequency domain uncertainty region with confidence level of at least α, derived from a PE identification experiment: if $dim(\theta) = k$, then construct the uncertainty region $\mathcal{L}(\chi_k)$ obtained by gluing together the ellipses $U(\omega, \chi_k)$ of size χ_k where, as usual, χ_k is the value for which $Pr(\chi^2(k) < \chi_k) = \alpha$. The procedure is significantly simpler and faster than that proposed in [9]; the penalty for this simplicity is that the delivered probability is larger than the desired level α. Our example in the next section will illustrate how conservative this bound can be.

[5]Note that (23) is expressed in terms of χ_2, but it holds for any value of χ.

Final comments. We have shown that, to construct a confidence region $\mathcal{L}(\chi)$ with probability α by gluing together the frequency ellipses $U(\omega, \chi)$, we need to work with the ellipses $U(\omega, \chi_k)$ rather than $U(\omega, \chi_2)$, where $k = dim(\theta)$. Both ellipses have, at each ω, the same center $\hat{g}(\omega) = g(e^{j\omega}, \hat{\theta})$ and the same axes, but $U(\omega, \chi_k)$ is a dilated version of $U(\omega, \chi_2)$ (for $k > 2$) since $\chi_k > \chi_2$. A schematic view of this is given in Figure 1 at an arbitrary frequency ω. As an example, when $\alpha = 0.95$ and $k = 6$, we have $\chi_6 = 12.6$ and $Pr(g(e^{j\omega}, \theta_0) \in U(\omega, \chi_k)) = 0.999$, whereas $\chi_2 = 5.99$ and $Pr(g(e^{j\omega}, \theta_0) \in U(\omega, \chi_2)) = 0.95$.

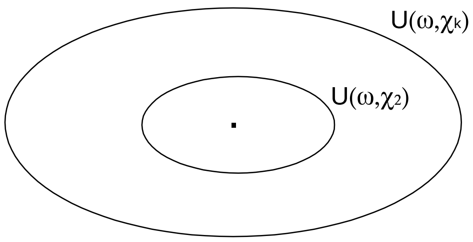

Figure 1: $U(\omega, \chi_2)$ and $U(\omega, \chi_k)$. The dot "." represents the center $\hat{g}(\omega)$.

6 Numerical illustration

Recall the identification experiment of Section 2, and consider here for simplicity that $v(t)$ and $u(t)$ in (6) are realizations of white noise signals of variance σ_e^2 and σ_u^2, respectively, and that the transfer functions $\Lambda_i(z)$ in (8) are given by z^{-i} $(i = 1, ..., k)$. Then the covariance matrix P_θ of the identified parameter vector $\hat{\theta}$ in (9) is the following diagonal matrix:

$$P_\theta = \frac{1}{N} \frac{\sigma_e^2}{\sigma_u^2} I_k, \tag{31}$$

where k is the size of the vector θ_0 and N the number data used in the identification. The results presented in the sequel are independent of the values of θ_0, $\hat{\theta}$, N, σ_e^2 and σ_u^2.

Suppose that we want to construct an uncertainty set for $G(z, \theta_0)$ with probability level $\alpha = 0.95$. Using the expression of P_θ given in (31), we have computed $Pr(G(z, \theta_0) \in \mathcal{L}(\chi)) = Pr(\theta_0 \in C_\theta(\mathcal{L}(\chi)))$ by approximating $C_\theta(\mathcal{L}(\chi))$ as the intersection of the volumes $C_\theta(U(\omega, \chi))$ at a number of different frequencies and by using a grid of points in this intersection to integrate the probability density function (9). We have done this for $k = 3, 4$ and 6, and for $\chi = \chi_2 = 5.99$, and $\chi = \chi_k$. For computational reasons we could not consider larger values for k since the number of points in the grid grows exponentially with k. The results for $k = 3$, 4 and 6 are presented in Tables 1, 2 and 3, respectively. Note that, in these tables, we compare the estimated $Pr(G(z, \theta_0) \in \mathcal{L}(\chi))$ with its upper and lower bound given in (30).

From these tables, we see how wrong it would be to choose $\chi = \chi_2 = 5.99$ as the size parameter for the construction of an uncertainty set $\mathcal{L}(\chi)$ for which $Pr(G_0(z) \in \mathcal{L}(\chi)) = 0.95$. Indeed, $Pr(G_0(z) \in \mathcal{L}(\chi_2))$ is only 0.74 when $k = 6$. In this generic example, we can also

χ	$Pr(\chi^2(k=3) < \chi)$	$Pr(G_0(z) \in \mathcal{L}(\chi))$	$Pr(\chi^2(2) < \chi)$
$\chi_2 = 5.99$	0.89	0.90	0.95
$\chi_{k=3} = 7.81$	0.95	0.96	0.98

Table 1: $k = 3$

χ	$Pr(\chi^2(k=4) < \chi)$	$Pr(G_0(z) \in \mathcal{L}(\chi))$	$Pr(\chi^2(2) < \chi)$
$\chi_2 = 5.99$	0.80	0.85	0.95
$\chi_{k=4} = 9.49$	0.95	0.97	0.99

Table 2: $k = 4$

see that this phenomenon worsens when k increases: $Pr(G_0(z) \in \mathcal{L}(\chi_2)) = 0.90$ when $k = 3$, 0.85 when $k = 4$, and only 0.74 when $k = 6$.

As expected by Theorem 5.1, for each considered k, the choice $\chi = \chi_k$ delivers an uncertainty region $\mathcal{L}(\chi_k)$ containing G_0 with a probability of at least 0.95. In this generic example, even though this probability $Pr(G_0(z) \in \mathcal{L}(\chi_k))$ is never relatively much larger than 0.95, we nevertheless notice that it increases when k increases: $Pr(G_0(z) \in \mathcal{L}(\chi_k))$ is equal to 0.96 when $k = 3$, 0.97 when $k = 4$ and 0.98 when $k = 6$. Note also that, since χ_k increases when k increases, the probability $Pr(g(e^{j\omega}, \theta_0) \in U(\omega, \chi_k)) = Pr(\chi^2(2) < \chi_k)$ also increases when k increases: it is equal to 0.98 when $k = 3$, 0.99 when $k = 4$ and 0.999 when $k = 6$.

7 Conclusions

We have developed a procedure, suitable for PE identification of 'linear in the parameter models', for the construction of a frequency domain uncertainty region to which the true system is guaranteed to belong with a probability level that is at least equal to some a priori fixed level α. The frequency region is based on the commonly used frequency domain ellipses, obtained at each frequency. Our main contribution has been to show how to choose the size of the ellipses in order to guarantee the required level α for the whole region.

χ	$Pr(\chi^2(k=6) < \chi)$	$Pr(G_0(z) \in \mathcal{L}(\chi))$	$Pr(\chi^2(2) < \chi)$
$\chi_2 = 5.99$	0.58	0.74	0.95
$\chi_{k=6} = 12.6$	0.95	0.98	0.999

Table 3: $k = 6$

Acknowledgments

We would like to thank Glenn Vinnicombe, Paresh Date and John Steele of the University of Cambridge who have motivated us to analyze this problem.

References

[1] P. Andersen, S. Töffner-Clausen, and T.S. Pedersen. Estimation of frequency domain model uncertainties with application to robust control design. In *Proc. IFAC Symposium on System Identification*, pages 603–608, Copenhagen, 1994.

[2] X. Bombois. *Connecting Prediction Error Identification and Robust Control Analysis: a new framework*. PhD thesis, Université Catholique de Louvain, 2000.

[3] D.K. de Vries and P.M.J. Van den Hof. Quantification of uncertainty in transfer function estimation: a mixed probabilistic - worst-case approach. *Automatica*, 31:543–558, 1995.

[4] G.C. Goodwin, M. Gevers, and B. Ninness. Quantifying the error in estimated transfer functions with application to model order selection. *IEEE Trans. Automatic Control*, 37:913–928, 1992.

[5] R.G. Hakvoort and P.M.J. Van den Hof. Identification of probabilistic system uncertainty regions by explicit evaluation of bias and variance errors. *IEEE Trans. Automatic Control*, 42(11):1516–1528, 1997.

[6] L. Ljung. Model validation and model error modeling. In *Proceedings of the Åström Symposium on Control. In B. Wittenmark and A. Rantzer Eds.*, pages 15–42, Lund, Sweden, 1999. Studentliteratur.

[7] L. Ljung. *System Identification: Theory for the User, 2nd Edition.* Prentice-Hall, Englewood Cliffs, NJ, 1999.

[8] L. Ljung. Model error modeling and control design. In *CD-Rom Proc. of IFAC Symposium on System Identification, paper WeAM1-3*, Santa Barbara, California, 2000.

[9] E.B. Manoukian. *Modern Concepts and Theorems of Mathematical Statistics*. Springer-Verlag, New York, 1986.

[10] B. Ninness and F. Gustafsson. A unifying construction of orthonormal bases for system identification. *IEEE Transactions on Automatic Control*, 42(4):515–521, April 1997.

[11] F. Tjärnström. Computing uncertainty regions with simultaneous confidence degree using bootstrap. In *Proc. IFAC Symposium on System Identification*, pages 1133–1138, Santa Barbara, 2000.

[12] F. Tjärnström and U. Forssell. Comparison of methods for probabilistic bounding. In *Proc. IEEE Conference on Decision and Control*, pages 522–527, Phoenix, 1999.

[13] S. Töffner-Clausen, P. Andersen, J. Stoustrup, and H.H. Niemann. Estimated frequency domain model uncertainties used in robust controller design - a μ-approach. In *Proc. IEEE Conference on Control Applications*, pages 1585–1590, Glasgow, 1994.

[14] B. Wahlberg. System identification using Laguerre models. *IEEE Transactions on Automatic Control*, 36(5):551–562, May 1991.

[15] B. Wahlberg. System identification using Kautz models. *IEEE Transactions on Automatic Control*, 39(6):1276–1282, June 1994.

[16] L. Wang and G.C. Goodwin. Integrating identification with robust control: a mixed H_2/H_∞ approach. In *CD-ROM Proc. 39th Conference on Decision and Control*, Sydney, Australia, 2000.

A Proof of Theorem 3.1

Let us first prove the following lemma that will be useful to prove Theorem 3.1.

Lemma A.1 *Consider the partitioned symmetric positive definite matrix* $P \in \mathbf{R}^{k\times k}$*:*

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{12}^T & P_{22} \end{pmatrix}$$

with $P_{11} \in \mathbf{R}^{n\times n}$*,* $P_{12} \in \mathbf{R}^{n\times(k-n)}$ *and* $P_{22} \in \mathbf{R}^{(k-n)\times(k-n)}$*. Consider also two real vectors* $x \in \mathbf{R}^n$ *and* $\bar{x} \in \mathbf{R}^{(k-n)}$ *and an ellipsoid* $U_{x\bar{x}}$ *defined as:*

$$U_{x\bar{x}} = \left\{ \begin{pmatrix} x \\ \bar{x} \end{pmatrix} \mid \begin{pmatrix} x \\ \bar{x} \end{pmatrix}^T P^{-1} \begin{pmatrix} x \\ \bar{x} \end{pmatrix} \leq 1 \right\}.$$

Then the set U_x

$$U_x = \{x \mid \begin{pmatrix} x \\ \bar{x} \end{pmatrix} \in U_{x\bar{x}}\} \tag{32}$$

is also an ellipsoid given by

$$U_x = \{x \mid x^T P_{11}^{-1} x \leq 1\} \tag{33}$$

Proof. The inverse of the block matrix P can be written (see e.g. [**?**, page 22])

$$P^{-1} = \begin{pmatrix} K_{11} & K_{12} \\ K_{12}^T & K_{22} \end{pmatrix}$$

where $K_{11} = P_{11}^{-1} + P_{11}^{-1}P_{12}\Delta^{-1}P_{12}^T P_{11}^{-1}$, $K_{12} = -P_{11}^{-1}P_{12}\Delta^{-1}$, $K_{22} = \Delta^{-1}$ and $\Delta = P_{22} - P_{12}^T P_{11}^{-1} P_{12}$. Using these notations and introducing the vector $z = K_{22}^{-1}K_{12}^T x + \bar{x}$, we have the following equivalences:

$$\begin{aligned} \begin{pmatrix} x \\ \bar{x} \end{pmatrix}^T P^{-1} \begin{pmatrix} x \\ \bar{x} \end{pmatrix} \leq 1 \quad &\Longleftrightarrow \quad x^T(K_{11} - K_{12}K_{22}^{-1}K_{12}^T)x + z^T K_{22} z < 1 \\ &\Longleftrightarrow \quad x^T P_{11}^{-1} x + z^T K_{22} z \leq 1 \end{aligned} \tag{34}$$

Using this last expression, we can now write that

1. if $(x^T \ \bar{x}^T)^T \in U_{x\bar{x}}$, then $x^T P_{11}^{-1} x < 1$. Indeed

$$\begin{pmatrix} x \\ \bar{x} \end{pmatrix}^T P^{-1} \begin{pmatrix} x \\ \bar{x} \end{pmatrix} < 1 \Longrightarrow x^T P_{11}^{-1} x < (1 - z^T K_{22} z) < 1$$

2. if $x^T P_{11}^{-1} x \leq 1$ then there exists $\bar{x}$ such that $(x^T \ \bar{x}^T)^T \in U_{x\bar{x}}$. Indeed, take as $\bar{x}$ the vector $\bar{x}$ such that $z = 0$ (i.e. $\bar{x} = -K_{22}^{-1} K_{12}^T x$). Then $\begin{pmatrix} x \\ -K_{22}^{-1} K_{12}^T x \end{pmatrix} \in U_{x\bar{x}}$.

This completes the proof. ■

Proof of Theorem 3.1. Let us first complete the mapping T with a matrix $\overline{T} \in \mathbf{R}^{(k-n)\times k}$ to generate a nonsingular mapping $\widetilde{T}$:

$$\begin{pmatrix} x \\ \bar{x} \end{pmatrix} = \overbrace{\begin{pmatrix} T \\ \overline{T} \end{pmatrix}}^{\widetilde{T}} y \tag{35}$$

such that $\tilde{T} \in \mathbf{R}^{k\times k}$ has rank k. Using $\widetilde{T}$, we have

$$y^T P_y^{-1} y \leq \chi \Longleftrightarrow \begin{pmatrix} x \\ \bar{x} \end{pmatrix}^T \overbrace{\widetilde{T}^{-T} P_y^{-1} \widetilde{T}^{-1}}^{P^{-1}} \begin{pmatrix} x \\ \bar{x} \end{pmatrix} \leq \chi \tag{36}$$

Proving Theorem 3.1 is thus equivalent to proving that (19) is the domain where x is constrained to lie when (36) holds. This follows immediately from Lemma A.1, noting that if $P = \widetilde{T} P_y \widetilde{T}^T$, then $P_x = P_{11} = T P_y T^T$. ■

B Proof of Theorem 3.2

We first prove that the inverse image of U_x by the mapping (17) is given by (21). This follows directly from:

$$x^T P_x^{-1} x < \chi \Longleftrightarrow y^T T^T P_x^{-1} T y \leq \chi. \tag{37}$$

The volume C_y is thus the inverse image of U_x since y has to satisfy the right-hand side of (37) in order to have x in U_x. It follows then from $R = T^T P_x^{-1} T \in \mathbf{R}^{k\times k}$ with T of rank $n < k$ that R has $k - n$ null eigenvalues and that the corresponding eigenvectors are in the null-space of the mapping T. ■

C Proof of Proposition 5.1

Let us first prove the following lemma that will be useful to prove this proposition.

Lemma C.1 *Let a be a real constant, $P_\theta \in \mathbf{R}^{3\times 3}$ be a symmetric positive definite matrix and $T \in \mathbf{R}^{2\times 3}$ be of rank 2. Then the intersection of*

$$x^T P_\theta^{-1} x = a \tag{38}$$

$$\textit{and} \;\; x^T T^T (T P_\theta T^T)^{-1} T x = a \tag{39}$$

is the contour of an ellipse.

Proof. Let us decompose P_θ in $P_\theta = U^T D U$, where U is orthogonal and D diagonal, and let us define $z \triangleq D^{-\frac{1}{2}} U x$ and $R \triangleq T U^T D^{\frac{1}{2}}$. Then, (38) and (39) become $z^T z = a$ and $z^T R^T (R R^T)^{-1} R z = a$, respectively. This implies that

$$z^T \overbrace{\left(I_3 - R^T (R R^T)^{-1} R\right)}^{\triangleq A} z = 0. \tag{40}$$

The matrix A is easily seen to be nonnegative and of rank 1 (i.e. since $AR^T = 0$). Hence, A can be decomposed in $A = \alpha\alpha^T$ for some vector $\alpha \neq 0$ of dimension 3. Then, from (40), we have $\alpha^T z = \alpha^T D^{-\frac{1}{2}} U x = 0$. This shows that the intersection of (38) and (39) lies in a plane and this plane intersects (38) and (39), obviously as the contour of an ellipse. ■

Proof of Proposition 5.1. We will establish that $C_\theta(\mathcal{L}(\chi)) \setminus U_\theta(\chi) \neq \emptyset$ for $k = dim(\theta) > 2$ in two steps. First, we analyze the case where the size k of the parameter vector θ is 3.

Step 1. When $k = 3$, the set $C_\theta(U(\omega, \chi))$ given in (22) is, at each ω, a cylinder containing $U_\theta(\chi)$ and with an infinite axis in the direction $\theta_{null}(\omega)$ which is the eigenvector corresponding to the null eigenvalue of $T(e^{j\omega}) \in \mathbf{R}^{2\times 3}$. Using (8) and (2), it is easy to see that

$$\theta_{null}(\omega) = \begin{pmatrix} 1 & -2cos(\phi(\omega)) & 1 \end{pmatrix}^T \tag{41}$$

where $\phi(\omega)$ is the phase of the all-pass filter $A(e^{j\omega})$. Note also that, according to Lemma C.1, the intersection $E_i(\omega, \chi)$ between the surface of $C_\theta(U(\omega, \chi))$ and the surface of $U_\theta(\chi)$:

$$E_i(\omega, \chi) = \{\theta \mid (\theta - \hat{\theta})^T P_\theta^{-1} (\theta - \hat{\theta}) = \chi\} \;\bigcap\; \{\theta \mid (\theta - \hat{\theta})^T R(\omega) (\theta - \hat{\theta}) = \chi\} \tag{42}$$

is the contour of an ellipse centered at $\hat{\theta}$.

We will prove the existence of vectors $\theta_{out} \in C_\theta(\mathcal{L}(\chi)) \setminus U_\theta(\chi)$ by contradiction. For this purpose, note that there would *not* exist such θ_{out} if and only if each parameter vector θ_s on the surface of $U_\theta(\chi)$ (i.e. each parameter vector θ_s such that $(\theta_s - \hat{\theta})^T P_\theta^{-1} (\theta_s - \hat{\theta}) = \chi$) belongs to the surface of (at least) one $C_\theta(U(\omega, \chi))$ [6]. Consequently, there would not exist such θ_{out} if and only if the union $\cup_{\omega \in [0\ \pi]} E_i(\omega, \chi)$ (see (42)) is equal to the surface of $U_\theta(\chi)$. This last condition is only possible if the axes of the different cylinders $C_\theta(U(\omega, \chi))$, when

[6] Indeed, under this condition, each $\theta_{out} \notin U_\theta(\chi)$ is also (at least) outside the cylinder $C_\theta(U(\omega, \chi))$ to the surface of which the vector $\theta_s - \hat{\theta}$ in the same direction as $\theta_{out} - \hat{\theta}$ belongs. The vector $\theta_s - \hat{\theta}$ in the same direction as $\theta_{out} - \hat{\theta}$ is the vector on the surface of $U_\theta(\chi)$ such that $\theta_s - \hat{\theta} = \beta(\theta_{out} - \hat{\theta})$ for some constant α $(0 < \beta < 1)$.

ω goes from 0 to π, are sufficiently different.

The direction of the axis of the cylinder $C_\theta(U(\omega,\chi))$ at a particular ω is given by the vector $\theta_{null}(\omega)$. By inspecting (41), we see that $\theta_{null}(\omega)$ always takes its value in one section of a plane delimited by $(1 \quad -2 \quad 1)^T$ and $(1 \quad 2 \quad 1)^T$. The angle between these two extreme vectors is 1.91 radians ($< \pi$ radians). The axes of the cylinders $C_\theta(U(\omega,\chi))$ can therefore not cover an angle of more than 1.91 radians when ω goes from 0 to π. This makes it impossible that $\cup_{\omega\in[0\ \pi]}E_i(\omega,\chi)$ equals the total surface of $U_\theta(\chi)$. Consequently, $C_\theta(\mathcal{L}(\chi)) \setminus U_\theta(\chi) \neq \emptyset$ for $k=3$.

Step 2. We will now extend the proof to the case where $k>3$ by using the fact that the strict inclusion has been proven for the case $k=3$. In this second step we will also use the following equivalence: $\theta_{out} \in C_\theta(\mathcal{L}(\chi)) \setminus U_\theta(\chi) \Leftrightarrow \theta_{out} \notin U_\theta(\chi)$ and there exists, at each frequency[7], a θ_{in} in $U_\theta(\chi)$ such that $G(e^{j\omega},\theta_{out}) = G(e^{j\omega},\theta_{in})$. This equivalence is a consequence of the relations (16) and (25).

Let us first introduce some notations: we partition the inverse of the matrix $P_\theta \in \mathbf{R}^{k\times k}$ defining $U_\theta(\chi)$ as follows

$$P_\theta^{-1} = \begin{pmatrix} P_{11}^{-1} & P_{12}^{-1} \\ P_{12}^{-T} & P_{22}^{-1} \end{pmatrix} \tag{43}$$

where $P_{11}^{-1} \in \mathbf{R}^{3\times 3}$, $P_{12}^{-1} \in \mathbf{R}^{3\times(k-3)}$ and $P_{22}^{-1} \in \mathbf{R}^{(k-3)\times(k-3)}$. Note that, since P_θ^{-1} is a symmetric positive-definite matrix, the matrix P_{11}^{-1} is also a symmetric positive-definite matrix. We also partition the identified vector $\hat{\theta}$ into $\hat{\theta} = \begin{pmatrix} \hat{\xi} & \hat{\eta} \end{pmatrix}^T$ with $\hat{\xi} \in \mathbf{R}^3$ and $\hat{\eta} \in \mathbf{R}^{(k-3)}$; and $\Lambda(z)$ into $\Lambda(z) = \begin{pmatrix} \Lambda_\xi(z) & \Lambda_\eta(z) \end{pmatrix}$ with $\Lambda_\xi(z)$ and $\Lambda_\eta(z)$ containing 3 and $k-3$ basis functions, respectively.

Consider now the following system description in ξ and the following ellipsoid defined by $P_{11}^{-1} \in \mathbf{R}^{3\times 3}$ and the vector $\hat{\xi}$:

$$G(z,\xi) = \Lambda_\xi(z)\xi \quad U_\xi(\chi) = \{\xi \mid (\xi-\hat{\xi})^T P_{11}^{-1}(\xi-\hat{\xi}) \leq \chi\}$$

where $\xi \in \mathbf{R}^3$ is a new parameter vector of dimension 3. Using the result of Step 1 and the equivalence given in the beginning of Step 2, we know that there exist vectors ξ_{out} (independent of frequency) having all the following property: $\xi_{out} \notin U_\xi(\chi)$ and $G(e^{j\omega},\xi_{out}) = G(e^{j\omega},\xi_{in})$ for some ξ_{in} at each frequency, with ξ_{in} in general frequency-dependent.

For **each** vector ξ_{out} of dimension 3 having the above property, we construct the vector $\theta_{out} \in \mathbf{R}^k \triangleq (\xi_{out}\ \hat{\eta})^T$. This vector θ_{out} does not lie in $U_\theta(\chi)$. Indeed, using (43) and $\xi_{out} \notin U_\xi(\chi)$, we have $(\theta_{out}-\hat{\theta})^T P_\theta^{-1}(\theta_{out}-\hat{\theta}) = (\xi_{out}-\hat{\xi})^T P_{11}^{-1}(\xi_{out}-\hat{\xi}) > \chi$. Moreover, we can define, at each ω, a vector $\theta_{in} \triangleq (\xi_{in}\ \hat{\eta})^T$, using the vector ξ_{in} corresponding to that ω. This vector θ_{in} lies in $U_\theta(\chi)$ since $(\theta_{in}-\hat{\theta})^T P_\theta^{-1}(\theta_{in}-\hat{\theta}) = (\xi_{in}-\hat{\xi})^T P_{11}^{-1}(\xi_{in}-\hat{\xi}) \leq \chi$ and has furthermore the property that $G(e^{j\omega},\theta_{out}) = G(e^{j\omega},\theta_{in})$ since $G(e^{j\omega},\xi_{out}) = G(e^{j\omega},\xi_{in})$. This shows that $C_\theta(\mathcal{L}(\chi)) \setminus U_\theta(\chi) \neq \emptyset$ and completes thus the proof for $k>3$.

[7] Note that there is however no single value of θ_{in} which applies at all frequencies.