Mean-squared error experiment design for linear regression models ⋆

Diego Eckhard * **Håkan Hjalmarsson** ** **Cristian R. Rojas** ** **Michel Gevers** ***

* *Department of Electrical Engineering, Universidade Federal do Rio Grande do Sul, Brazil (e-mail: diegoeck@ece.ufrgs.br)*
** *Automatic Control Lab and ACCESS, School of Electrical Engineering, KTH, SE-100 44 Stockholm, Sweden (e-mail: hjalamrs,crro@kth.se)*
*** *CESAME, Université catholique de Louvain, B1348 Louvain la Neuve, Belgium, and Dept ELEC, Vrije Universiteit Brussel, Pleinlaan 2, B1050 Brussels, Belgium (e-mail: Michel.Gevers@uclouvain.be)*

Abstract: This work solves an experiment design problem for a linear regression problem using a reduced order model. The quality of the model is assessed using a mean square error measure that depends linearly on the parameters. The designed input signal ensures a predefined quality of the model while minimizing the input energy.

Keywords: Input and Excitation Design, Maximum Likelihood Methods.

1. INTRODUCTION

Experiment design is a key component of system identification: important improvements in the quality of the estimates or, alternatively, important reductions in experiment time can be obtained by a clever choice of the excitation signals applied for the estimation of parametric models. In system identification, experiment design has its roots in the 1970's with pioneering work from Goodwin and Payne [1977], Zarrop [1979], Mehra [1974] and others. That work dealt with the optimization, with respect to the input signal, of a quality measure of the covariance of the parameter estimates.

In the 1980's the objective of the experiment design changed from a quality measure on the estimated parameters to a quality measure on an application oriented measure of the transfer function estimate. A number of optimal design results were obtained in Yuan and Ljung [1985] for open loop identification and in Gevers and Ljung [1986] for closed loop identification. This was made possible by new approximate frequency domain formulas for the bias and variance of transfer function estimates derived by Ljung in the mid-eighties under the assumption that the model order tends to infinity: see Ljung [1999]. The work of Yuan and Ljung [1985] considered the situation of reduced order models.

Optimal experiment design has seen a major revival in the last decade, with three major advances: (i) a return to the covariance formulas that do not assume that the model order goes to infinity; (ii) a significant expansion of the cost criteria that can be handled; (iii) the adoption of the dual concept where the optimal experiment is one that minimizes the cost of the identification experiment subject to achieving a prescribed quality constraint: see e.g. Jansson [2004], Bombois et al. [2006], Hjalmarsson [2009] and many others.

As far as we know, all of the work on optimal experiment design for finite order models makes the assumption that the system is in the model set, i.e. that there are no bias errors. The optimal criteria are therefore a function of the asymptotic covariance of the estimated quantity. The treatment of the case where the system is not in the model set is much harder to solve.

In this paper we present a first attempt at solving the problem with a reduced order model, by considering the simple case where the true system is a linear regression system with n parameters, and where the model contains only a subset m of these n parameters with $m < n$. The quality criterion is the Mean Square Error between a linear measure of all parameters and the same measure involving only the subset of m estimated parameters. The optimal experiment design problem consists of finding the regression signal with the smallest energy subject to this quality criterion being below some prescribed bound. We show that the optimal solution does not depend on the estimated parameter vector, but that it depends on the unmodeled parameter vector, say θ_2. We then provide an optimal solution that satisfies the required quality constraint for all θ_2 vectors whose norm is below some bound. The optimal solution is such that the smaller the prior uncertainty on θ_2, the smaller the energy that is required of the regression signal.

In Section 2 we set up the Least Squares linear regression problem with a reduced order model and we propose an

⋆ This work was supported in part by CAPES, by CNPq/Brasil, by the European Research Council under the advanced grant LEARN, contract 267381 and by the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office.

Preprint submitted to 16th IFAC Symposium on System Identification.
Received November 14, 2011.

"applications oriented" quality measure. The MSE input design problem is formulated precisely in Section 3 and its solution is presented in Section 4, including for the case where a structure is imposed on the covariance of the regression signal. Several examples are presented in Section 5 which, we believe, provide a lot of intuition into the optimal solution. Section 6 is a brief conclusion. The proofs of all results are presented in the Appendix.

2. PRELIMINARIES

Consider a linear regression system described by

$$y_0(t) = \varphi^T(t)\theta_0 + \nu(t) \tag{1}$$

where $y(t) \in \mathbb{R}$ is the system output, $\varphi(t) \in \mathbb{R}^n$ is the input vector, $\nu(t) \in \mathbb{R}$ is a zero mean white-noise sequence with variance σ^2 and $\theta_0 \in \mathbb{R}^n$ is the vector of parameters.

We would like to estimate a low-order model for the process (1). The model has the following structure:

$$y(t,\theta) = \varphi_1^T(t)\theta \tag{2}$$

where $\theta \in \mathbb{R}^m (m < n)$ is the vector of parameters that we want to estimate. Notice that the model comprises only the m first elements of θ_0 and $\varphi(t)$, so the model is a low-order approximation of the system. Let us assume that the input and the parameter vector have the structure

$$\varphi^T(t) = \left[\, \varphi_1^T(t)\ \varphi_2^T(t) \,\right]^T, \ \ \theta_0^T = \left[\, \theta_1^T\ \theta_2^T \,\right]^T.$$

Assume that N samples are collected from the system. This setup can be described in a compact form as

$$Y = \Phi\theta_0 + V, \tag{3}$$

where

$$Y = [\, y_0(1)\ \cdots\ y_0(N) \,]^T, \ \ V = [\, \nu(1)\ \cdots\ \nu(N) \,]^T,$$
$$\Phi_1^T = [\, \varphi_1(1)\ \cdots\ \varphi_1(N) \,], \ \ \Phi_2^T = [\, \varphi_2(1)\ \cdots\ \varphi_2(N) \,],$$
$$R = \begin{bmatrix} R_{11} & R_{12} \\ * & R_{22} \end{bmatrix} = \Phi^T\Phi = \begin{bmatrix} \Phi_1^T\Phi_1 & \Phi_1^T\Phi_2 \\ * & \Phi_2^T\Phi_2 \end{bmatrix}, \ \ \Phi = [\, \Phi_1\ \Phi_2 \,].$$

The least-squares technique can be used to obtain an estimate of θ based on the N samples collected from the system. The estimate is given by

$$\hat{\theta} = \left(\Phi_1^T\Phi_1\right)^+ \Phi_1^T Y \tag{4}$$

where A^+ is the Moore-Penrose pseudo-inverse of A [Horn and Johnson, 1999]. The pseudo-inverse is used here because in some applications the matrix $\Phi_1^T\Phi_1$ is not full rank. The estimate $\hat{\theta}$ is optimal in the sense that it minimizes the quadratic criterion $\sum_{t=1}^{N}(y_0(t) - y(t,\theta))^2$. However, notice that $\hat{\theta}$ is a biased estimate of θ_1.

One of the simplest quantities used to evaluate the quality of $\hat{\theta}$ is the linear composition of the parameters:

$$J(\hat{\theta}) = \Lambda_1^T\hat{\theta}$$

where $\Lambda_1 \in \mathbb{R}^m$ is chosen according to the purpose of the model. For example, if the model is an FIR and Λ_1 is a vector of ones, then $J(\hat{\theta})$ is the DC-gain of the model. We can define a similar quantity for the process

$$J(\theta_0) = \Lambda_1^T\theta_1 + \Lambda_2^T\theta_2.$$

where $\Lambda_2 \in \mathbb{R}^{n-m}$; this quantity takes into account both θ_1 and θ_2.

In this work we propose the Mean-Square Error as a measure to evaluate the quality of the estimated model. The measure is defined as

$$MSE(\hat{\theta}) = E\left[\left(J(\hat{\theta}) - J(\theta_0)\right)^2\right]. \tag{5}$$

Notice that the criterion considers both bias and variance errors. When (5) is evaluated using (4) we get

$$MSE(\hat{\theta}) = \left\|\Lambda_1^T R_{11}^+ \left(R_{11}\theta_1 + R_{12}\theta_2\right) - \Lambda_1^T\theta_1 + \Lambda_2^T\theta_2\right\|^2 + \sigma^2\Lambda_1^T R_{11}^+\Lambda_1. \tag{6}$$

We can see that the measure $MSE(\hat{\theta})$ depends strongly on the input signal. A natural way to reduce the value $MSE(\hat{\theta})$ and, therefore, to improve the quality of the model, is to choose properly the input vector $\varphi(t)$. The next section covers the design of the input signal.

3. THE MSE INPUT DESIGN PROBLEM

In this section, we propose a method to design $\varphi(t)$ which ensures that the quality measure is smaller than a specified value γ.

It is worth noticing that there is a trade-off between the quality measure and "energy" of the signal $\varphi(t)$. If a signal with more "energy" is used, it is possible to achieve a lower value of γ. In this work, we are interested in finding the signal with smallest "energy" which ensures that $MSE(\hat{\theta}) \leq \gamma$. The energy of the signal will be measured by the trace of R, i.e. trace of $\Phi^T\Phi$.

The design of the input signal can be described by the following optimization problem

$$\min_R trR \tag{7}$$
$$\text{s. t.} \quad MSE(\hat{\theta}) \leq \gamma. \tag{8}$$

In order to ensure that we can construct a signal $\varphi(t)$ from R we need to add one more constraint to the problem:

$$R \geq 0. \tag{9}$$

Note that in (6) the quality measure $MSE(\hat{\theta})$ depends on θ_1 and θ_2, which are assumed to be unknown to the designer. In principle, one can only compute the optimal input signal $\varphi(t)$ if the values θ_1 and θ_2 are known. On the other hand, we will show in the sequel that the solution of the problem does not involve θ_1. However, the optimal solution depends on θ_2. In this work, we will relax the above optimization problem, considering that an upper-bound to the norm of θ_2 is known and that we want to ensure that

$$MSE(\hat{\theta}) \leq \gamma \ \ \forall\theta_2 : \|\theta_2\| \leq \alpha.$$

In other words, we will not solve the problem just for a specific θ_2, but for any θ_2 such that $\|\theta_2\| \leq \alpha$. The advantage of this robust/worst case formulation is that the user only needs to specify the upper-bound α.

We thus have the problem:

$$\begin{aligned} R_{opt} = {} & \arg\min_R trR \\ \text{s. t.} \quad & MSE(\hat{\theta}) \leq \gamma \ \ \forall\theta_2 : \|\theta_2\| \leq \alpha \\ & R \geq 0. \end{aligned} \tag{10}$$

Preprint submitted to 16th IFAC Symposium on System Identification.
Received November 14, 2011.

In the next section, we will show that this optimization problem has an explicit solution, which can be easily computed and that has several interesting properties.

In some cases, it is desirable that the matrix R has a specified structure. For example, consider that system (1) has an FIR structure such that

$$\varphi^T(t) = \left[\, \varphi_1^T(t)\ \varphi_1^T(t-1)\ \varphi_1^T(t-2)\ \varphi_1^T(t-3) \,\right]^T.$$

Under this consideration, the matrix R is Toeplitz. In order to ensure that the solution of the problem has the desired structure, we should include one more constraint in the problem. In this case, the problem may not have an explicit solution, although a numeric solution can be found. In this work, we present also a solution to the problem when the matrix R is structured.

4. SOLUTION TO THE MSE INPUT DESIGN PROBLEM

4.1 Unstructured R

In this section we will compute the explicit solution to the optimization problem proposed before.

Lemma 1. Problem (10) is equivalent to following problem with only one variable $\delta \in \mathbb{R}^{n-m}$ and with one constraint:

$$\delta_{opt} = \arg\min_{\delta} F(\delta) \tag{11}$$

$$\text{s.t.} \quad F(\delta) = \frac{\sigma^2 \left(\|\Lambda_1\|^2 + \|\delta + \Lambda_2\|^2 \right)}{\gamma - \|\delta\|^2 \alpha^2} \tag{12}$$

$$\gamma > \|\delta\|^2 \alpha^2, \tag{13}$$

where

$$R_{opt} = \frac{\sigma^2 \Lambda_{opt} \Lambda_{opt}^T}{\gamma - \|\delta_{opt}\|^2 \alpha^2} \tag{14}$$

$$\Lambda_{opt}^T = \left[\Lambda_1^T \quad (\delta_{opt} + \Lambda_2) \right]^T. \tag{15}$$

Proof: See the appendix. ■

Note that if the constraint is active, then $F(\delta)$ tends to infinity, which is not the solution of the minimization problem. Hence, the solution lies in the interior of the admissible set. The solution of the problem can thus be found by computing the values of δ such that the gradient of $F(\delta)$ is equal to zero. We can compute all the possible solutions and then check, one by one, if they respect the constraint of the problem. It will be shown that this problem has two solutions, but that only one of them respects the constraint. The result is presented in the next theorem.

Theorem 2. The solution of the problem (11) is

$$\delta = k\Lambda_2$$

where

$$k = \frac{-\left(\|\Lambda_1\|^2 + \|\Lambda_2\|^2 + \frac{\gamma}{\alpha^2} \right) + \sqrt{\left(\|\Lambda_1\|^2 + \|\Lambda_2\|^2 + \frac{\gamma}{\alpha^2} \right)^2 - 4\,\|\Lambda_2\|^2 \frac{\gamma}{\alpha^2}}}{2\,\|\Lambda_2\|^2}$$

Proof: See the appendix. ■

This result has several interesting properties which are presented on the next remarks.

Remark 1. The solution of the problem does not depend on the value of θ_1. The upper-bound α is the only system property that affects the result.

Remark 2. The optimal input signal is proportional to the noise ($R_{opt} \propto \sigma^2$). This is the usual case in input design. However, note that we are identifying a low-order model, and we have both errors from bias and variance. Even so, the optimal solution is proportional to the level of noise.

Remark 3. There is a trade-off between the energy of the signal ($tr(R)$) and the quality of the estimate γ. The asymptotic result is

$$\lim_{\gamma\to 0} tr(R) = \infty \qquad \lim_{\gamma\to\infty} tr(R) = 0.$$

The result shows that in order to have a "perfect model" described by $\gamma \to 0$, it is necessary to have an input signal with unbounded energy. On the other hand, if the quality measure can be very large ($\gamma \to \infty$) then the optimum input signal is zero. There is a relation between the quality of the model and the necessary amount of energy used in the input signal; we get a better model if we use a signal with more energy.

Remark 4. The asymptotic relation between R and the parameter α is given by:

$$\begin{aligned}
&\lim_{\alpha\to 0} R_{11} = \frac{\sigma^2 \Lambda_1 \Lambda_1^T}{\gamma} && \lim_{\alpha\to\infty} R_{11} = \frac{\sigma^2 \Lambda_1 \Lambda_1^T}{\gamma} \\
&\lim_{\alpha\to 0} R_{12} = 0 && \lim_{\alpha\to\infty} R_{12} = \frac{\sigma^2 \Lambda_1 \Lambda_2^T}{\gamma} \\
&\lim_{\alpha\to 0} R_{22} = 0 && \lim_{\alpha\to\infty} R_{22} = \frac{\sigma^2 \Lambda_2 \Lambda_2^T}{\gamma}.
\end{aligned}$$

The result shows that if we know that θ_2 is very small ($\alpha \to 0$) then the energy of the input signal is concentrated on R_{11}, which means that it is not necessary to excite the part of the process related to θ_2. On the other hand, even if nothing is known about θ_2 ($\alpha \to \infty$) the energy of the input signal is finite.

4.2 Structured R

If we want to solve the optimization problem considering that R is structured, we need to include one more constraint to the optimization problem. The next theorem presents an equivalent problem to (10) with the extra constraint, described by Bilinear Matrix Inequalities.

Theorem 3. If there are a matrix Z, a structured matrix R and a scalar x which satisfy the conditions

$$\begin{bmatrix} \gamma - x & \sigma\Lambda_1^T \\ * & R_{11} \end{bmatrix} \geq 0 \tag{16}$$

$$\begin{bmatrix} x & \alpha\left(\Lambda_1^T Z - \Lambda_2^T\right) \\ * & I \end{bmatrix} \geq 0 \tag{17}$$

$$R_{11} Z = R_{12} \tag{18}$$

$$\begin{bmatrix} R_{11} & R_{12} \\ * & R_{22} \end{bmatrix} > 0 \tag{19}$$

$$tr\left\{ \begin{bmatrix} R_{11} & R_{12} \\ * & R_{22} \end{bmatrix} \right\} < t \tag{20}$$

then

$$\begin{gathered} MSE(\theta) \leq \gamma, \ \ \forall \theta_2 : \|\theta_2\| \leq \alpha. \\ R > 0, \\ tr(R) < t. \end{gathered}$$

Preprint submitted to 16th IFAC Symposium on System Identification.
Received November 14, 2011.

Proof: From the conditions of the theorem, R_{11} is positive definite, and the problem can be simplified to

$$\left\|\left(\Lambda_1^T (R_{11})^{-1} R_{12} - \Lambda_2^T\right)\theta_2\right\|^2 + \sigma^2 \Lambda_1^T (R_{11})^{-1} \Lambda_1 \leq \gamma \quad \forall \theta_2 : \|\theta_2\| \leq \alpha$$

$$R > 0$$

$$tr(R) < t.$$

The first condition is satisfied $\forall \theta_2 : \|\theta_2\| \leq \alpha$ if and only if

$$\left\|\left(\Lambda_1^T (R_{11})^{-1} R_{12} - \Lambda_2^T\right)\right\|^2 \alpha^2 + \sigma^2 \Lambda_1^T (R_{11})^{-1} \Lambda_1 \leq \gamma.$$

Using the Schur complement and introducing the slack variable x, the condition becomes

$$\begin{bmatrix} \gamma - x & \sigma\Lambda_1^T \\ * & R_{11} \end{bmatrix} \geq 0 \tag{21}$$

$$\begin{bmatrix} x & \alpha\left(\Lambda_1^T R_{11}^{-1} R_{12} - \Lambda_2^T\right) \\ * & I \end{bmatrix} \geq 0. \tag{22}$$

The change of variables

$$Z = R_{11}^{-1} R_{12}$$

completes the proof. ■

Note that the present problem is not convex because the equality (18) presents a multiplication between two variables. A local solution of this problem can be found using iterative LMIs:

- Fix Z and solve for the other variables;
- Fix R_{11} and solve for other variables.

We propose two optimization criteria to be used with this theorem.

(1) Minimize γ for a fixed t;
(2) Minimize t for a fixed γ.

In the first problem we have a fixed "energy" and we want to maximize the quality of the estimate.

In the second problem we want to compute the minimal "energy" of the input signal such that we ensure that the model has a minimum level of quality.

5. EXAMPLES

5.1 Example 1

Consider the following problem

$$\Lambda_1 = 1, \quad \Lambda_2 = 1, \quad \gamma = 1, \quad \sigma = 1.$$

We will compute the solution to the optimization problem considering several values of α.

Figure 1 shows the values of R_{11}, R_{12} and R_{22} as a function of the parameter α. The figure also shows the trace of R, indicated as t.

As expected, $\lim_{\alpha\to 0} R_{11} = \lim_{\alpha\to\infty} R_{11} = 1$ and the trace of R increases with α. If $\alpha = 1$ then the solution is

$$R_{11} = 1.1708, \quad R_{12} = 0.7239,$$
$$R_{22} = 0.4478, \quad t = tr(R) = 1.6189.$$

The results presented in the Figure 1 confirm Remark 4. For small values of α both R_{12} and R_{22} tend to zero,

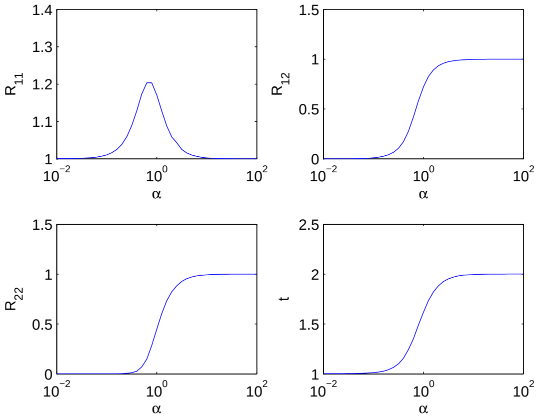

Fig. 1. Solution for varying α.

and for large values of α the matrix R is bounded. It is interesting to note that the results almost do not change for values of α smaller than 10^{-1}. The same happens for values of α larger than 10^1. The size of this band of values for which the results change is under investigation.

5.2 Example 2

Consider that the following system

$$y(t) = 2u_1(t) - 1u_2(t) + v(t)$$

will be identified by the model

$$y(t) = \hat{\theta} u_1(t)$$

and the experiment design constraints are the same as in the previous example with $\alpha = 1$. As presented before, the optimal solution is

$$R_{11} = 1.1708, \quad R_{12} = 0.7239, \quad R_{22} = 0.4478.$$

The input signal can be realized in many different ways. We realized the signal as the periodic signals with 10000 samples and period equal to 2:

$$u_1(t) = [0.00057\ 0.01529\ 0.00057\ 0.01529...]$$
$$u_2(t) = [0.00057\ 0.00944\ 0.00057\ 0.00944...].$$

To check the quality of the experiment design, a Monte-Carlo experiment with 10000 runs was performed.

The mean value of the estimated parameter was $\frac{1}{10000}\sum_{i=1}^{10000}\hat{\theta} = 1.3674$ and the mean square value of the quality measure was $\frac{1}{10000}\sum_{i=1}^{10000}\left(J(\theta_0) - J(\hat{\theta})\right)^2 = 0.9883$. Note that as expected the value is approximately γ.

5.3 Example 3

Consider the following problem

$$\Lambda_1 = 1, \quad \Lambda_2 = 1, \quad \gamma = 1, \quad \sigma = 1, \quad \alpha = 1$$

where the system is

$$y(t) = 2u(t) - 1u(t-1) + v(t)$$

and the model is

$$y(t) = \hat{\theta} u(t)$$

Note that because the system has an FIR structure, we need to ensure that R is Toeplitz, and to include one more constraint in the optimization. The solution of the problem is

$$R = \begin{bmatrix} 1.000 & 1.000 \\ 1.000 & 1.000 \end{bmatrix}.$$

It is interesting that the solution of this problem is the same for any value of α. This is, however, not true in general (see Example 4).

Note that the trace of R is 2.000, which is larger than the value obtained for the unstructured problem (see Example 1). The extra constraint, which was included to ensure the structure of the matrix R, made the problem more restricted, resulting in a larger "energy" needed for the input signal.

The input signal was realized as a constant $u(t) = 0.01$ for $t = 1, ..., 10000$.

The solution of the problem was used in 10000 Monte-Carlo runs. The mean value of the estimated model parameter is $\frac{1}{10000}\sum_{i=1}^{10000}\hat{\theta} = 1.0246$ and the mean value of the quality criterion is given by $\frac{1}{10000}\sum_{i=1}^{10000}\left(J(\theta_0) - J(\hat{\theta})\right)^2 = 0.9972$. Note that as expected the value is approximately γ. The discrepancy is due to the variance of the Monte-Carlo runs.

5.4 Example 4

Consider the following problem

$$\Lambda_1 = 1, \ \ \Lambda_2 = 2, \ \ \gamma = 1, \ \ \sigma = 1.$$

We compute the solution of the optimization problem, considering that R is Toeplitz. The solution is plotted in figure 2 as a function of α.

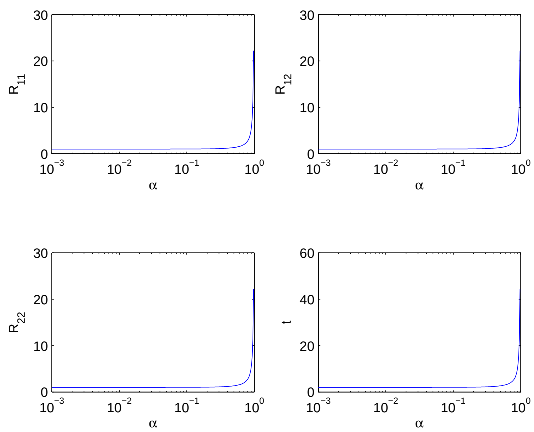

Fig. 2. Solution varying α.

The values of R_{11}, R_{12} and R_{22} increase with α, and all of them tend to infinity when α tends to 1, as presented in Figure 2. The problem is unfeasible for $\alpha \geq 1$.

When R is unstructured, the optimization problem has a finite solution for every value of α. The present example has a structured R, and the extra constraint - used to ensure the structure of the matrix - makes the problem unfeasible for large values of α.

6. CONCLUSION

This article presented a solution to an experiment design problem for a linear regression model of reduced order. The energy of the input signal was minimized and the designed input ensured a predefined quality of the model which was assessed using a mean square error measure.

7. PROOF OF LEMMA 1

The problem can be described as

$$\min_{R_{11},R_{12},R_{22}} tr(R_{11}) + tr(R_{22})$$

$$\begin{aligned}
\text{s. t. } & \|\beta\|^2 + \sigma^2\Lambda_1^T R_{11}^+ \Lambda_1 \leq \gamma \ \ \forall\theta_2 : \|\theta_2\| \leq \alpha && (23)\\
& R_{11} \geq 0 && (24)\\
& R_{22} \geq R_{12}^T R_{11}^+ R_{12} && (25)\\
& (I - R_{11}R_{11}^+)R_{12} = 0 && (26)\\
& \beta = \Lambda_1^T R_{11}^+ (R_{11}\theta_1 + R_{12}\theta_2) - \Lambda_1^T\theta_1 - \Lambda_2^T\theta_2. && (27)
\end{aligned}$$

Since we are searching for the matrix R_{22} with smallest trace, we know that the constraint (25) is active, so that it becomes an equation, $R_{22} = R_{12}^T R_{11}^+ R_{12}$.

In order to ensure that the constraint (23) is satisfied $\forall\theta_2 : \|\theta_2\| \leq \alpha$ we will relax this constraint to

$$h(\beta) + \sigma^2\Lambda_1^T R_{11}^+ \Lambda_1 \leq \gamma \tag{28}$$

where

$$h(\beta) = \max_{\theta_2} \|\beta\|^2 \quad s.t. \quad \|\theta_2\| \leq \alpha.$$

However, because R_{11} has the smallest possible trace (28) becomes also an equation. The problem can be simplified to

$$\min_{R_{11},R_{12}} tr(R_{11}) + tr(R_{12}^T R_{11}^+ R_{12})$$

$$\begin{aligned}
\text{s. t. } & h(\beta) + \sigma^2\Lambda_1^T R_{11}^+ \Lambda_1 = \gamma && (29)\\
& R_{11} \geq 0 && (30)\\
& R_{12} = R_{11}R_{11}^+ R_{12}. && (31)
\end{aligned}$$

If we replace (31) in the criteria we get

$$tr(R_{11})+tr(R_{12}^T R_{11}^+ R_{12}) = tr(R_{11})+tr(R_{12}^T R_{11}^+ R_{11} R_{11}^+ R_{12})$$

but, from the definition of pseudo-inverse $R_{11}^+ R_{11} R_{11}^+ = R_{11}^+$, so the criteria can be simplified again to

$$tr(R_{11}) + tr(R_{12}^T R_{11}^+ R_{12}).$$

Replacing R_{12} in the definition of β we get

$$\begin{aligned}
\beta &= \Lambda_1^T R_{11}^+ \left(R_{11}\theta_1 + R_{11}R_{11}^+ R_{12}\theta_2\right) - \Lambda_1^T\theta_1 - \Lambda_2^T\theta_2\\
&= \Lambda_1^T R_{11}^+ (R_{11}\theta_1 + R_{12}\theta_2) - \Lambda_1^T\theta_1 - \Lambda_2^T\theta_2
\end{aligned}$$

We see that the constraint (31) does not change the problem and it can be dropped without any loss.

Equation (29) is respected for every X if

$$R_{11}^+ = \left(\frac{\sigma^2\Lambda_1\Lambda_1^T}{\gamma - h(\beta)}\right)^+ + (I - \Lambda_1^+\Lambda_1^T)X. \tag{32}$$

If we plug (32) into (27) we get

$$\beta = (\Lambda_1^T R_{11}^+ R_{12} - \Lambda_2^T)\theta_2 \tag{33}$$

and then

Preprint submitted to 16th IFAC Symposium on System Identification.
Received November 14, 2011.

$$h(\beta) = \|\delta\|^2\alpha^2 \quad (34)$$
$$\delta^T = \Lambda_1^T R_{11}^+ R_{12} - \Lambda_2^T. \quad (35)$$

From (32) and (24)

$$R_{11} = \frac{\sigma^2\Lambda_1\Lambda_1^T}{\gamma - \|\delta\|^2\alpha^2} + MM^T \quad if \quad \gamma > \|\delta\|^2\alpha^2. \quad (36)$$

Remember that we search for the minimum trace matrix R, hence the solution of the problem has $M = 0$.

From (35) and (36) it is possible to simplify the problem to an optimization problem with just one variable (δ) and with one constraint:

$$\min_{\delta} \frac{\sigma^2\left(\|\Lambda_1\|^2 + \|\delta + \Lambda_2\|^2\right)}{\gamma - \|\delta\|^2\alpha^2}$$
$$\text{s.t. } \gamma > \|\delta\|^2\alpha^2.$$

8. PROOF OF THEOREM 2

Let us first compute the gradient of F

$$\nabla F = \frac{2\sigma^2(\delta + \Lambda_2)}{\gamma - \|\delta\|^2\alpha^2} + \frac{2\sigma^2\delta\alpha^2\left(\|\Lambda_1\|^2 + \|\delta + \Lambda_2\|^2\right)}{\left(\gamma - \|\delta\|^2\alpha^2\right)^2} \quad (37)$$

and let us now evaluate for which values of δ the gradient is zero:

$$\nabla F = \left(\gamma - \|\delta\|^2\alpha^2\right)(\delta + \Lambda_2) + \delta\alpha^2\left(\|\Lambda_1\|^2 + \|\delta + \Lambda_2\|^2\right) = 0.$$

We can reorganize the terms of this equation to

$$\delta\left(\alpha^2\|\Lambda_1\|^2 + \alpha^2\|\delta + \Lambda_2\|^2 + \gamma - \|\delta\|^2\alpha^2\right) = -\left(\gamma - \|\delta\|^2\alpha^2\right)\Lambda_2$$

and describe in a convenient form

$$\delta = -\frac{\left(\gamma - \|\delta\|^2\alpha^2\right)}{\left(\alpha^2\|\Lambda_1\|^2 + \alpha^2\|\delta + \Lambda_2\|^2 + \gamma - \|\delta\|^2\alpha^2\right)}\Lambda_2.$$

So, we can prove the first part of the theorem, which says that the vector δ is a scaling version of the vector Λ_2

$$\delta = k\Lambda_2,$$

where $k \in \mathbb{R}$.

We can use this fact, to simplify the equation $\nabla F = 0$:

$$k^2\|\Lambda_2\|^2 + k\left(\|\Lambda_1\|^2 + \|\Lambda_2\|^2 + \frac{\gamma}{\alpha^2}\right) + \frac{\gamma}{\alpha^2} = 0.$$

This is a second order polynomial on the parameter k. There are two solutions to this equation:

$$k_1 = \frac{-\left(\|\Lambda_1\|^2 + \|\Lambda_2\|^2 + \frac{\gamma}{\alpha^2}\right) + \sqrt{\left(\|\Lambda_1\|^2 + \|\Lambda_2\|^2 + \frac{\gamma}{\alpha^2}\right)^2 - 4\|\Lambda_2\|^2\frac{\gamma}{\alpha^2}}}{2\|\Lambda_2\|^2}$$
$$k_2 = \frac{-\left(\|\Lambda_1\|^2 + \|\Lambda_2\|^2 + \frac{\gamma}{\alpha^2}\right) - \sqrt{\left(\|\Lambda_1\|^2 + \|\Lambda_2\|^2 + \frac{\gamma}{\alpha^2}\right)^2 - 4\|\Lambda_2\|^2\frac{\gamma}{\alpha^2}}}{2\|\Lambda_2\|^2}.$$

Remember that we need to verify if the constraints of the problem are not violated. We need to ensure that

$$\|\delta\|^2 = k^2\|\Lambda_2\|^2 < \frac{\gamma}{\alpha^2}.$$

However, this expression involves k^2. To simplify the condition, we will use the following equation,

$$k^2\|\Lambda_2\|^2 = -k\left(\|\Lambda_1\|^2 + \|\Lambda_2\|^2 + \frac{\gamma}{\alpha^2}\right) - \frac{\gamma}{\alpha^2}$$

and then, the constraint is

$$-k\left(\|\Lambda_1\|^2 + \|\Lambda_2\|^2 + \frac{\gamma}{\alpha^2}\right) - \frac{\gamma}{\alpha^2} < \frac{\gamma}{\alpha^2}$$

which has the following simplified form

$$-k\left(\|\Lambda_1\|^2 + \|\Lambda_2\|^2 + \frac{\gamma}{\alpha^2}\right) < 2\frac{\gamma}{\alpha^2}. \quad (38)$$

We need to verify if the solutions k_1 and k_2 respect the constraint (38).

The constraint evaluated with k_1 is equivalent to

$$\left(\|\Lambda_1\|^2 + \|\Lambda_2\|^2 + \frac{\gamma}{\alpha^2}\right)^2 - 4\|\Lambda_2\|^2\frac{\gamma}{\alpha^2} <$$
$$\left(\|\Lambda_1\|^2 + \|\Lambda_2\|^2 + \frac{\gamma}{\alpha^2}\right)\sqrt{\left(\|\Lambda_1\|^2 + \|\Lambda_2\|^2 + \frac{\gamma}{\alpha^2}\right)^2 - 4\|\Lambda_2\|^2\frac{\gamma}{\alpha^2}} \quad (39)$$

which is true. Hence, the first solution does not violate the constraints of the problem.

The second solution gives the constraint

$$\left(\|\Lambda_1\|^2 + \|\Lambda_2\|^2 + \frac{\gamma}{\alpha^2}\right)^2 - 4\|\Lambda_2\|^2\frac{\gamma}{\alpha^2} <$$
$$-\left(\|\Lambda_1\|^2 + \|\Lambda_2\|^2 + \frac{\gamma}{\alpha^2}\right)\sqrt{\left(\|\Lambda_1\|^2 + \|\Lambda_2\|^2 + \frac{\gamma}{\alpha^2}\right)^2 - 4\|\Lambda_2\|^2\frac{\gamma}{\alpha^2}} \quad (40)$$

which is is false. Hence, the second solution violates the constraints of the problem.

Now we know that the solution of the problem is given by $\delta = k_1\Lambda_2$, what concludes the proof.

REFERENCES

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

M. Gevers and L. Ljung. Optimal experiment designs with respect to the intended model application. *Automatica*, 22:543–554, 1986.

G. C. Goodwin and R. L. Payne. *Dynamic System Identification: Experiment Design and Data Analysis.* Academic Press, New York, 1977.

H. Hjalmarsson. System identification of complex and structured systems. *European Journal of Control*, 15 (3-4):275–310, 2009.

R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, 1999.

H. Jansson. *Experiment design with applications in identification for control.* PhD thesis, Royal Institute of Technology (KTH), December 2004. TRITA-S3-REG-0404.

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

R. K. Mehra. Optimal input signals for parameter estimation in dynamic systems - survey and new results. *IEEE Trans. on Automatic Control*, AC-19(6):753–768, December 1974.

Z.D. Yuan and L. Ljung. Unprejudiced optimal open loop inut design for identification of transfer functions. *Automatica*, 21(6):697–708, 1985.

M. Zarrop. *Optimal experiment design for dynamic system identification.* Lecture Notes in Control and Information Sciences, Vol. 21, Springer Verlag, Berlin, New York, 1979.

Preprint submitted to 16th IFAC Symposium on System Identification.
Received November 14, 2011.