Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Finite-time experiment design with multisines

X. Bombois * **M. Barenthin** ** **P.M.J. Van den Hof** *

* *Delft Center for Systems and Control, Delft University of Technology, Delft (e-mail: x.j.a.bombois@tudelft.nl).*
** *Automatic Control, School of Electrical Engineering, KTH, Stockholm.*

Abstract: When the number of data used for an identification experiment is small, the data contain transient effects and these effects contribute to the accuracy of the identified model. Consequently, when designing the optimal input signal for an identification experiment in the case of small data set, these transient effects have to be taken into account. In this paper, we present a methodology for optimal experiment design which, unlike other approaches such as the asymptotic approach, takes explicitly the transient effects into account. This paper is restricted to model structures that are linear in the parameter and to multisine input signals.

Keywords: optimal experiment design; system identification; prediction error methods; convex optimization; sine waves

1. INTRODUCTION

This paper deals with the optimal design of the input signal of an identification experiment when the number N of data available for this identification is (possibly) small.

The typical approach to optimal experiment design has been to maximize the accuracy of the identified model for a given experiment time and under prespecified constraints on input power (see e.g. (Goodwin and Payne, 1977; Ljung, 1999)). In recent contributions this trade-off has been addressed from the dual perspective; this is the dual perspective that we will here consider. In this novel framework, assuming that the experiment duration N is fixed, the optimal (open-loop) identification experiment is defined as the experiment whose input signal power is minimized under the constraint that the accuracy of the identified model $G(z,\hat{\theta}_N)$ remains above some prespecified threshold R_{adm} (Bombois et al., 2006):

$$\begin{aligned} &arg \min_{u(1)\ldots u(N)} \frac{1}{N}\sum_{t=1}^{N} u^2(t) \\ &subject\ to\ P_\theta^{-1} \geq R_{adm} \end{aligned} \tag{1}$$

As can be seen in (1), the accuracy of $G(z,\hat{\theta}_N)$ is here measured (for simplicity) via the inverse of the covariance matrix P_θ of $\hat{\theta}_N$ and R_{adm} is therefore also a matrix. However, more sophisticated accuracy constraints can also be used (see e.g. (Jansson and Hjalmarsson, 2005; Bombois et al., 2006)). Due to the fact that expressions for P_θ can generally be determined only for the case where the number N of data tends to infinity, this problem is generally treated under this assumption (Ljung, 1999; Jansson and Hjalmarsson, 2005; Bombois et al., 2006). The experiment design problem is then solved by determining the power spectrum of the input signal using this asymptotic expression of P_θ and assuming thus that this asymptotic expression represents a good approximation of the actual variance obtained with a finite number of data. However, since, in this paper, we are particularly interested in the case where N is small, the asymptotic theory is not the most appropriate.

Expressions for P_θ that are also valid for finite (small) N can be determined if the model structure used for the identification is linear in the parameter (FIR, Laguerre model structures) (Ljung, 1999). In (Bombois and Gilson, 2006), it was shown that (1) could be made convex by determining $u(t)$ within the (restricted) class of PRBS input sequences. In the present paper, our objective will be to extend the results in (Bombois and Gilson, 2006) by considering a more extended class of signals (Pintelon and Schoukens, 2001) i.e. the class of multisines (sum of sinusoids) having a fundamental period exactly equal to the number N of data. More precisely, assuming that the phases of the multisine are given, we compare different (convex) methods to determine the amplitudes of the multisine solving (1). For the considered class of input signals, the power $\frac{1}{N}\sum_{t=1}^{N} u^2(t)$ can be written as a linear function of the squared amplitudes of the multisine. Unfortunately, P_θ^{-1} is not affine in the squared amplitudes. Consequently, to be able to use convex optimization to solve (1), we cannot use the exact expression for P_θ^{-1}, but we must instead use an approximation of this matrix which is affine in the squared amplitudes. This can be done by using the asymptotic theory or by supposing that the system has reached steady-state (see e.g. (Jansson, 2004)[Corollary 4.2]). In fact, these approximations neglect the transient effects. Transient effects are nevertheless always present when the system is excited for identification purpose after having been at rest (the case considered in this paper) and they contribute to a better identification by increasing the amount of excitation. Consequently, the available ap-

proximations of P_θ^{-1} obtained by neglecting the transient effects are therefore not the most appropriate when N is small. We therefore introduce new approximations of P_θ^{-1} as affine functions of the squared amplitudes that explicitly take these transient effects into account. Based on these affine approximations of P_θ^{-1}, the experiment design problem can be solved using convex optimization. However, due to the approximation of P_θ^{-1}, it is not guaranteed that the constraint in (1) is effectively met with the determined signal. Another contribution of this paper is therefore to introduce a methodology to scale the determined input signal in order to ensure that the constraint $P_\theta^{-1} \geq R_{adm}$ is effectively met. The input signal determined using this convex approach in two steps is suboptimal. However, even though it is absolutely not a necessity, the resulting multisine can thereafter be refined by using it as the starting value for a non-linear optimization algorithm (Goodwin and Payne, 1977) which will allow to further approach the global optimum of (1). Note that a non-linear optimization algorithm generally needs to start sufficiently near the global optimum to lead to a reliable result. The multisine obtained via our suboptimal algorithm is therefore a good candidate for this starting value.

2. IDENTIFICATION FRAMEWORK

We consider the identification of a linear time-invariant single input single output system:

$$\mathcal{S}: \quad y(t) = G(z,\theta_0)u(t) + e(t) \tag{2}$$

with $G(z,\theta_0)$ a transfer function with an unknown parameter vector $\theta_0 \in \mathbf{R}^k$ that we wish to identify. Here, to simplify the notations, the additive noise $e(t)$ is assumed to be the realization of a Gaussian white noise of variance σ_e^2. However, colored noise can be treated as well (see Section 7). We further assume that $G(z,\theta_0)$ is linear in θ_0 i.e. $G(z,\theta_0) = \Lambda(z)\theta_0$ where $\Lambda(z) = (\Lambda_1(z)\ \Lambda_2(z)...\Lambda_k(z))$ is a row vector containing the first k elements of a series of basis functions $\Lambda_i(z)$ $(i = 1...\infty)$. Examples of those basis functions are the FIR basis with $\Lambda_i(z) = z^{-i}$ and the Laguerre basis with $\Lambda_i(z) = z^{-1}\frac{\sqrt{1-\xi^2}}{1-\xi z^{-1}}\left(\frac{-\xi+z^{-1}}{1-\xi z^{-1}}\right)^{i-1}$ for some real-valued pole ξ (see Heuberger et al. (2005) for other examples and a discussion on the generality of this parametrization).

We consider the case where the system (2) is at rest for $t \leq 0$ i.e. $u(t) = 0$ for $t \leq 0$ and where the data required for the identification of a model $G(z,\hat{\theta}_N) = \Lambda(z)\hat{\theta}_N$ of the true system $G(z,\theta_0)$ are collected by applying, from $t = 1$ to $t = N$, an input sequence $u(t)$ $(t = 1...N)$ to (2) and by measuring the corresponding output signal $y(t)$ $(t = 1...N)$. Based on this (possibly small) input-output data set, the parameter vector $\hat{\theta}_N$ of the identified model is computed in the prediction error framework (Ljung, 1999):

$$\hat{\theta}_N = M^{-1}\frac{1}{N}\sum_{t=1}^{N}\phi(t)y(t) \tag{3}$$

$$with\ M = \frac{1}{N}\sum_{t=1}^{N}\phi(t)\phi^T(t)\ \ and\ \ \phi^T(t) = \Lambda(z)u(t). \tag{4}$$

By using the fact that $y(t) = \phi^T(t)\theta_0 + e(t)$, (3) can be rewritten as: $\hat{\theta}_N = \theta_0 + M^{-1}\frac{1}{N}\sum_{t=1}^{N}\phi(t)e(t)$. Since $e(t)$ is supposed to be Gaussian white noise, we see that the identified parameter vector $\hat{\theta}_N$ is normally distributed with mean θ_0 and its covariance matrix is given by:

$$P_\theta = \frac{\sigma_e^2}{N}\left(\frac{1}{N}\sum_{t=1}^{N}\phi(t)\phi^T(t)\right)^{-1} = \frac{\sigma_e^2}{N}M^{-1} \tag{5}$$

It is important to note that the statistical properties of $\hat{\theta}_N$ are exact even for small data sets. Note also that P_θ is dependent on the data set only via the input sequence $u(t)$ $(t = 1...N)$. In the sequel, the matrix M in (5) will be called, with some abuse, the information matrix and we will assume that we know (an estimate of) σ_e^2.

3. EXPERIMENT DESIGN

In this paper, our objective is to determine an optimal input sequence $\{u(t)\ \ t = 1...N\}$ for the identification of $\hat{\theta}_N$. The optimal input sequence is here defined as the solution of the optimal experiment design problem (1). As mentioned in the introduction, we consider the case where the input sequence has to be determined in the class of multisines whose fundamental period is equal to the number N of data. This class is given by:

$$u(t) = \sum_{i=1}^{m} A_i\ cos(i\omega_0 t + \alpha_i)\ \ \text{with}\ \omega_0 = \tfrac{2\pi}{N} \tag{6}$$

where the amplitudes $A_i \geq 0$ $(i = 1...m)$ are the variables to be designed in order to achieve the required objective while the phase shifts α_i $(i = 1...m)$ are supposed given (e.g. randomly chosen). The fundamental frequency ω_0 is chosen equal to $2\pi/N$ in such a way that the input sequence $\{u(t) \mid t = 1...N\}$ exactly represents one period of $u(t)$ $(t = -\infty...+\infty)$. The number m of cosines in (6) is a user-choice and can be chosen up to $N/2$. Note that, with $m = N/2$, the parametrization would allow to represent any zero-mean signal of length N if the phase shifts could be tuned with the amplitudes.

With an input signal such as in (6), the power of the input sequence which is the cost function of the considered optimal experiment design problem (1) can be rewritten as:

$$\frac{1}{N}\sum_{t=1}^{N}u^2(t) = \sum_{i=1}^{m}\frac{A_i^2}{2} \tag{7}$$

and the information matrix M (see (5)) has a particular structure:

$$M = \frac{1}{N}\sum_{t=1}^{N}\phi(t)\phi^T(t) = \sum_{i=1}^{m}\sum_{j=1}^{m}A_i\ A_j\ M_{ij} \tag{8}$$

$$\text{with}\ \ M_{ij} = \frac{1}{N}\sum_{t=1}^{N}\phi_i(t)\phi_j^T(t) \tag{9}$$

where $\phi_i^T(t) = \Lambda(z)cos(i\omega_0 t + \alpha_i)$ $(i = 1...m)$.

Based on the above expressions, the optimal multisine for the identification is the multisine having the amplitudes A_i which solve the following optimization problem:

$$arg \min_{A_i \ (i=1...m)} \sum_{i=1}^{m} \frac{A_i^2}{2}$$
$$subject\ to\ \frac{N}{\sigma_e^2} \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{m} A_i\ A_j\ M_{ij}}_{=M} \geq R_{adm} \tag{10}$$

This optimization problem is nonconvex since M is not affine in the squared amplitudes A_i^2 $(i = 1...m)$ like the cost function. If it would have been the case, (10) would boil down to an LMI-based optimization problem with decision variables A_i^2 $(i = 1...m)$. Since solving this problem directly is not tractable, we will proceed in two steps. In a first step, we will replace M in (10) by a (close) approximation M_{appro} which is well affine in A_i^2 $(i = 1...m)$ and solve the modified optimization problem. The input sequence $u_1(t)$ obtained by solving this modified problem is designed to meet the approximated constraint $\frac{N}{\sigma_e^2} M_{appro} \geq R_{adm}$ and thus not the actual accuracy constraint $\frac{N}{\sigma_e^2} M \geq R_{adm}$. As a consequence, $u_1(t)$ could be not enough powerful to meet this constraint or, conversely, the power of $u_1(t)$ could be larger than strictly necessary for this constraint to hold. The second step will therefore consist of scaling $u_1(t)$ in order to meet (just) $\frac{N}{\sigma_e^2} M \geq R_{adm}$. In this two step method, the first step is classical while the second, even though obvious, seems new. This procedure is summarized in the following algorithm:

Algorithm 1. Consider the optimal experiment design problem (10) and a symmetric and positive-definite matrix $M_{appro} \in \mathbf{R}^{k \times k}$ that is affine in A_i^2 $(i = 1...m)$ and that is such that $M_{appro} \approx M$. A (suboptimal) solution of this experiment design problem is given by the multisine $u(t) = \xi^{opt} u_1(t)$ where ξ^{opt} and $u_1(t)$ are defined as follows:

1) The signal $u_1(t)$ is a multisine (6) whose amplitudes A_i $(i = 1...m)$ are determined via the following convex optimization problem with decision variables A_i^2 $(i = 1...m)$

$$arg \min_{A_i^2 \ (i=1...m)} \sum_{i=1}^{m} \frac{A_i^2}{2}$$
$$subject\ to\ \frac{N}{\sigma_e^2} M_{appro} \geq R_{adm} \tag{11}$$

2) The scaling factor ξ^{opt} is a scalar determined as the solution of the following convex optimization problem:

$$\xi^{opt} = arg \min_{\xi} \xi$$
$$subject\ to\ \frac{N}{\sigma_e^2}\ M(\xi u_1) \geq R_{adm} \tag{12}$$

where $M(u)$ represents the information matrix M (see (8)) that is obtained for a signal $u(t)$ $(t = 1...N)$ and $u_1(t)$ is the multisine determined via (11). ■

The correction factor ξ^{opt} can be easily determined since we have that $M(\xi u) = \xi^2 M(u)$. It is important to note that, due to the approximation $M_{appro} \approx M$, the input sequence determined via Algorithm 1 is only a suboptimal solution of the actual experiment design problem (10). However, the better the approximation $M_{appro} \approx M$, the closer to the optimum the determined input sequence will be. In the sequel, we will therefore investigate how an appropriate approximation for the information matrix can be found. Recall that $M_{appro} \approx M$ must be affine in the squared amplitudes A_i^2 $(i = 1...m)$. A very common way to obtain such a matrix M_{appro} is to use the asymptotic theory i.e. to assume that $N \to \infty$ (see e.g. (Bombois et al., 2006)). However, asymptotic approximations could not be appropriate for this paper since the number N of data is possibly small. Another possibility to obtain an approximation M_{appro} of M affine in the squared amplitudes is to neglect the transient effects generated by the excitation of the system. We review this in the next section and we show that the approximation M_{appro} obtained as such is in fact not different from the one we obtain by assuming that $N \to \infty$. Finally, in Section 5, we propose an approximation M_{appro} which is more appropriate to the case where N is (possibly) small.

4. EXPERIMENT DESIGN WITH THE STATIONARITY ASSUMPTION

In this section, we present the approximation M_{appro} for the information matrix (8) which is obtained by neglecting the transient effects due to the excitation of the true system i.e. by assuming that the signal vector $\phi(t)$ present in the expression (8) of M has already reached its steady-state. This assumption has been used e.g. in (Jansson, 2004)[Corollary 4.2]. Under this assumption (approximation), the matrices M_{ij} for $i \neq j$ (see (8)) are all identically equal to 0 and the matrices M_{ii} are equal to the real part of $\frac{\Lambda(e^{j(i\omega_0)})\Lambda^*(e^{j(i\omega_0)})}{2}$ $(i = 1...m)$. To show this, notice that, when $\phi_i^T(t) = \Lambda(z)cos(i\omega_0 t + \alpha_i)$ has reached steady state, its p^{th} element is, in the time interval $[1\ N]$, exactly equal to $\nu_{p,i}\ cos(i\omega_0 t + \lambda_{p,i})$ with $\nu_{p,i} = |\Lambda_p(e^{j(i\omega_0)})|$ and $\lambda_{p,i} = \alpha_i + \angle\Lambda_p(e^{j(i\omega_0)})$. Consequently, the entry (p, l) of the matrix M_{ij} is:

$$\nu_{p,i}\ \nu_{l,j}\ \frac{1}{N}\sum_{t=1}^{N}(cos(i\omega_0 t + \lambda_{p,i}) \times cos(j\omega_0 t + \lambda_{l,j})) =$$
$$\frac{\nu_{p,i}\ \nu_{l,j}}{2}\ \frac{1}{N}\sum_{t=1}^{N}(cos((i+j)\omega_0 t + \lambda_{p,i} + \lambda_{l,j})$$
$$+cos((i-j)\omega_0 t + \lambda_{p,i} - \lambda_{l,j}))$$

which is identically 0 when $i \neq j$ since N represents a period for each cosine at harmonics of $\omega_0 = \frac{2\pi}{N}$. Consequently, $M_{ij} = 0$ for all $i \neq j$. From the above equation, we see also that the entry (p, l) of the matrix M_{ii} is given by $\frac{\nu_{p,i}\ \nu_{l,i}}{2} cos(\lambda_{p,i} - \lambda_{l,i})$ which is equal to $Re\left(\frac{\Lambda_p(e^{j(i\omega_0)})\Lambda_l^*(e^{j(i\omega_0)})}{2}\right)$. Thus:

$$M_{ii} = Re\left(\frac{\Lambda(e^{j(i\omega_0)})\Lambda^*(e^{j(i\omega_0)})}{2}\right) \tag{13}$$

Consequently, the information matrix that would be obtained if $\phi(t)$ had effectively reached steady-state would be:

$$M_{st} = \sum_{i=1}^{m} A_i^2 \underbrace{Re\left(\frac{\Lambda(e^{j(i\omega_0)})\Lambda^*(e^{j(i\omega_0)})}{2}\right)}_{M_{ii}} \quad (14)$$

We denote this matrix by M_{st} to distinguish it from the actual information matrix M. The matrix M_{st} is indeed only an approximation of M since the transient effects that are necessarily present in the vector $\phi(t)$ are here neglected. We observe that M_{st} is linear in A_i^2 $(i = 1...m)$. Consequently, this matrix can be used as the matrix M_{appro} in Algorithm 1 to determine the input sequence for the identification. We finish this section by two important comments.

Comment 1. The matrix M_{st} is nevertheless equal to the actual information matrix M in the particular case (not considered here) where, even though $u(t)$ and $y(t)$ are measured between $t = 1$ and $t = N$, the signal $u(t)$ has been applied since $t = -\infty$. This fact will be used to determine via an example if there is something to gain (or to loose) from an identification point of view by collecting the data only when they have reached steady-state compared to the situation where the data are collected with $u(t) = 0$ for $t \leq 0$ and which is considered in this paper.

Comment 2. There is in fact no difference between the approximation $M_{appro} = M_{st}$ and the approximation $M_{appro} = M_{as}$ obtained via the asymptotic theory. To evidence this, let use recall that the asymptotic information matrix is given by $M_{as} = \bar{E}\phi(t)\phi^T(t)$ and that, using Parseval theorem, it can be rewritten as: $M_{as} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\Lambda(e^{j\omega})\Lambda^*(e^{j\omega})\Phi_u(\omega)$ where $\Phi_u(\omega)$ is the power spectrum of the multisine (6) i.e. $\Phi_u(\omega) = \frac{\pi}{2}\sum_{i=1}^{m} A_i^2 \ (\delta(\omega - i\omega_0) + \delta(\omega + i\omega_0))$ Replacing the expression of $\Phi_u(\omega)$ within the expression of M_{as} yields (14). Consequently, when using $M_{appro} = M_{st}$ in Algorithm 1 for small N, the designed input sequence could be far from optimal.

5. EXPERIMENT DESIGN TAKING THE TRANSIENT EFFECTS INTO ACCOUNT

The main drawback of the approximation $M \approx M_{st}$ presented in Section 4 is that the transient effects are neglected while they can be important certainly when N is small. In this section, we present new approximations M_{appro} for M which will take these transient effects into account. For this purpose, note that the information matrix M in (8) can be rewritten as follows:

$$M = \left(\sum_{i=1}^{m} A_i^2 M_{ii}\right) + \left(\sum_{i=1}^{m}\sum_{j=1(j\neq i)}^{m} A_i A_j M_{ij}\right) \quad (15)$$

Remember that M_{appro} must be affine in the squared amplitudes A_i^2 $(i = 1...m)$. Consequently, by looking at (15), we can straightforwardly obtain the following approximation:

$$M_{appro,1} = \sum_{i=1}^{m} A_i^2 M_{ii} \quad (16)$$

In (16), the elements of M_{ii} are computed with expression (9) with $j = i$. The transient effects present in the vector $\phi(t)$ are thus taken into account: we do not indeed use the expression (13) for M_{ii} as in the previous section. However, the transient effects are only partially reflected in (16) since the matrices M_{ij} for $i \neq j$ are neglected. This approximation can be justified by the fact that, when N is taken larger and larger, $M_{ij} \to 0$ for $i \neq j$ while M_{ii} converges to a nonzero value. We therefore expect that the contribution to M of the matrices M_{ii} are more important than the one of the matrices M_{ij} for $i \neq j$.

The approximation $M_{appro,1}$ for the information matrix is affine in the squared amplitudes A_i^2 $(i = 1...m)$ and can therefore be used in Algorithm 1. Denote by β_i $(i = 1...m)$ the amplitudes A_i^2 $(i = 1...m)$ obtained by using Algorithm 1 with $M_{appro,1}$. This initial set of amplitudes can be used to determine two refined approximations for M which, unlike $M_{appro,1}$, take explicitly the contribution to M of the matrices M_{ij} for $i \neq j$ and this without loosing the property of affinity in A_i^2 $(i = 1...m)$. Indeed, the information matrix M obtained with a multisine (6) with amplitudes $A_i \approx \beta_i$ $(i = 1...m)$ can be approximated by:

$$M_{appro,2} = \left(\sum_{i=1}^{m} A_i^2 M_{ii}\right) + \underbrace{\left(\sum_{i=1}^{m}\sum_{j=1(j\neq i)}^{m} \beta_i\beta_j M_{ij}\right)}_{\mathcal{M}_\beta} \quad (17)$$

In this approximation, the contribution of the matrices M_{ij} for $i \neq j$ for the multisine with amplitude A_i is taken equal to the contribution $\mathcal{M}_\beta$ of these matrices for the multisine with the (initial) amplitudes β_i.

A more sophisticated approximation for M (see (15)) can be obtained by linearizing the contribution of the matrices M_{ij} around β_i^2:

$$\begin{aligned}\sum_{i=1}^{m}\sum_{j=1(j\neq i)}^{m} A_i A_j M_{ij} &= \sum_{i=1}^{m}\sum_{j=1(j\neq i)}^{m} \sqrt{A_i^2}\sqrt{A_j^2} M_{ij} \\ &\approx \left(\sum_{i=1}^{m} \Delta M_i (A_i^2 - \beta_i^2)\right) + \mathcal{M}_\beta\end{aligned}$$

where ΔM_i is the partial derivative of $\sum\sum \sqrt{A_i^2}\sqrt{A_j^2} M_{ij}$ with respect to A_i^2 and evaluated at β_i^2:

$$\Delta M_i = \sum_{i=1}^{m} \frac{1}{2\beta_i}\left(\sum_{j=1(j\neq i)}^{m} \beta_j M_{ij}\right)$$

for $\beta_i \neq 0$. Consequently, the information matrix M obtained with a multisine (6) with amplitudes $A_i \approx \beta_i$ $(i = 1...m)$ can thus also be approximated by

$$M_{appro,3} = \left(\sum_{i=1}^{m} A_i^2 \left(M_{ii} + \Delta M_i\right) - \Delta M_i \beta_i^2\right) + \mathcal{M}_\beta \quad (18)$$

Like (16), the refined approximations (17) and (18) are affine in the squared amplitudes A_i^2 $(i = 1...m)$ and can therefore be used in Algorithm 1 in order to refine the input sequence with amplitudes β_i $(i = 1...m)$ obtained by using Algorithm 1 with $M_{appro,1}$. We now summarize the overall procedure:

Procedure 1. The multisine (6) for the identification is determined via the following steps:

a) Apply Algorithm 1 with $M_{appro} = M_{appro,1}$. It yields $u_a(t)$.

b) Apply Algorithm 1 with $M_{appro} = M_{appro,2}$ and β_i given by the amplitudes of $u_a(t)$. It yields $u_b(t)$. If the power of $u_b(t)$ is smaller than $u_a(t)$, then pose $u_a(t) = u_b(t)$. Otherwise $u_a(t)$ remains unchanged.

c) Apply Algorithm 1 with $M_{appro} = M_{appro,3}$ and β_i given by the amplitudes of $u_a(t)$. It yields $u_c(t)$. If the power of $u_c(t)$ is smaller than $u_a(t)$, then pose $u_a(t) = u_c(t)$. Otherwise $u_a(t)$ remains unchanged.

d) Go back to step (b) and stop when a power reduction is no longer observed. The input sequence for the identification is then $u_a(t)$ ■

We have observed in numerical examples that steps (b) and (c) do not always provide an improvement with respect to step (a). This generally happens when the scaling factor ξ_{opt} to apply to $u_1(t)$ obtained with $M_{appro} = M_{appro_1}$ is very close to one; showing that the contribution of the matrices M_{ij} for $i \neq j$ is indeed negligeable. Moreover, when steps (b) and (c) are providing an improvement, it is generally sufficient to stop after one iteration.

6. NUMERICAL ILLUSTRATION

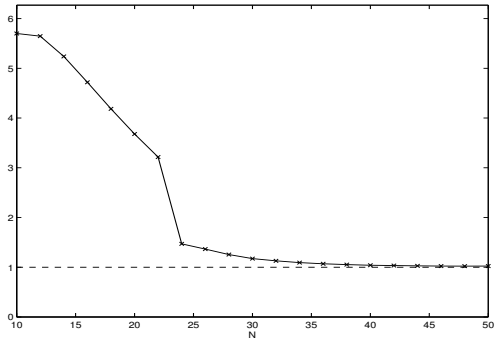

Fig. 1. First example: $\frac{\mathcal{P}_{st}}{\mathcal{P}_{finite}}$ (crosses) for different values of N

In this section, we will illustrate the results presented in this paper. As a first example, we consider the following true system: $G(z,\theta_0) = 0.1766\Lambda_1(z) + 0.1187\Lambda_2(z) + 0.32\Lambda_3(z)$ where $\Lambda_i(z)$ are Laguerre basis functions with a pole at $\xi = 0.8147$. For this $G(z,\theta_0)$, we want to determine the least powerful multisine of the type (6) (with $m = ceil(N/2) - 1$) which guarantees that the model identified with this multisine has a relative error $(G(z,\hat{\theta}_N) - G(z,\theta_0))/G(z,\theta_0)$ which is smaller than 0.1 in the frequency range [0 0.1] (see Jansson and Hjalmarsson (2005) to see how this constraint can be expressed similarly as the constraint in (10)).

We will compare two methods to determine this multisine. The first method consists of using Algorithm 1 with the approximation $M_{appro} = M_{st}$ which neglects the transient effects (see Section 4). The second method presented in Procedure 1 (see Section 5) involve successive applications of Algorithm 1 and takes the transient effects into account. Denote by $\mathcal{P}_{st}$ and $\mathcal{P}_{finite}$ the power (7) of the multisine obtained via these two approaches. Note that, with these two multisines, the accuracy constraint is guaranteed to be met. These multisines and their respective powers are computed for different values of the number N of data. Figure 1 represents the ratio $\frac{\mathcal{P}_{st}}{\mathcal{P}_{finite}}$ for these different N. We observe, as expected, that the approach which explicitly takes into account the transient effects (i.e. the one of Section 5) delivers a multisine which is much closer to the optimal solution of the experiment design problem (10) than the one which neglects these effects. As expected also, the discrepancy is the largest for small values of N and decreases for larger values of N. For these larger values of N, the transient effects are indeed more and more negligeable. In fact, when looking at the impulse response $h(t)$ of $G_0(z)$, we observe that $h(t)$ is significant for $t < 40$ and we see that $N < 40$ is approximately also the region where the transient effects can not be neglected for experiment design. We are grateful to Johan Schoukens for having pointed us to this relation.

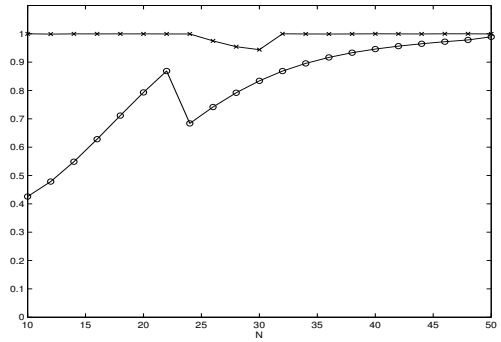

Fig. 2. First example: scaling factor ξ^{opt} required for the approximation of Section 4 (circle) and scaling factor ξ^{opt} corresponding to the last step of Procedure 1 in Section 5 (crosses) for different values of N

That, for small N, the multisine obtained with the methodology of Section 5 is much closer to the optimal multisine than the one obtained via the approach of Section 4 can also be deduced by looking at the scaling factor ξ^{opt} that had to be applied to be sure to meet the accuracy constraint (see Algorithm 1). As can be seen in Figure 2, this scaling factor is almost equal to one for all values of N for the methodology of Section 5 while being relatively far away from one with the method of Section 4.

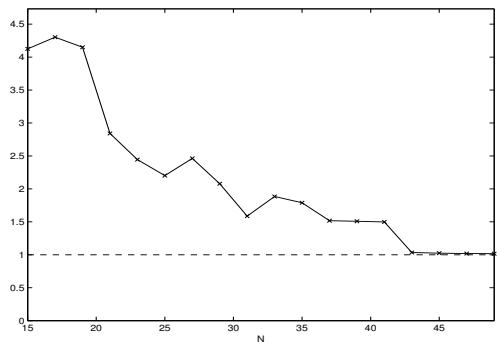

Fig. 3. Same figure as in Figure 1 but for the second example

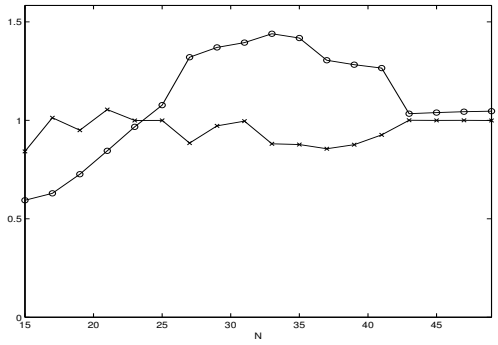

Fig. 4. Same figure as Figure 2 but for the second example

We now consider a second true system where five parameters are to be identified: $G_0(z) = 1.766\Lambda_1(z)+1.187\Lambda_2(z)+0.2\Lambda_3(z)+0.1\Lambda_4(z)+0.05\Lambda_5(z)$. Here, the $\Lambda_i(z)$ are Laguerre basis functions with a pole at $\xi = 0.65$. The results are presented in Figures 3 and 4. We observe the same phenomena as in the first example.

We will now compare the experimental condition considered in this paper i.e. the case where $u(t) = 0$ for $t \leq 0$ and the case where, even though the data are only measured from $t = 1$ to $t = N$, the input signal has been applied from $t = -\infty$. In the latter case, the information matrix M is exactly given by M_{st} and the optimization problem (10) can be exactly solved without passing through Algorithm 1. Denote by $\mathcal{P}_\infty$ the power of the optimal multisine designed in this way. Note that $\mathcal{P}_\infty \neq \mathcal{P}_{st}$ since the multisine corresponding to $\mathcal{P}_\infty$ must not be scaled by ξ^{opt}. We compute this power $\mathcal{P}_\infty$ for different values of N and we compare them with the powers $\mathcal{P}_{finite}$ in the case where $u(t) = 0$ for $t \leq 0$ and which had been computed previously. In Figure 5, the ratio $\frac{\mathcal{P}_\infty}{\mathcal{P}_{finite}}$ corresponding to the first true system is represented for these different values of N. For the second true system, the results are similar and are thus omitted. We observe that the power $\mathcal{P}_\infty$ required to meet the accuracy constraint when the signals are all in steady-state is, for small value of N, much larger than the power $\mathcal{P}_{finite}$ which is required to meet this constraint when the data contain transient effects. This shows that transient effects are useful for the identification. This phenomenon can also be seen by looking at the designed multisine in both cases. In the stationary case, the optimal input sequence is made up of (at least) two cosines for the first example and of (at least) three cosines for the second example as required by the persistence of excitation condition since θ_0 lies in $\mathbf{R}^3$ and $\mathbf{R}^5$, respectively. In the case where $u(t) = 0$ for $t \leq 0$, the designed input sequence contains for small values of N one cosine in the first example and two cosines in the second example and this would in both cases imply a non-informative experiment in the absence of these transient effects.

7. CONCLUDING REMARKS

Until now, we have assumed that the true system was corrupted by a white noise $e(t)$ (see (2)). In the case of colored noise i.e. when $y(t) = G(z, \theta_0)u(t)+H_0(z)e(t)$, we can then proceed as follows for the identification of $G(z, \theta_0)$. Instead of using, as usual, the data set $\{u(t),\ y(t) \mid t = 1...N\}$, the

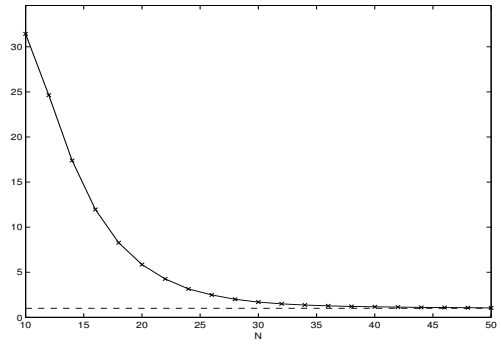

Fig. 5. First example: $\frac{\mathcal{P}_\infty}{\mathcal{P}_{finite}}$ (crosses) for different values of N

identification will be performed using a modified data set $\{u_F(t),\ y_F(t) \mid t = 1...N\}$. The signal $u_F(t)$ $(t = 1...N)$ is the signal that will be optimally designed, but will not be directly applied to the true system. Indeed, this signal will be first filtered by (an estimate of) the noise model $H_0(z)$ yielding $u(t) = H_0(z)u_F(t)$ before being applied to the true system. The corresponding output $y(t)$ $(t = 1...N)$ is measured and is subsequently filtered by (an estimate of) $H_0^{-1}(z)$ to deliver the signal $y_F(t)$ $(t = 1...N)$ which will be used for the identification. The reason why the data $u_F(t)$ and $y_F(t)$ are used for the identification instead of the data $u(t)$ and $y(t)$ is that the relation between $y_F(t)$ and $u_F(t)$ is (approximately) given by $y_F(t) = G(z, \theta_0)u_F(t)+e(t)$ such as in (2) and that, consequently, the design of $u_F(t)$ can be done as presented in the previous sections of this paper (with a slight adaptation of the cost function since it is the power of $u(t) = H_0(z)u_F(t)$ which has to be minimized).

Future development of the results in this paper should consider the optimization of the phase shifts α_i in (6) and alternative relaxation techniques to improve Procedure 1.

REFERENCES

X. Bombois and M. Gilson. Cheapest identification experiment with guaranteed accuracy in the presence of undermodeling. In *14th IFAC Symposium on System Identification, Newcastle*, 2006.

X. Bombois, G. Scorletti, M. Gevers, P. Van den Hof, and R. Hildebrand. Least costly identification experiment for control. *Automatica*, 42(10):1651–1662, 2006.

G.C. Goodwin and R.L. Payne. *Dynamic system identification. Experiment design and data analysis.* Academic Press, 1977.

P. Heuberger, P. Van den Hof, and B. Walhberg (Eds). *Modelling and Identification with Rational Orthogonal Basis Functions.* Springer Verlag, 2005.

H. Jansson. *Experiment design with applications in Identification for Control.* PhD thesis, Royal Institute of Technology, Stockholm, Sweden, 2004.

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

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

R. Pintelon and J. Schoukens. *System Identification - A Frequency Domain Approach.* IEEE Press, New York, 2001.