Published version: https://dx.doi.org/10.1016/j.automatica.2013.08.027

Input design as a tool to improve the convergence of PEM ⋆

Diego Eckhard [a] Alexandre S. Bazanella [b] Cristian R. Rojas [c] Håkan Hjalmarsson [c]

[a] *Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul, Brazil*

[b] *Department of Electrical Engineering, Universidade Federal do Rio Grande do Sul, Brazil*

[c] *Automatic Control Lab and ACCESS Linnaeus Centre, School of Electrical Engineering, KTH Royal Institue of Technology*

Abstract

The Prediction Error Method (PEM) is related to an optimization problem built on input/output data collected from the system to be identified. It is often hard to find the global solution of this optimization problem because the corresponding objective function presents local minima and/or the search space is constrained to a nonconvex set. The shape of the cost function, and hence the difficulty in solving the optimization problem, depends directly on the experimental conditions, more specifically on the spectrum of the input/output data collected from the system. Therefore, it seems plausible to improve the convergence to the global minimum by properly choosing the spectrum of the input; in this paper, we address this problem. We present a condition for convergence to the global minimum of the cost function and propose its inclusion in the input design. We present the application of the proposed approach to case studies where the algorithms tend to get trapped in nonglobal minima.

Key words: Identification methods; Experiment design

1 Introduction

The Prediction Error Method (PEM) for parameter identification uses input-output data collected from the process to form a cost function. The parameters are then estimated as the solution of the optimization of this cost function. Under mild assumptions, the global minimum of the cost function is a consistent estimate of the model parameters, and the asymptotic variance equals the limit of the Cramér-Rao Bound. Therefore, identification by means of PEM provides a consistent and otherwise statistically appealing estimate of the system parameters and transfer function, provided that the global minimum of the cost function is obtained by the optimization procedure [22].

One difficulty in applying the PEM method is that in many cases achieving the global minimum may prove difficult [24, 38], for two main reasons: the cost function is usually not convex and the problem is constrained to a nonconvex set - namely, the set of parameters which yield stable predictors. Currently adopted solutions to this problem consist mainly in searching for good initial conditions to initialize the optimization, which is performed with some standard algorithm - steepest descent, Newton-Raphson, Levenberg-Marquardt, and the like. A "good" initial condition is one that is close enough to the global minimum to begin with, that is, closer to it than any local maxima or minima that would prevent convergence to the global minimum if they happened to be in the optimization path. Although there seem to be no firmly established guarantees that these solutions yield the global minimum, they have been successfully applied for many years. They are able to achieve the global minimum of the objective function in most cases, but failure to do so is not such an uncommon occurrence either. As the model order becomes larger, the trend to get trapped in local minima or to "converge" to the boundary of the search space, thus providing a useless model, seems to grow; we provide a couple of such examples in this paper. This problem has received renewed interest in the past few years, as pointed out in [24], and different approaches have emerged to cope with it, such as the use of resampling schemes to the input-output data [12] and the approximation of the process model by (potentially high order) linearly parameterized model structures to

⋆ An early version of this paper has been presented at SYSID 2012 - the 16^{th} IFAC Symposium on System Identification. This work was supported by CAPES and CNPq/Brasil, by the European Research Council under the advanced grant LEARN, contract 267381, and by the Swedish Research Council under contract 621-2009-4017.
E-mail address: diego@eckhard.com.br (D. Eckhard)

obtain convex cost functions [15] [19].

In this work, we present a different approach to the convergence problem. We focus on the cost function itself and its shape, trying to avoid the very existence of local minima and maxima. If the cost function has a "good" shape, meaning that it does not have local extrema other than the global minimum or inflection points, then it is easier for any gradient based optimization algorithm to converge to the global minimum, even if initialized far from the global minimum. We analyze what are the conditions under which the cost function has a "good" shape. We show that this property depends on the experimental conditions, more specifically on the spectrum of the input/output data collected from the system. Therefore, it is possible to improve convergence to the global minimum by properly choosing the spectrum of the input signal.

As a next step, we apply this concept to experiment design: we present the use of *input design* as a tool to ensure that the cost function's shape will be amenable to optimization. Typically, input design is formulated as an optimization problem on the input, where several different constraints may be applied. These constraints usually refer to the cost of the identification procedure and/or the quality of the model obtained at the global minimum of the cost function. We propose the addition of a *convergence constraint* to the input design, aiming to obtain an input spectrum such that the cost function will not present local extrema within a given set, which can be made as large as desired. In doing so, we will improve the convergence of iterative algorithms to the global minimum of the PE criterion.

The paper is organized as follows. Section 2 presents basic definitions and the problem formulation. Section 3 presents desired properties of the optimization problem. Conditions about the convergence of the methods are described in Section 4. Input design as a tool to improve the convergence to the global minimum is shown in Section 5. Section 6 presents case studies, and concluding remarks are given in Section 7.

2 Problem formulation

We consider prediction error identification of a linear time-invariant discrete-time single-input single-output "true system":

$$\mathcal{S}: \quad y(t) = G_0(z)u(t) + H_0(z)e(t) \tag{1}$$

where $G_0(z)$ and $H_0(z)$ are the process transfer functions, $u(t)$ is the input and $e(t)$ is white noise with variance σ_e^2. Both transfer functions are rational and proper; furthermore, $H_0(z)$ is monic, i.e. $H_0(\infty) = 1$. To be precise, we shall define $\mathcal{S} \triangleq [G_0(z) \;\; H_0(z)]$. The signal $u(t)$ is assumed to be quasistationary [22]. We also assume that the data is collected in open loop such that $\bar{E}[u(t)e(s)] = 0 \;\; \forall t, s$ where

$$\bar{E}[f(t)] \triangleq \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} E[f(t)]$$

with $E[\cdot]$ denoting expectation [22].

In this paper we consider the identification of linear models

$$y(t) = G(z,\theta)u(t) + H(z,\theta)e(t) \tag{2}$$

where $G(z,\theta)$ and $H(z,\theta)$ are rational and proper transfer functions and θ represents the parameters to be identified. The set $\mathcal{M} \triangleq \{[G(z,\theta)\; H(z,\theta)], \forall\theta \in D_\theta\}$ is called the model set. The search space D_θ is usually constrained to be such that the predictors are BIBO-stable $\forall\theta \in D_\theta$.

We assume that the numerator and denominator of the model are affine functions of the unknown parameters, so that the transfer functions $G(z,\theta)$ and $H(z,\theta)$ have the structure

$$G(z,\theta) = \frac{B^T(z)\theta}{1 + F^T(z)\theta}, \;\; H(z,\theta) = \frac{1 + C^T(z)\theta}{1 + D^T(z)\theta}, \tag{3}$$

where the vectors $B(z), C(z), D(z), F(z) \in \mathbb{R}^n$ are composed of fixed transfer functions and $\theta \in \mathbb{R}^n$ corresponds to the unknown parameter vector of the model. This structure is more general than the structures used in [14, 36, 37] and encompasses all the classical model structures: Box-Jenkins, Output Error (OE), ARMAX, ARX.

We also consider that the system being identified can be exactly described within the model class considered, as formalized by the following assumption.

Assumption 1 $\mathcal{S} \in \mathcal{M}$.

The real system $\mathcal{S}$ belongs to the model set $\mathcal{M}$, i.e. $\exists\, \theta_0 \in D_\theta$ such that

$$G(z,\theta_0) = G_0(z) \;\; \textit{and} \;\; H(z,\theta_0) = H_0(z).$$

$\diamond$

Prediction error identification based on N input-output data consists in finding, among all the models in the pre-specified model set, one that provides the minimum value for the prediction error criterion, that is, one that

solves the following optimization

$$\hat{\theta}_N = \arg\min_{\theta} V_N(\theta) \tag{4}$$

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}[\hat{y}(t,\theta) - y(t)]^2 \tag{5}$$

where the optimal one-step-ahead predictor is given by

$$\hat{y}(t,\theta) = H^{-1}(z,\theta)G(z,\theta)u(t) + (1 - H^{-1}(z,\theta))y(t) \tag{6}$$

PEM has the property [22] that under mild conditions the parameter estimate $\hat{\theta}_N$ converges w.p.1, for $N \to \infty$, to a set

$$\Theta^* = \{\theta^* \triangleq \arg\min_{\theta \in D_\theta} V(\theta)\}, \tag{7}$$

with

$$V(\theta) \triangleq \bar{E}[y(t) - \hat{y}(t,\theta)]^2. \tag{8}$$

If $\mathcal{S} \in \mathcal{M}$, then $\theta_0 \in \Theta^*$. Moreover, under appropriate conditions the cardinality of the set Θ^* is one, that is, θ_0 is the unique global minimum of $V(\theta)$. Necessary and sufficient conditions for uniqueness of the global minimum are local identifiability of the model structure and local informativity of the experiment, as shown recently in [4]. Under these conditions, the parameter error converges to a Gaussian random variable:

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \overset{N\to\infty}{\longrightarrow} \mathcal{N}(0, P_{\theta_0}), \tag{9}$$

where

$$P_{\theta_0} = \left(\frac{1}{\sigma_e^2}\bar{E}\left[\left(\left.\frac{\partial\hat{y}(t,\theta)}{\partial\theta}\right|_{\theta=\theta_0}\right)\left(\left.\frac{\partial\hat{y}(t,\theta)}{\partial\theta}\right|_{\theta=\theta_0}\right)^T\right]\right)^{-1}.$$

In this paper, we study the properties of the optimization problem defined in (4), and approximate the properties of $V_N(\theta)$ by those of $V(\theta)$, due to the uniform convergence of the former to the latter [22, 33, 27]. This approximation is very useful because $V_N(\theta)$ is stochastic and $V(\theta)$ is a deterministic function. By the application of Parseval's theorem this function can be written as:

$$V(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|H^{-1}(e^{j\omega},\theta)H_0(e^{j\omega})\right|^2 \sigma_e^2 + \left|H^{-1}(e^{j\omega},\theta)(G_0(e^{j\omega}) - G(e^{j\omega},\theta))\right|^2 \Phi_u(\omega)d\omega$$

where $\Phi_u(\omega)$ is the spectrum of the input signal $u(t)$.

The solutions of the optimization problems (4) and (7) are the points in the search space D_θ which minimize the respective cost functions, and we will henceforth refer to as the *global minimum*; see the following definition.

Definition 1 *The solution θ^* (resp. $\hat{\theta}_N$) of the optimization problem (7) (resp. (4)) is called a **global minimum** of the function $V(\theta)$ (resp. $V_N(\theta)$). A point θ^+ is called a **local minimum** if it is not a global minimum and $\exists\delta > 0$ such that $V(\theta) \geq V(\theta^+)\ \forall\theta :\ \|\theta - \theta^+\| < \delta, \theta \neq \theta^+$. (or equivalently for $V_N(\theta)$).* ⋄

When a cost function presents local minima, there is a risk that local search methods converge to such points. Even if there are no local minima within the search space, it is often the case that the cost function decreases towards the boundary of the search space D_θ, possibly leading to "convergence" to this boundary. The statistical properties asserted previously are valid only for the global minimum $\hat{\theta}_N$, so whenever this happens, the whole analysis above does not hold, because the identification procedure failed to yield $\hat{\theta}_N$ as its result.

In this paper, we analyze the shape of the cost function $V(\theta)$, assuming that N is large enough so that $V_N(\theta)$ inherits these properties. We will show that this shape is determined by the spectrum of the input signal, and thus propose to apply *input design* as a tool to ensure that the cost function $V(\theta)$ has the appropriate shape, therefore facilitating the convergence to the global minimum.

3 Optimization properties

The optimization problem (7) usually does not have a closed-form solution, so it is common to apply iterative algorithms to solve it. These algorithms have the following generic form:

$$\theta_{i+1} = f(\theta_i) \qquad i = 1, 2, \ldots, \infty.$$

As the iterations proceed, the algorithm may – and should – converge to some point in D_θ; these points are called **limit points**, defined formally below.

Definition 2 *A point $\bar{\theta} \in D_\theta$ is called a **limit point** of an iterative algorithm if for some initial condition*

$$\bar{\theta} = \lim_{i\to\infty} \theta_i.$$

A given algorithm can have – and usually has – different limit points, corresponding to different initial conditions. Most optimization algorithms are based on the gradient of the cost function, and as a consequence have all local minima as limit points (see Chapter 4 in [5]). It is worth mentioning en passant that for some algorithms, Newton-Raphson among them, even maxima are limit points. Therefore, when the cost function presents local minima, each one of these local minima will be a

limit point of the algorithm for some initial conditions and avoiding the highly undesirable convergence to one such non-global minimum may be a very difficult task. It is thus very relevant to determine, in any given identification problem, whether or not the cost function $V(\theta)$ presents local minima. Although there is no definitive answer to this problem in general, the literature has been concerned with this issue for a long time and does provide the answer for some particular cases. The cost function $V(\theta)$ has no local minima in the following cases:

- for ARX models [22] or any linear parameterization (which is standard textbook knowledge, since $V(\theta)$ is quadratic in θ);
- for ARMA models [2];
- for Box-Jenkins models if the denominator of $G(z)$ is a first order polymonial [31];
- for Output Error (OE) models if the input signal is white noise [32].

Examples of convergence to local minima can be found, for instance, in [10, 14, 31]. The example in [31] is an OE model where the input is colored noise. Let us give an example of convergence to the boundary of the search space.

Example 1 *Consider the identification of the "real" system*

$$G_0(z) = \frac{0.033157}{(z-0.7)(z-0.8)(z-0.9)} \qquad H_0(z) = 1 \tag{10}$$

where the noise variance is $\sigma_e^2 = 0.1$. *The following model structure will be used:*

$$G(z,\theta) = \frac{\theta_1}{z^3 + \theta_2 z^2 + \theta_3 z + \theta_4} \qquad H_0(z) = 1 \tag{11}$$

where $\theta = [\theta_1 \;\; \theta_2 \;\; \theta_3 \;\; \theta_4]^T$, *so that* $\mathcal{S} \in \mathcal{M}$ – *Assumption 1 is satisfied. Data will be collected in an experiment where the input* $u(t)$ *consists of* $N = 1,000$ *samples of white noise with variance* $\sigma_e^2 = 1$ *and* D_θ *is the set of parameters that yield stable predictors.*

The present example fits in the last case listed above, since the model is OE and the input signal is white noise, so we know that there are no local minima. However, using the command `oe` *of the System Identification Toolbox from MATLAB [23] we get the estimate:*

$$\bar{\theta} = [-0.0005708 \;\; 1.5911 \;\; 0.2946 \;\; -0.36012]^T;$$

$$G(z,\bar{\theta}) = \frac{-0.0005708}{(z-0.3607)(z^2+1.952z+0.9985)}.$$

which is clearly quite different from the real parameter value θ_0. *Computing the poles of the resulting model, it is seen that there is a complex pair whose module is* 0.9992, *that is, the algorithm stopped very close to the border of the search space* D_θ. *Observe that, whether the algorithms converge to the wrong parameters or not depends on the realization of the noise; in this example the algorithms converged to the wrong parameters in circa* 10% *of trials.*

◇

Most algorithms will converge to the global minimum if their *initial condition* is close enough to it. A set of initial conditions for which the algorithm converges to the global minimum is called a *domain of attraction* (DOA) of the algorithm.

Definition 3 *Assume* θ^* *to be the unique global minimum of the function* $V(\theta) : \mathbb{R}^n \to \mathbb{R}^+$. *A set* $\Omega \subset \mathbb{R}^n$ *is a domain of attraction of an algorithm if* $\lim_{i\to\infty} \theta_i = \theta^*$ $\forall \theta_1 \in \Omega$.

A "good" initialization for an algorithm will be one that is inside a DOA, which will happen if the initial condition is sufficiently close to the global minimum. Different methods for the choice of "good" initial conditions are proposed in the literature, and current commercial identification software packages usually incorporate such methods. The System Identification Toolbox for Matlab uses initial conditions, e.g., based on instrumental variables identification and the University of Newcastle Identification Toolbox [26] applies the Steiglitz-McBride method to this end [34].

An alternative approach, which is what we propose in this paper, is to make the DOA as large as possible, similarly to what has been proposed for controller tuning in [6, 5]. Instead of, or in addition to, looking for a "good" initial condition, we will try to obtain a large DOA. The DOA depends on the algorithm and on the shape of the cost function $V(\theta)$, so to enlarge the DOA one can modify the cost function or/and the algorithm. We will explore in this article how the shape of the cost function affects the DOA and how one can design the experiment to enlarge this set. An important concept related to the shape of the cost function is called *Candidate Domain of Attraction* (CDOA), also known as *Decreasing Euclidean Parameter Error Norm* (DEPEN) region [14, 36], which is defined next.

Definition 4 *Let* θ^* *be the global minimum of the function* $V(\theta)$. *A set* Λ *is a Candidate Domain of Attraction for the function* $V(\theta)$ *if*

$$(\theta - \theta^*)^T \nabla V(\theta) > 0 \;\forall \theta \in \Lambda,\; \theta \neq \theta^*. \tag{12}$$

◇

A nice property of a CDOA is that the gradient of the cost function $\nabla V(\theta)$ is zero only at the global minimum

- there are no local minima or maxima inside it. The most relevant property of a CDOA is that the angle α between $\nabla V(\theta)$ and the vector $(\theta - \theta^*)$ is smaller than $\frac{\pi}{2} rad$ for all θ inside the set, as illustrated in Fig. 1.

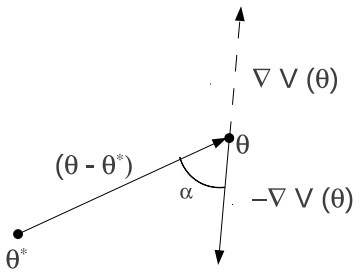

Fig. 1. Angle between $\nabla V(\theta)$ and $(\theta - \theta^*)$.

As a consequence, for all θ inside the CDOA, a sufficiently prudent step in the negative gradient direction of the cost function takes θ closer (in a Euclidean sense) to the global minimum θ^*. Since most algorithms are based on the gradient of the cost function, if the initial condition is inside the CDOA then it is easier to converge to the global minimum, because the negative gradient is always pointing towards the global minimum, and never away from it. One of the most used iterative algorithms is the steepest descent method, for which we can prove the convergence, as seen in the following lemma.

Lemma 3.1 *[5] Let θ^* be the global minimum of the cost function $V(\theta)$ and let $\mathcal{B}(\theta^*)$, defined as*

$$\mathcal{B}(\theta^*) = \left\{\theta : (\theta - \theta^*)^T(\theta - \theta^*) < \beta\right\},$$

be a Candidate Domain of Attraction.

Then the set $\mathcal{B}(\theta^*)$ is a Domain of Attraction of the steepest descent algorithm

$$\theta_{i+1} = \theta_i - \gamma_i \nabla V(\theta_i) \tag{13}$$

provided that

$$0 < \gamma_i < \frac{2(\theta_i - \theta^*)^T \nabla V(\theta_i)}{\nabla V^T(\theta_i)\nabla V(\theta_i)}. \tag{14}$$

$\diamond$

The lemma shows that if the steepest descent algorithm is used with appropriate step sizes - that is, respecting (14) - then any ball that is a Candidate DOA will be also a DOA. Similar results can be obtained for other gradient based optimization algorithms, such as the Newton-Raphson, and the robustness of this convergence property to numerical imprecisions in the algorithm's implementation has also been established [11, 5]. Therefore, it is fair to say that larger Candidate DOAs result in larger DOAs, as a generic property for gradient based algorithms, hence the relevance of the CDOA concept.

4 Convergence Condition

In this section we will explore the properties of the CDOAs to obtain a condition which ensures that a given set is a CDOA.

Theorem 1 *Consider that the model has the structure (3) and that $\mathcal{S} \in \mathcal{M}$. A set $\Lambda \ni \theta_0$ is a CDOA if and only if:*

$$\begin{aligned}
&\frac{1}{\pi}\int_{-\pi}^{\pi} \Phi_u(\omega)\left|H^{-1}(e^{j\omega},\theta)(G_0(e^{j\omega}) - G(e^{j\omega},\theta))\right|^2 \\
&\cdot\Re\left\{\frac{1 + C^T(e^{j\omega})\theta_0}{1 + C^T(e^{j\omega})\theta} + \frac{1 + F^T(e^{j\omega})\theta_0}{1 + F^T(e^{j\omega})\theta} - \frac{1 + D^T(e^{j\omega})\theta_0}{1 + D^T(e^{j\omega})\theta}\right\} \\
&\quad + \sigma_e^2\left|H_0(e^{j\omega})(H^{-1}(e^{j\omega},\theta) - H_0^{-1}(e^{j\omega}))\right|^2 \\
&\quad\cdot\Re\left\{\frac{1 + C^T(e^{j\omega})\theta_0}{1 + C^T(e^{j\omega})\theta}\right\} d\omega > 0
\end{aligned} \tag{15}$$

for all $\theta \in \Lambda$, such that $\theta \neq \theta_0$. In the expression, $\Re\{\cdot\}$ is the real part of a complex number.

Proof. Let us first compute the gradient of the cost function $V(\theta)$ with respect to θ:

$$\begin{aligned}
\nabla V(\theta) = \frac{1}{\pi}\int_{-\pi}^{\pi} \Phi_u \Re\Big\{&\left[H^{-1}(\theta)\left(G_0 - G(\theta)\right)\right]^* \\
&\cdot\left[\nabla H^{-1}(\theta)\left(G_0 - G(\theta)\right) - H^{-1}(\theta)\nabla G(\theta)\right]\Big\} \\
&+ \sigma_e^2 \Re\left\{\left(H_0(H^{-1}(\theta) - H_0^{-1})\right)^* H_0 \nabla H^{-1}(\theta)\right\} d\omega
\end{aligned}$$

where the dependence on the variable ω was omitted to improve readability. Using the structure of the model, we have that

$$\begin{aligned}
\nabla G(z,\theta) &= \frac{B(z)}{1 + F^T(z)\theta} - \frac{B^T(z)\theta F(z)}{(1 + F^T(z)\theta)^2} \\
\nabla H^{-1}(z,\theta) &= \frac{D(z)}{1 + C^T(z)\theta} - \frac{(1 + D^T(z)\theta)C(z)}{(1 + C^T(z)\theta)^2}.
\end{aligned}$$

The gradient of the cost function can be written as

$$\nabla V(\theta) = \frac{1}{\pi}\int_{-\pi}^{\pi} \Phi_u \Re\left\{\left[\left(\frac{1+D^T\theta}{1+C^T\theta}\right)\left(G_0 - \frac{B^T\theta}{1+F^T\theta}\right)\right]^* \cdot \left[\left(\frac{D}{1+C^T\theta} - \frac{(1+D^T\theta)C}{(1+C^T\theta)^2}\right)\left(G_0 - \frac{B^T\theta}{1+F^T\theta}\right) - \left(\frac{1+D^T\theta}{1+C^T\theta}\right)\left(\frac{B}{1+F^T\theta} - \frac{B^T\theta F}{(1+F^T\theta)^2}\right)\right]\right\} + \sigma_e^2 \Re\left\{\left(H_0(H^{-1}(\theta) - H_0^{-1})\right)^* H_0 \left(\frac{D}{1+C^T\theta} - \frac{(1+D^T\theta)C}{(1+C^T\theta)^2}\right)\right\} d\omega. \quad (16)$$

Using the assumption $\mathcal{S} \in \mathcal{M}$ in (16) where $G_0(z) = G(z,\theta_0)$ and $H_0(z) = H(z,\theta_0)$, and reorganizing the terms, we can write

$$(\theta-\theta_0)^T \nabla V(\theta) = \frac{1}{\pi}\int_{-\pi}^{\pi} \Phi_u \left|H^{-1}(\theta)(G_0 - G(\theta))\right|^2 \cdot \Re\left\{\frac{1+C^T\theta_0}{1+C^T\theta} + \frac{1+F^T\theta_0}{1+F^T\theta} - \frac{1+D^T\theta_0}{1+D^T\theta}\right\} + \sigma_e^2 \left|H_0(H^{-1}(\theta) - H_0^{-1})\right|^2 \Re\left\{\frac{1+C^T\theta_0}{1+C^T\theta}\right\} d\omega$$

Using the condition of the theorem we ensure that

$$(\theta-\theta_0)^T \nabla V(\theta) > 0, \forall \theta \in \Lambda,\ \theta \neq \theta_0.$$

which concludes that the set Λ is a CDOA. □

Condition (15) in Theorem 1 expresses the relation between a Candidate DOA and the fundamental elements that are involved in the identification: the transfer functions of the system G_0 and H_0, the model structure $G(z,\theta)$ and $H(z,\theta)$, the input spectrum Φ_u and the variance of the noise σ_e. Among these elements that determine whether or not a given set is a Candidate DOA, the only one that the user can manipulate in order to shape the CDOA is the input spectrum Φ_u. The next section will present the design of the input spectrum with this aim.

5 Input Design

Input Design consists in choosing the input spectrum to be applied in the identification experiment. It is generally posed as an optimization problem [21]

$$\begin{aligned} &\min_{\Phi_u} \text{criterion} \\ &\text{s.t. constraints} \end{aligned} \quad (17)$$

where several constraints may apply. It is usual to include criteria and constraints related to:

- the power of the input signal

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_u(\omega) d\omega < c$$

- the variance of the estimate, represented by the matrix P_{θ_0}:
  - A-optimality - $trace(P_{\theta_0}) < c$
  - D-optimality - $det(P_{\theta_0}) < c$
  - E-optimality - $\lambda_{max}(P_{\theta_0}) < c$
  - L-optimality - $trace(W P_{\theta_0}) < c$

The above are classical optimality constraints criteria, used since the early days of experiment design [25, 35]. More contemporary criteria, also aiming at more sophisticated objectives, have appeared in the past 15 years [16, 3, 17, 18, 20]. An active topic of research is how to obtain useful models to design controllers, and it has been called *identification for control* [16].

The optimization problem (17) is, in general, not convex and infinite dimensional (since the entire input spectrum is to be designed) [20]. However, it is possible in many cases to parametrize the input spectrum and to describe the constraints in such a way that the problem becomes convex and finite dimensional. To do so, the input spectrum is parametrized as

$$\Phi_u(\omega) = \sum_{k=1}^{M} c_k \beta_k(e^{j\omega}) \quad (18)$$

where $\beta_k(e^{j\omega}),\ k = 1,\ldots,M$ are pre-specified basis functions and $c_k,\ k = 1,\ldots,M$ are the parameters to be optimized. The use of a finitely parameterized input spectrum is a standard procedure in input design, which does not constitute a mayor limitation, since it is known that a continuous spectrum can be arbitrarily well approximated by a finite linear combination of basis functions such as $\beta_k(e^{j\omega}) = e^{-j\omega k}, k = 0,1,2,\ldots,M-1$. Note that using this parametrization we have a finite number M of parameters to be optimized. Fortunately, using this parametrization of the input spectrum, several constraints can be rewritten as convex ones, and the optimization problem can be efficiently solved by numerical solvers.

In this paper, we propose to impose an additional condition on the experiment design aiming to improve the convergence of the iterative algorithms to the global minimum of the criterion. The previous section has shown that if the cost function has a large Candidate Domain of Attraction, then it is easier for the algorithm to converge. Hence, if the user could ensure that for a chosen set Λ the condition (15) is respected, then she/he would ensure the convergence for all initial condition inside this set.

The optimization problem (17) with the parametrization (18) and restriction (15) is convex, but it is infinite dimensional because the condition (15) should be verified for all $\theta \in \Lambda, \theta \neq \theta_0$. This kind of problem, that has a finite number of optimisation variables and infinite number of constraints, is called *semi-infinite optimisation problem* in the mathematical programming literature. It has been shown in [10] that this problem can be tackled by using the S-procedure technique, though with conservative results.

In [7], an innovative method called "scenario approach" has been introduced to deal with semi-infinite convex programming at a very general level. Since it is very difficult to deal with the constraint (15) for all $\theta \in \Lambda, \theta \neq \theta_0$, the method proposes that one should naively concentrate attention on just a few constraints on the set Λ. The Scenario Approach [7, 9] technique proposes that Q randomly generated values of $\theta \in \Lambda$ should be selected, generating Q respective scenarios for (15) with these θ to represent the original infinite dimensional set. The number Q should be large, so that the set of the points is sufficiently representative of the whole set Λ.

Specifically, condition (15) is replaced by

$$\phi(\theta_i) + \sum_{k=1}^{M} c_k \psi_k(\theta_i) > 0 \quad i = 1, \ldots, Q \qquad (19)$$

where

$$\psi_k(\theta) = \frac{1}{\pi} \int_{-\pi}^{\pi} \beta_k \left| H^{-1}(\theta)(G_0 - G(\theta)) \right|^2 \cdot \Re \left\{ \frac{1 + C^T\theta_0}{1 + C^T\theta} + \frac{1 + F^T\theta_0}{1 + F^T\theta} - \frac{1 + D^T\theta_0}{1 + D^T\theta} \right\} d\omega$$

$$\phi(\theta) = \sigma_e^2 \frac{1}{\pi} \int_{-\pi}^{\pi} \left| H_0(H^{-1}(\theta) - H_0^{-1}) \right|^2 \Re \left\{ \frac{1 + C^T\theta_0}{1 + C^T\theta} \right\} d\omega$$

and θ_i is selected randomly such that $\theta_i \in \Lambda$ according to a given probability distribution on Λ, which in this paper we assume to be a uniform distribution.

The optimisation problem (17) with the restriction (19) can be efficiently solved with a low computational cost if the number Q is not too large. Observe that the number of scenarios Q is very relevant: the larger Q is, the better condition (19) represents condition (15). However, the computational cost also is a function of Q. The works [7, 9] present a method to design the number of scenarios Q. This method will be described in the sequence.

Before that, we provide some key concepts of the scenario approach:

- Condition (15) is infinite dimensional because it should be verified for all $\theta \in \Lambda$.
- Let us assume we could neglect a fraction of the set Λ verifying condition (15). Let us define that the fraction of the set that would be neglected has probability $\epsilon \in (0, 1)$.
- Observe that the Q scenarios are randomly selected, and therefore there is an uncertainty related to condition (19).

The article [9] describes the relation between the number of scenarios Q and the probability that the constraint (19) is verified for a fraction of Λ.

Lemma 5.1 *[9] Select a* violation parameter $\epsilon \in (0, 1)$ *and a* confidence parameter $\beta \in (0, 1)$. *If (19) is satisfied and*

$$\sum_{i=0}^{n-1} \binom{Q}{i} \epsilon^i (1 - \epsilon)^{Q-i} < \beta$$

then, with probability no smaller than $1 - \beta$*, condition (15) is satisfied for all* Λ *neglecting a fraction of probability at most* ϵ*.*

The above result relates the number of scenarios Q with the probability that the condition will be satisfied to the set Λ if we accept to neglect a fraction of the set proportional to ϵ. This result can be used to design the number of scenarios Q to be used with the input design. For example, one can conclude, using the condition of the lemma, that $1,000$ scenarios are enough to ensure that the conditions will be verified to 98% of the set Λ with 99.92% of confidence. An important observation is that the number of scenarios Q is not very sensitive to β, so we can select β to be such a small number as 10^{-10}, in practice zero, and still Q does not grow significantly [8].

The proposed input design procedure, like most approaches, relies on the knowledge of the true system. This apparent contradiction can be solved in several different ways. For example, it is possible to assume, as prior knowledge, that the true parameter vector lies in a known compact set, and formulate the input design problem as a robust/worst case optimization problem [30, 28]. Another possibility is to solve the input design in an adaptive fashion, by re-designing the input signal recursively, as more data becomes available [13, 29]. For reasons of space, we will not address this alternative in the present paper. However, in Section 6.2 we show, via a case study, that the proposed approach can be based on a preliminary, rough estimate of θ, due to the relative insensitivity of the resulting optimal input to this prior knowledge.

6 Case Studies

In this section we will present two case studies that demonstrate how the method proposed in this article

can be applied. The first case study presents a comparison between the input design with and without the convergence constraints used to improve the convergence of the optimisation algorithms to the global minimum of the criterion. The second case study presents a more realistic example where it is possible to observe the robustness of the method with respect to knowledge of the system parameters.

6.1 Input design with and without convergence constraints

In this example we will compare the convergence when the input signal is computed using input design with and without convergence constraints.

Consider the true system described by

$$G_0(z) = \frac{-0.2073z^{-1} - 0.5083z^{-2} + 0.5564z^{-3} + 0.1592z^{-4}}{1 - 3.125z^{-1} + 3.977z^{-2} - 2.382z^{-3} + 0.5677z^{-4}}$$

$$H_0(z) = 1, \qquad \sigma_e^2 = 20.$$

This model represents a power system consisting of two synchronous generators connected to the power grid, with data taken from [1]. The transfer function $G_0(z)$ describes the relation between the field voltage applied to the first generator and the angular speed of the same generator at a given operating condition.

Let us identify a model to this real system with the Output Error structure

$$G(z,\theta) = \frac{B^T(z)\theta}{1 + F^T(z)\theta} \qquad H(z,\theta) = 1$$

where

$$B(z) = [z^{-1}\ z^{-2}\ z^{-3}\ z^{-4}\ 0\ 0\ 0\ 0\ ]^T,$$

$$F(z) = [0\ 0\ 0\ 0\ z^{-1}\ z^{-2}\ z^{-3}\ z^{-4}]^T.$$

The real system belongs to the model set defined by this model structure, that is, $G_0(z) = G(z,\theta_0)$ with the "real parameter" θ_0 given by

$$\theta_0 = [-0.2073 \quad -0.5083 \quad 0.5564 \quad 0.1592 \quad -3.125 \quad 3.977 \quad -2.382 \quad 0.5677]^T.$$

Before running the experiment, one should choose the input signal to be applied to the process. This input signal will be generated by an experiment design procedure. The number of samples used in the experiment will be $N = 1,000$. To do so, we need to choose a parametrization for the input spectrum. In this example, we will use a discrete spectrum described by

$$\Phi(\omega) = \sum_{k=1}^{40} c_k \left(\delta(\omega - \omega_k) + \delta(\omega + \omega_k)\right)$$

where

$$\omega_k = 10^{\left(2\frac{(k-1)}{40} - 2\right)}\pi.$$

Actually, we will design two different input spectra. The first spectrum will be denoted $\Phi_1(\omega)$ and it will be designed as the solution of the following standard E-optimal input design problem:

$$\max_{c_1,\ldots,c_{40}} \lambda_{min}\left(P_{\theta_0}^{-1}\right) \quad \text{s.t.} \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_u(\omega)d\omega < 1. \tag{20}$$

This optimization problem is convex and it is equivalent to minimize the maximum eigenvalue of the covariance matrix. The squares in Figure 2 are the solution of this problem. The minimum eigenvalue of $P_{\theta_0}^{-1}$ is 0.0044.

The second spectrum will be denoted $\Phi_2(\omega)$ and it will be designed as the solution of the following problem:

$$\max_{c_1,\ldots,c_{40}} \lambda_{min}\left(P_{\theta_0}^{-1}\right) \quad \text{s.t.} \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_u(\omega)d\omega < 1, \tag{21}$$

$$\text{constraint (19).} \tag{22}$$

This optimization problem is convex and it is equivalent to minimize the maximum eigenvalue of the covariance matrix. The circles in Figure 2 are the solution of this problem. The set Λ was chosen as

$$\Lambda = \left\{\theta : (\theta - \theta_0)^T 10^3 P_{\theta_0}^{-1}(\Phi_1)(\theta - \theta_0) < \chi_8^2(95\%)\right\}$$

so that it represents a confidence interval of 95% of the identification procedure obtained with 10^3 samples of a realization of $\Phi_1(\omega)$. The number of points inside Λ used to compose the constraint is $Q = 10^3$ sampled uniformly from Λ. Therefore, the scenario approach ensures that with 99.92% of confidence the condition (15) is satisfied for at least 98% of the set Λ.

This last problem can be formulated as a LMI as described in Section 5, and the spectrum related to the solution of this problem is also plotted in the Figure 2. The solution of the problem is related to a covariance matrix such that the minimum eigenvalue of $P_{\theta_0}^{-1}$ is 0.0037. Note that this minimum eigenvalue is smaller than the value obtained with the first spectrum. The additional

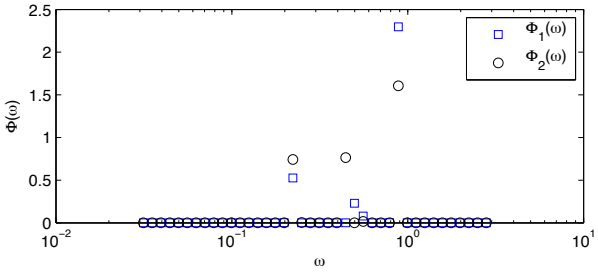

Fig. 2. Input spectra defined by the solution of the optimization problems (20) and (21).

constraint in the problem forces a solution with larger covariance matrix.

The parameter θ has been estimated using the University of Newcastle Identification Toolbox. For each input signal, 100 Monte-Carlo runs have been performed, thus providing 100 parameter estimates. The results of the first signal are summarized in Figure 3 and 4.

In Figure 3 the ordinate axis presents the value of each one of the eight elements of the parameter vector θ and each abscissa corresponds to one Monte Carlo run. The Monte-Carlo runs were re-ordered in terms of the smallest θ. It is seen that after the 19^{th} point of the plot, the values are close to the real parameter value θ_0, indicating that in these 82 runs the algorithm converged to the global minimum of the cost function. However, in the first 18 Monte Carlo runs, the values are close to another point in the parameter space – far away from the real value θ_0 – indicating that the algorithm has yielded a local (nonglobal) minimum of the cost function.

Figure 4 presents the Boxplot of the parameters where the ordinate axis indicates the element of the parameter vector and the abscissa presents the values obtained at each Monte Carlo run. It is possible to see that the algorithm did not convege to the global minimum for every Monte-Carlo run.

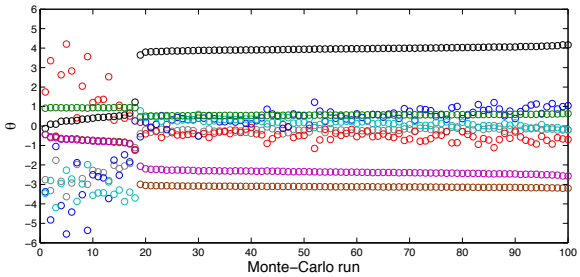

Fig. 3. Parameters estimated at each Monte-Carlo run using $\Phi_1(\omega)$. Each colour represent one parameter of the model.

In Figures 5 and 6 we present the same plots for the second input signal, where it is seen that convergence to the global minimum has been achieved at every run.

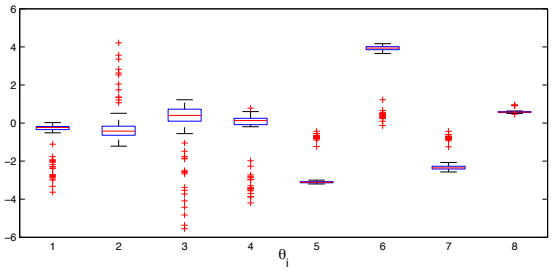

Fig. 4. Boxplot of Monte-Carlo runs with $\Phi_1(\omega)$.

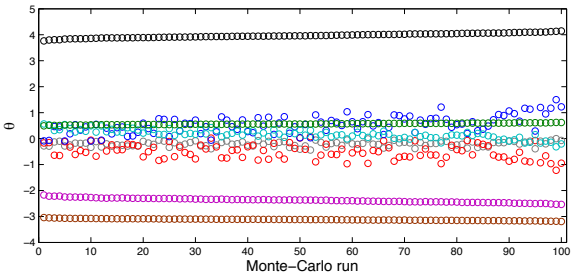

Fig. 5. Parameters estimated at each Monte-Carlo run using $\Phi_2(\omega)$. Each colour represent one parameter.

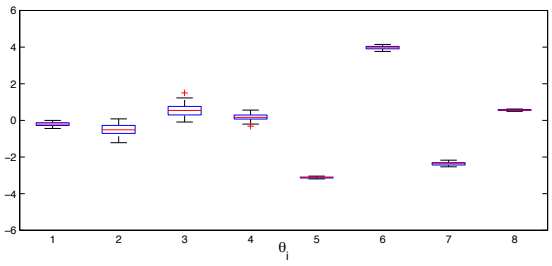

Fig. 6. Boxplot of Monte-Carlo runs with $\Phi_2(\omega)$.

In order to further illustrate the results, let us also compute, for each method, the mean value $\theta_m = \frac{1}{100}\sum_{i=1}^{100}\hat{\theta}_i$ of the model parameters, where each $\hat{\theta}_i$ represents the estimate obtained at the $i-th$ Monte Carlo run.

$$\theta_m(\Phi_1) = [-0.61811 \quad -0.1173 \quad -0.1005 \quad -0.43944 \quad -2.691 \quad 3.323 \quad -2.083 \quad 0.63166]^T$$

$$\theta_m(\Phi_2) = [-0.20804 \quad -0.51126 \quad 0.56363 \quad 0.1546 \quad -3.1264 \quad 3.9807 \quad -2.3859 \quad 0.56916]^T.$$

Comparing these values with the real parameter θ_0 we see that the estimate obtained with the proposed method is, on average, much closer to the real parameter vector.

We have also computed the sample covariance matrix resulting from the 100 Monte-Carlo runs: $P_m =$

$\frac{1}{101}\sum_{i=1}^{100}(\hat{\theta}_i - \theta_m)(\hat{\theta}_i - \theta_m)^T$. When we have used the input spectrum $\Phi_2(\omega)$ the minimum eigenvalue of the P_m^{-1} was 0.00412469, which is close to the theoretical value 0.0037. However, when the spectrum $\Phi_1(\omega)$ was used, the minimum eigenvalue of the P_m^{-1} was computed as 0.00015350, which is much smaller than the theoretical value 0.0044. As expected, the 18 Monte-Carlo runs which do not converge to the global minimum make the sample covariance much larger than expected (close to 30 times larger).

6.2 Robustness of the method

The input design uses in its formulations constraints related to the system parameters which are unknown to the user. In order to design the input spectrum, it is necessary to obtain an estimate to the system parameters that will be used to compute the constraints of the input design. In this section we will explore the robustness of the proposed method to uncertainty on the system parameters. This example shows that the convergence constraints proposed in this article do not seem very sensitive to uncertainty on the estimated model.

Consider the true system described by

$$G_0(z) = \frac{1}{(1-0.8z^{-1})(1-0.85z^{-1})(1-0.9z^{-1})}$$

$$H_0(z) = 1, \qquad \sigma_e^2 = 100.$$

Let us identify a model to this real system with the Output Error structure

$$G(z,\theta) = \frac{B^T(z)\theta}{1+F^T(z)\theta} \qquad H(z,\theta) = 1$$

where

$$B(z) = [1\ 0\ 0\ 0\ 0\ ]^T,$$
$$F(z) = [0\ z^{-1}\ z^{-2}\ z^{-3}]^T.$$

The real system belongs to the model set defined by this model structure, that is, $G_0(z) = G(z,\theta_0)$ with the "real parameter" θ_0 given by

$$\theta_0 = [1 \quad -2.55 \quad 2.165 \quad -0.612]^T.$$

Before performing the experiment design, let us use white noise with unitary variance as input signal. We have run 1,000 Monte-Carlo experiments to observe the convergence of the identification algorithms. We have used the `oe` command of the *System Identification Toolbox* to estimate the model parameters. Figure 7 presents the same plot for the white noise input signal.

In Figure 7 the ordinate axis presents the value of each one of the four elements of the parameter vector θ and each abscissa corresponds to one Monte Carlo run. The Monte-Carlo runs were re-ordered in terms of the smallest θ. With the white noise input signal 118 runs do not converg to the global minimum.

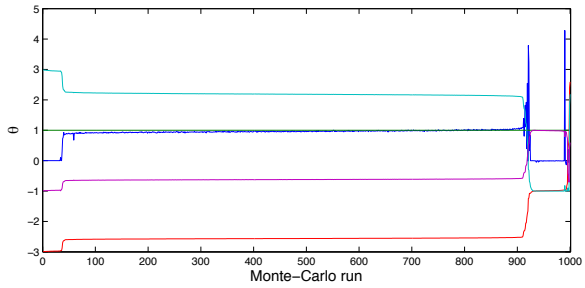

Fig. 7. Parameters estimated at each Monte-Carlo run using white noise as input spectrum. Each colour represent one parameter of the model.

One of these Monte-Carlo runs was used to obtain a model using the instrumental variable technique. The following model was obtained:

$$G_{IV}(z) = \frac{1.1112}{(1-0.9906z^{-1})(1-1.379z^{-1}+0.4928z^{-2})}.$$

The transfer function $G_{IV}(z)$ is just an rough approximation of $G_0(z)$, but it will be shown in the sequl that this approximation is enough to design the input signal. The Figure 8 shows the Bode Diagram of $G_{IV}(z)$ and $G_0(z)$ for comparison. The step responses of $G_{IV}(z)$ and $G_0(z)$ are shown in Figure 9.

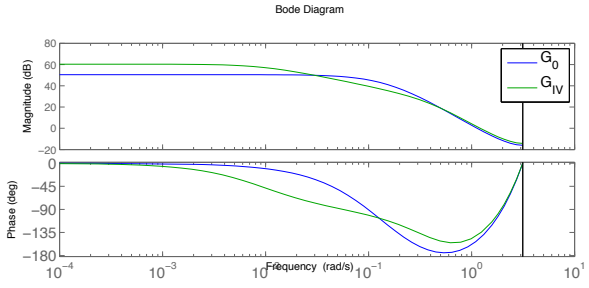

Fig. 8. Bode plot of $G_{IV}(z)$ and $G_0(z)$

The model $G_{IV}(z)$ will be used in the input design to generate an input signal with 1,000 samples. To do so, we need firstly choose a parametrization for the input spectrum. In this example, we will use a discrete spectrum described by

$$\Phi_u(\omega) = \sum_{k=1}^{40} c_k\left(\delta(\omega-\omega_k)+\delta(\omega+\omega_k)\right)$$

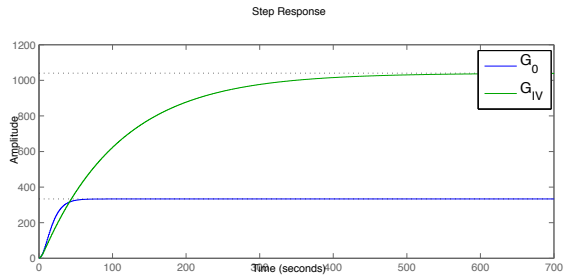

Fig. 9. Step response of $G_{IV}(z)$ and $G_0(z)$

where

$$\omega_k = 10^{\left(2\frac{(k-1)}{40}-2\right)}\pi.$$

The spectrum $\Phi_u(\omega)$ will be designed as the solution of the following problem:

$$\max_{c_1,\ldots,c_{40}} \lambda_{min}\left(P_{\theta_0}^{-1}\right)$$
$$\text{s.t.} \frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_u(\omega)d\omega < 1. \quad (23)$$
$$\text{constraint (19)} \quad (24)$$

This optimization problem is convex and it is equivalent to minimize the maximum eigenvalue of the covariance matrix. The convergence constraint (19) was included in this problem as proposed in this work. The set Λ was chosen as

$$\Lambda = \left\{\theta : (\theta-\theta_0)^T(\theta-\theta_0) < 0.0001\right\}$$

so that it represents a ball with radius 0.01. The number of points inside Λ used to compose the constraint is $Q = 1,000$, where the points were sampled uniformly from Λ. Therefore, the scenario approach ensures that with 99.92% of confidence the condition (15) is satisfied for at least 98% of the set Λ.

This problem can be formulated as a LMI as described in Section 5, and the spectrum related to the solution of this problem is plotted in the Figure 10.

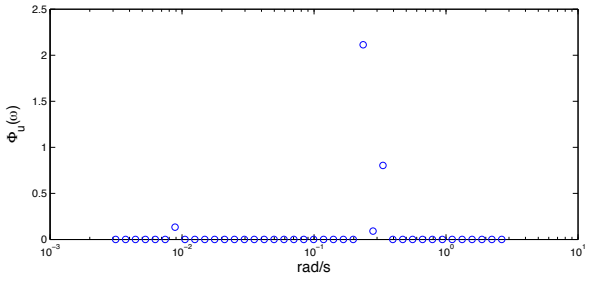

Fig. 10. Spectrum.

Again, the System Identification Toolbox was used and 1000 Monte-Carlo runs were performed, thus providing 1000 parameter estimates. In Figure 11 the ordinate axis presents the value of each one of the four elements of the parameter vector θ and each abscissa corresponds to one Monte Carlo run. The Monte-Carlo runs were re-ordered in terms of the smallest θ. In 4 Monte Carlo runs the values are close to another point in the parameter space – far away from the real value θ_0 – indicating that the algorithm has yielded a local (nonglobal) minimum of the cost function. All the other 996 runs the algorithm converged to the global minimum of the criterion.

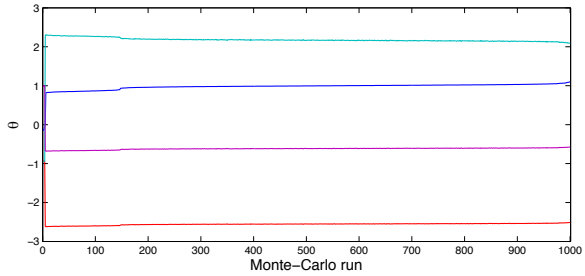

Fig. 11. Parameters estimated at each Monte-Carlo run using designed input spectrum. Each colour represent one parameter of the model.

7 Conclusion

In this paper we proposed a solution to the problem of convergence to "false minima" in identification through the Prediction Error Method. This solution is based on a sufficient condition for the convergence to the global minimum of the Prediction Error cost function, a condition which has been given and demonstrated in the paper. The spectra of the input signals are key factors in this convergence condition, so our approach consists in designing the spectra such that the convergence condition is respected. This can be accomplished by including additional constraints in the input design problem, and we have provided a way to include these constraints in a convex manner. Case studies show that this approach to shape the PEM cost function can be very effective in guaranteeing convergence of the PEM to its global optimum.

References

[1] P. M. Anderson and A. A. Fouad. *Power System Control and Stability*. IEEE Press, 1977.

[2] K. J. Åström and T. Söderström. Uniqueness of the maximum likelihood estimates of the parameters of an arma model. *IEEE Transactions on Automatic Control*, 18(6):769–773, 1974.

[3] M. Barenthin and H. Hjalmarsson. Identification and control: Joint input design and state feedback with ellipsoidal parametric uncertainty via LMIs. *Automatica*, 44(2):543–551, 2008.

[4] A. S. Bazanella, X. Bombois, and M. Gevers. Necessary and sufficient conditions for uniqueness of the

minimum in prediction error identification. *Automatica*, 48(8):1621 – 1630, 2012.

[5] A. S. Bazanella, L. Campestrini, and D. Eckhard. *Data-driven Controller Design: The H2 Approach*. Springer, Netherlands, 2012.

[6] A. S. Bazanella, M. Gevers, L. Miskovic, and B. D. O. Anderson. Iterative minimization of H_2 control performance criteria. *Automatica*, 44(10):2549–2559, 2008.

[7] G. Calafiore and M. C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *IMathematical Programming*, 102(1):25–46, 2005.

[8] M. Campi, S. Garatti, and M. Prandini. The scenario approach for systems and contol design. In *Proc. 17th IFAC World congress*, Seoul, Korea, 2008.

[9] M. C. Campi, S. Garatti, and M. Prandini. The scenario approach for systems and control design. *Annual Reviews in Control*, 33(2):149 – 157, 2009.

[10] D. Eckhard and A. S. Bazanella. On the global convergence of identification of output error models. In *Proc. 18th IFAC World congress*, Milan, Italy, 2011.

[11] D. Eckhard and A. S. Bazanella. Robust convergence of the steepest descent method for data-based control. *International Journal of Systems Science*, iFirst:1–7, 2011.

[12] S. Garatti and R. R. Bitmead. On resampling and uncertainty estimation in linear system identification. *Automatica*, 46(5):785 – 795, 2010.

[13] L. Gerencsér, H. Hjalmarsson, and J. Mårtensson. Identification of arx systems with non-stationary inputs - asymptotic analysis with application to adaptive input design. *Automatica*, 45(3):623–633, 2009.

[14] G. C. Goodwin, J. C. Agüero, and R. Skelton. Conditions for local convergence of maximum likelihood estimation for armax models. In *Proc. 13th IFAC Symposium on System Identification*, pages 797–802, Rotterdam, Holand, August 2003.

[15] C. Grossmann, C. N. Jones, and M. Morari. System identification via nuclear norm regularization for simulated moving bed processes from incomplete data sets. In *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pages 4692 –4697, dec. 2009.

[16] H. Hjalmarsson. From experiment design to closed-loop control. *Automatica*, 41(3):393–438, 2005.

[17] H. Hjalmarsson, M. Gevers, and F. De Bruyne. For model-based control design, closed loop identification gives better performance. *Automatica*, 32(12):1659–1673, 1996.

[18] H. Hjalmarsson and H. Jansson. Closed loop experiment design for linear time invariant dynamical systems via LMIs. *Automatica*, 44(3):623–636, 2008.

[19] H. Hjalmarsson, J. Welsh, and C. Rojas. Identification of box-jenkins models using structured ARX models and nuclear norm relaxation. In *Proc. 16th IFAC Symposium on System Identification*, pages 322–327, Brussels, Belgium, July 2012.

[20] H. Jansson and H. Hjalmarsson. Input design via LMIs admitting frequency-wise model specifications in confidence regions. *Automatic Control, IEEE Transactions on*, 50(10):1534–1549, oct. 2005.

[21] K. Lindqvist and H. Hjalmarsson. Identification for control: adaptive input design using convex optimization. In *Decision and Control, 2001. Proceedings of the 40th IEEE Conference on*, volume 5, pages 4326–4331, 2001.

[22] L. Ljung. *System identification: theory for the user*. Prentice Hall, Upper Saddle River, NJ, USA, 1999.

[23] L. Ljung. *System Identification Toolbox - for use with MATLAB, User's Guide*. The Mathworks, 4th edition edition, 2000.

[24] L. Ljung. Perspectives on system identification. *Annual Reviews in Control*, 34(1):1–12, 2010.

[25] R. Mehra. Optimal input signals for parameter estimation in dynamic systems–survey and new results. *Automatic Control, IEEE Transactions on*, 19(6):753–768, dec 1974.

[26] B. Ninness, A. Wills, and S. Gibson. The university of newcastle identification toolbox (unit). In *IFAC World Congress*, pages 1–6, jul 2005.

[27] R. Pintelon and J. Schoukens. *System Identification: A Frequency Domain Approach*. IEEE Press, New York, 2001.

[28] C. R. Rojas, J. C. Agüero, J. S. Welsh, G. C. Goodwin, and A. Feuer. Robustness in experiment design. *EEE Transactions on Automatic Control*, 57(4):860–874, 2012.

[29] C. R. Rojas, H. Hjalmarsson, L. Gerencsér, and J. Mårtensson. An adaptive method for consistent estimation of real-valued non-minimum phase zeros in stable lti systems. *Automatica*, 47(7):1388–1398, 2011.

[30] C. R. Rojas, J. S. Welsh, G. C. Goodwin, and A. Feuer. Robust optimal experiment design for system identication. *Automatica*, 43(6):993–1008, 2007.

[31] T. Söderström. On the uniqueness of maximum likelihood identification. *Automatica*, 11(2):193–197, 1975.

[32] T. Söderström and P. Stoica. Some properties of the output error method. *Automatica*, 18(1):93–99, 1982.

[33] T. Söderström and P. Stoica. *System Identification*. Prentice-Hall International, Hemel Hempstead, Hertfordshire, 1989.

[34] K. Steiglitz and L. Mcbride. A technique for the identification of linear systems. *Automatic Control, IEEE Transactions on*, 10(4):461–464, 1965.

[35] M. Zarrop. *Optimal Experiment Design for Dynamic System Identification*. Springer Verlag, Berlin, 1979.

[36] Y. Zou and W. P. Heath. Conditions for attaining global minimum in maximum likelihood sys-

tem identification. In *Proc. 15th IFAC Symposium on System Identificationl*, pages 1110–1115, Saint-Malo, France, July 2009.

[37] Y. Zou and W. P. Heath. Global convergence conditions in maximum likelihood estimation. *International Journal of Control*, 85(5):475–490, 2012.

[38] Y. Zou, X. Tang, and Z. Ding. Counter-examples design to global convergence of maximum likelihood estimators. In *UKACC International Conference on Control*, pages 864–869, 2012.