# TWO TIME SCALED PARAMETER IDENTIFICATION BY COORDINATION OF LOCAL IDENTIFIERS

F.Chaplais K.Alaoui

*Centre Automatique et Systèmes*
*École Nationale Supérieure des Mines de Paris*
*35 rue Saint Honoré*
*77305 Fontainebleau Cedex FRANCE*
*Fax: (33) 1 64 69 47 01, Vocal: (33) 1 64 69 48 71*
*e-mail: chaplais@cas.ensmp.fr*

**Abstract**

Contradictory requirements on the precision in $L^\infty$ and $L^2$ norms prevent the classical least square procedure from identifying two time scaled systems accurately. A new identification procedure, explicitly using the time scale structure of the system, is proposed here, and is proved to solve the problem of accurate parameter estimation.

*Keywords:* Parameter estimation, Two time scaled systems, Least-squares method, Prefiltering

## 1 Introduction and summary

It is well known that systems that have a wide dispersion of poles and zeroes are hard to identify. Among those "wide band" systems, we shall study two time scaled systems (see Luse and Khalil 1985). The poles and zeroes of such systems can be grouped into two distinct sets, that we shall call the slower and faster parts of the system.

The reason why the classical least squares method fails to give an accurate estimation of the parameters is that it has to deal with data that is very scattered in the frequency domain.

⋆ This work was financed by the department of Electrical Applications of Électricité de France, EDF-ADE, Les Renardières, France.

To get around this, it is natural to identify separately the slow and fast parts of the system. The only trouble is that two independent identifications may lead to a couple of models that cannot be traced back to a unique two time scaled system. This is avoided by requiring that the static gain of the fast submodel must be equal to the high frequency gain of the slow one; it is a consequence of an approximation theorem on two-time scaled systems.

The previous ideas suggest a new identification scheme for two time scaled systems, which can be outlined as follows. First, the signals are prefiltered twice in parallel, to provide the slow and fast components of the data. After a suitable time rescaling, the prefiltered signals are used to compute two partial excitation matrices. The two time scaled identifier uses then these matrices to minimize the sum of the slow and fast model prediction errors, under the previous equality constraint on the gains.

Because it uses some transfer approximations, this method is not a model matching method for two time scaled systems. Hence, it has a worse performance than the classical least squares when there is not an important separation of time scales, and/or when the noise/signal ratio is small.

The situation is reversed, however, when the time scales are quite distinct and the data is corrupted. The reason is that each subsystem identifier works, at its own level, with homogeneous data. In that case, this seems to be more important than the issue of model matching. From a theoretical point of view, we show that the two time scaled identification method provides asymptotically exact parameters when the time scales ratio $\epsilon$ tends to 0. We also show that, with suitable signals and filters, it is possible to maintain a bounded modulus of continuity between the two time scaled excitation matrices and the associate identified parameters, while $\varepsilon$ tends to 0. This proves that the two time scaled method is more robust than the classical least squares method. Such a result is achieved at the expense of the method's optimal accuracy.

A few numerical results illustrate then the compared performances of the classical least squares method and of its two time scaled equivalent. They have been obtained by using a MATLAB® program; its source code can be obtained by e-mail on simple request to the author.

## 2 Two time scaled systems and their identification by the least squares method

In this section we define two time scaled systems, and explain briefly why the classical least squares method fails to give an robust estimation of the parameters for this class of systems.

**Technical settings** The identification problems will expressed in continuous time. The reason is that it greatly simplifies the formulation of time rescaling. The time scales will be the usual one and a slower one. While the tradition in the theory of singular perturbations is to consider a fast v.s. normal time scale decomposition, we made this choice because it is closer to the situation in discrete time [1] : in that framework, the basic time unit is that of the faster subsystem. Because we will consider plants with very small poles and zeroes, the identification horizon will be taken equal to the infinity.

### *2.1 Definition of two time scaled systems*

**Definition 1** *Following (Luse and Khalil 1985), we call* continuous two time scaled systems*, systems with a rational transfer of the form*

$$T_\epsilon(s) = T_s(\frac{s}{\varepsilon})T_n(s) \tag{1}$$

*where $\varepsilon$ is a (small) scaling parameter.*

$T_n$ and $T_s$, of course, are assumed to be independent of $\varepsilon$.

**Remarks**

- the poles and zeroes of $T_n(s)$ are of order 1, while those of $T_s(s/\varepsilon)$ are of order $\varepsilon$. Hence, for a small $\varepsilon$, $T_n$ describes (up to a constant) the behavior of $T_\varepsilon$ in the medium frequency range, while $T_s$ describes it in the low frequency range [2] .
- $T_\varepsilon(s)$ is actually a rational fraction with two variables $s$ and $\varepsilon$. In general, $\varepsilon$ itself is not identifiable, because only one plant $T_\varepsilon$ is available to the user. Hence the numerical value of $\varepsilon$ has to be drawn from some a priori knowledge, or perhaps from some recursive tuning of the identifier.

### *2.2 Asymptotic degeneracy of the global least squares method*

We briefly show here why, in the case of the identification of two time scaled systems, the least squares methods fails to provide accurate results.

[1] By the way, the two points of view exist in the theory of averaging.

[2] If we refer to the notations of Luse and Khalil, we have $H(s, \varepsilon) = T_\varepsilon(\varepsilon s)$. This is because Luse and Khalil use a fast/medium time scale decomposition, which is classical in the theory of singular perturbations.

This is essentially due to the scattering of the input in the frequency domain.

**About the shape of the input** In order to excite both parts of the system, we shall assume that the input $u$ is a two time scaled signal $u(\varepsilon, t)$, of the form:

$$u(\varepsilon, t) = u_1(t) + \varepsilon^\alpha u_2(\varepsilon t) \tag{2}$$

The extra parameter $\varepsilon^\alpha$ is here to allow different amplitude ratios between the two components, a matter that we shall investigate in the next section.

**Remark:** since we use a continuous time, differential formulation, we shall have to assume, technically, that $u_1$ and $u_2$ are regular enough, e.g., they belong to the Sobolev space $H^{n-1}(\mathbb{R})$ [3].

### *2.2.1 The amplitude v.s. energy dilemma*

If we take $\alpha > 0$, it is clear that the slow part of the signal will vanish at $\varepsilon = 0$ in $L^\infty$ norm; this *a fortiori* true for its derivatives. This means that, unless we compute with arbitrary precision, all that concerns the slow signals in the excitation matrix will be mistaken for 0. If we take $\alpha < 0$, the same can be said for the faster part, if we restrict ourselves to the lower derivatives.

We study next the case $\alpha = 0$, e.g., when the slow and fast signals are of equivalent magnitude.

### *2.2.2 Asymptotic degeneracy of the excitation matrix*

We assume now that $\alpha = 0$, e.g. $u(t) = u_1(t) + u_2(\varepsilon t)$.

**Theorem 2 (Ill-conditioning of the excitation matrix)** *Let $U_1(i\omega)$ and $U_2(i\omega)$ the Fourier transforms of $u_1$ and $u_2$. Assume that there exists $n_1 > 0$ and $n_2 > 1/2$ such that $U_1(i\omega) \leq |\omega|^{n_1}$ in the neighborhood of 0, and that $|U_2(i\omega)| \leq |\omega|^{-n_2}$ in a neighborhood of $\infty$.*

*Let $y = T_\varepsilon u$ the output of the plant (1), $Z = (y^{(n-1)}, \ldots, u, ^{(n-1)}, \ldots, u)^T$, $Exc(\varepsilon) \stackrel{\text{def}}{=} \int_0^{+\infty} Z(t)Z^T(t)dt$ the corresponding excitation matrix of order $n-1$. Then there exists a matrix $Exc(0)$ such that*

– *$Exc(0)$ is degenerate and is independent of $\varepsilon$*

[3] $H^{n-1}(\mathbb{R})$ is the space of the functions whose $n-1$ first derivatives, in the sense of distributions, belong to $L^2(\mathbb{R})$. Its norm is taken as the sum of the $L^2$ norms of these derivatives.

– $Exc(\varepsilon) = Exc(0) + o(\varepsilon)$

**Sketch of proof:** considering the assumptions on $U_1$ and $U_2$, the inner product $< u_1(t), u_2(\varepsilon t) >$ tends to zero with $\varepsilon$. Extending this analysis to all of the elements of the excitation matrix, one sees that the dominating terms in $Exc(\varepsilon)$, for a small $\varepsilon$, are obtained by replacing $u_1$ by 0, replacing the derivatives of $u_2$ and $T_\varepsilon u_2$ by 0, and the replacing the response $T_\varepsilon u_2$ by its slow approximation $T_n(0)T_s u_2$. This defines the matrix $Exc(0)$, which is indeed degenerate.

**Remark** This proves that using a least squares method to identify a two time scaled system will give a poor accuracy on the system parameters. From the structure of $Exc(0)$, one can see that the only parameter that is asymptotically identifiable is the static gain. This is confirmed qualitatively by the experiments of section 4.

In the next section, we are going to expose a new scheme which, by explicitly modeling the time scales, eliminates this ill-conditioning problem when $\varepsilon$ is small.

## 3 A bifocal method for identifying two time scaled systems

**Presentation** The problems described by theorem 2 are essentially due to the inhomogeneity of the signals (some are slow, and some are fast), and of the system parameters (some are small, others are bigger). This homogeneity can be recovered by handling separately the slow and the fast parts of the system. While this will require some model approximations within each time scale, we will be nonetheless able to identify the parameters with a suitable accuracy when $\varepsilon$ is small enough.

### *3.1 Local approximations for two time scaled systems*

We have the following trivial (but useful) limit results:

**Lemma 3**

$$\lim_{\varepsilon \to 0} T_\epsilon(i\omega) = T_s(i\infty)T_n(i\omega) \stackrel{\text{def}}{=} \tau_n(i\omega) \tag{3}$$

$$\lim_{\varepsilon \to 0} T_\epsilon(\varepsilon i\omega) = T_s(i\omega)T_n(0) \stackrel{\text{def}}{=} \tau_s(i\omega) \tag{4}$$

$\tau_n$ will be called the fast transfer, and $\tau_s$ the slow transfer. Note that the static gain of the fast transfer is equal to the high frequency gain of the slow transfer. Conversely, for a given $\varepsilon$, and assuming that the static gain of $\tau_n$ is equal to the high frequency gain of $\tau_s$, then there exists a unique $T_\varepsilon$ of the form (1) which satisfies the approximations (3) and (4).

**Remarks** The important point in lemma 3 is that the parameters of $\tau_n$ and $\tau_s$ *do not depend on* $\varepsilon$.

Also, the kind of convergence in (3) and (4) is, with respect to $\omega$, a pointwise convergence, and not a uniform one. Indeed, we shall use the prefilters to eliminate the parts of the signals where the convergence fails to be uniform.

**Additional assumptions** For approximations (3) and (4) to give non trivial results, we shall assume that the static gain of $T_n$ is defined and non zero, and that the high-frequency gain of $T_s$ is neither zero (nor infinity), which means that we shall assume that $T_s$ is biproper. If this was not verified, then either the slow or the fast transfer would be completely negectible with respect to the other.

### *3.2 The two time scaled identification problem*

**An outline of the method** The previous considerations lead to the idea what we call a bifocal [4] method to identify the two time scaled system $T_\varepsilon$:

- use a high-pass and a low pass prefilter to generate data for which the approximations (3) and (4) are uniform
- using this data, identify separately $\tau_n$ and $\tau_s$ under the constraint that the static gain of the first transfer should be equal to the high frequency gain of the second one.

Note that the identification of $\tau_s$ involves a time rescaling of the model, e.g., the use of an operator that is different from the one used to identify $\tau_n$.

#### *3.2.1 Prefilters requirements*

Prefiltering has been known to increase the performance of identification schemes by transforming the shape of noises and reconditioning the excita-

[4] This is a reference to spectacles with bifocal lenses, which allow their user to see both the far and the close view.

tion matrix (Middleton and Goodwin 1990, Landau 1988, Ljung 1987). We use it here to validate the use of the reduced models (3) and (4).

From a theoretical point of view, here is what we will require from the pre-filters:

**Definition 4** *A high-pass filter $F_h$ is asymptotically suited to the identification of the fast subsystem if*

$$\lim_{\varepsilon \longrightarrow 0} \max_{\omega \in \mathbb{R}} \left| (T_s(\frac{i\omega}{\varepsilon}) - T_s(i\infty)) F_h(i\omega) \right| = 0 \tag{5}$$

*A low-pass filter $F_l$ is asymptotically suited to the identification of the slow subsystem if $F_l(0)$ is defined and $\neq 0$, and if, for $0 \leq k < n_s$,*

$$\lim_{\varepsilon \longrightarrow 0} \max_{\omega \in \mathbb{R}} \left| (T_n(\varepsilon i\omega) - T_n(0)) \omega^k F_l(i\omega) \right| = 0 \tag{6}$$

These asymptotic conditions can be achieved by taking

- $F_h(s) = s^k \tilde{F}_h(s)$, with $\tilde{F}_h$ rational, causal, without pole or zero at 0, and $k$ strictly greater than the number of poles or zeroes at 0 of $T_s$
- $F_l$ rational with relative degree greater or equal to $n_s$, and no pole or zero at 0

### *3.2.2 The two time scaled identification problem*

The identification procedure, as in the least squares method, will be in fact an optimization procedure.

It will consist in the local identification of $\tau_n$ and $\tau_s$ by minimization of the sum of their prediction errors, under the constraint that the high frequency gain of $\tau_s$ be equal to the static gain of $\tau_n$.

We denote by $\Psi_n(s)$ the polynomial vector $\Psi_n(s) = [s^{n-1}, s^{n-2}, \ldots, s, 1]^T$, by $n_n$ the order of $T_n$, and by $n_s$ the order of $T_s$. We parameterize $\tau_n$ as

$$\tau_n(s) = \frac{\nu_n^T \Psi_{n_n}(s)}{1 - s\delta_n^T \Psi_{n_n}(s)} \tag{7}$$

which is not classical, but makes the constraint linear in the optimization problem, and we parameterize $\tau_s$ as

$$\tau_s(s) = \frac{\nu_s^T \Psi_{n_s}(s)}{s^{n_s} - \delta_s^T \Psi_{n_s}(s)} \tag{8}$$

the parameters being $\delta_s$, $\nu_s$, $\delta_n$ and $\nu_n$.

The identification problem is the following:

**Definition 5 (Two time scaled least squares)** *Let $F_l$ a suitable low-pass filter, $F_h$ a suitable high-pass filter. The two time scaled identification problem is:*

$$\begin{aligned} \min_{\delta_s, \nu_s \in \mathbb{R}^{n_s}, \delta_n, \nu_n \in \mathbb{R}^{n_n}} \quad & J_s(\delta_s, \nu_s) + J_n(\delta_n, \nu_n) \\ \nu_s(1) = \nu_n(n_n + 1) & \end{aligned} \tag{9}$$

*with*

$$J_s(\delta_s, \nu_s) = \frac{1}{\|u_s\|^2_{H^{n_s-1}}} \times \int_0^{+\infty} \left[ \frac{d^{n_s} y_s}{d\tau^{n_s}} - \delta_s^T \Psi_{n_s}\left(\frac{d}{d\tau}\right) y_s - \nu_s^T \Psi_{n_s}\left(\frac{d}{d\tau}\right) u_s \right]^2 d\tau \tag{10}$$

$y_s(t) = F_l[y(t/\varepsilon)]$, $u_s(t) = F_l[u(t/\varepsilon)]$, *and*

$$J_n(\delta_n, \nu_n) = \frac{1}{\|u_n\|^2_{H^{n_n-1}}} \times \int_0^{+\infty} \left[ y_n - \delta_n^T \Psi_{n_n}\left(\frac{d}{dt}\right) \frac{dy_n}{dt} - \nu_n^T \Psi_{n_n}\left(\frac{d}{dt}\right) u_n \right]^2 dt \tag{11}$$

$y_n(t) = [F_h y](t)$, $u_n(t) = [F_h u](t)$.

**Comments:**

- as we noticed after lemma 3, we can recover, for a given $\varepsilon$, the parameters of $T_\varepsilon$ from the solution of this problem. This conversion has a unique result.
- $J_n$ and $J_s$ are the prediction error costs on the reduced models $y_n = \tau_n u_n$ and $y_s = \tau_s u_s$, using the prefiltered signals.
- The prediction errors are essentially linear with respect to the norm of the inputs in the Sobolev space $H^{n_n-1}$; this is why we use the latter to normalize the cost.

**About model matching in the two time scaled method** The reader will notice that, for $\varepsilon \neq 0$, the previous identification procedure is not a model matching one for systems of the form (1), that is, there exists no set of parameters that gives a zero cost in (9) for all inputs. If it was, this would imply that we have two exact models with an order which is lower [5] than that

[5] the prefilters cancel each other in the slow and fast transfers

of $T_\varepsilon$. We will show in section 3.4 that, instead, the two time scaled method is *asymptotically* model matching when $\varepsilon$ tends to 0.

### *3.2.3 Strong convexity and existence of persistent signals*

In the least squares method, persistent excitation is another formulation for the strong convexity of the optimization cost. We study here some sufficient conditions, under which the slow and fast excitation matrices in problem (9) have a non zero lower bound on their eigenvalues which is independent of $\varepsilon$, that is, when we have a non vanishing excitation property.

Let us consider again the case where the input is of the form $u = u_1(t)+u_2(\varepsilon t)$. We assume that the Fourier transforms $U_1$ and $U_2$ of $u_1$ and $u_2$ satisfy

$$|\omega^{-n_1} U_1(i\omega)| \underset{\varepsilon \to 0}{\longrightarrow} 0 \text{ for some } n_1 > 0 \tag{12}$$

$$|\omega^{n_2} U_2(i\omega)| \underset{\varepsilon \to 0}{\longrightarrow} 0 \text{ for some } n_2 > \tfrac{1}{2} \tag{13}$$

Because of the requirements that we made on the low pass and high pass filters, the terms coming from $u_2$ vanish in $J_n$ and the terms coming from $u_1$ vanish in $J_s$ when $\varepsilon$ tends to 0.

Hence, the existence of persistent excitation in $J_s$ and $J_n$ is asymptotically equivalent to the same question where one considers only the signals *relevant to the corresponding time scale*. As a consequence, persistent excitation for the two time scaled identification problem is satisfied if the slow part $F_l[u_2(t)]$ (resp. fast part $F_h[u_1(t)]$) of the input data provides persistent excitation with respect to the slow (resp. fast) subsystem. This condition does not involve $\varepsilon$. It is satisfied if $u_1$ and $u_2$ are persistently exciting with respect to their subsystem, and if the filters, while eliminating the undesired part of the signal, still preserve this excitation.

As a consequence, it is possible to design $F_l$, $F_h$, $u_1$ and $u_2$ in order to give $J_s$ and $J_n$ a strong convexity constant $\mu$ that does not depend on $\varepsilon$.

## *3.3 An optimization-identification algorithm*

We give here an example of a solving algorithm for problem (9).

**Principle of the algorithm** A nice feature of problem (9) is that it is a quadratic problem with linear constraints; this is due to our special parameterization (7) of $\tau_n$. Hence, it has a saddle point, and we can replace the constrained problem (9) by:

$$\max_{p\in\mathbb{R}} \min_{\delta_s,\nu_s\in\mathbb{R}^{n_s},\delta_n,\nu_n\in\mathbb{R}^{n_n}} \mathcal{L}(\delta_s,\nu_s,\delta_n,\nu_n,p) \tag{14}$$

the Lagrangian $\mathcal{L}$ being defined by

$$\mathcal{L}(\delta_s,\nu_s,\delta_n,\nu_n,p) = J_s(\delta_s,\nu_s) + J_n(\delta_n,\nu_n) + p(\nu_s(1) - \nu_n(n_n+1)) \tag{15}$$

$\mathcal{L}$ can be decomposed as $\mathcal{L}_s(\delta_s,\nu_s,p) + \mathcal{L}_n(\delta_n,\nu_n,p)$ with

$$\mathcal{L}_s(\delta_s,\nu_s,p) = J_s(\delta_s,\nu_s) + p\nu_s(1) \tag{16}$$
$$\mathcal{L}_n(\delta_n,\nu_n,p) = J_n(\delta_n,\nu_n) - p\nu_n(n_n+1) \tag{17}$$

We can use then a Uzawa algorithm to solve (14). It uses a gradient algorithm on the max problem as follows:

$$(\delta_s^{k+1},\nu_s^{k+1}) = \arg\min \mathcal{L}_s(\delta_s^k,\nu_s^k,p^k) \tag{18}$$
$$(\delta_n^{k+1},\nu_n^{k+1}) = \arg\min \mathcal{L}_n(\delta_n^k,\nu_n^k,p^k) \tag{19}$$
$$p^{k+1} = p^k + \rho(\nu_s^k(1) - \nu_n^k(n_n+1)) \tag{20}$$

The minimizing arguments are uniquely defined if we have the persistent excitation property for each reduced model. We shall some study sufficient conditions for this excitation/convexity requirements in section 3.2.3.

**Explicitation of the algorithm** Let us denote by $Q_s$ and $Q_f$ the excitation matrices (or Hessians) in the slow and fast prediction costs, and by $R_s$ and $R_f$ the column vectors that define the parts of the costs which are linear with respect to the parameters.

The Uzawa algorithm can be written now as:

**Algorithm 1 (Identification algorithm)**

$$\begin{bmatrix} \delta_s^{k+1} \\ \hline \\ \nu_s^{k+1} \end{bmatrix} = -\frac{1}{2} Q_s^{-1} \left\{ R_s + \begin{bmatrix} 0 \\ \hline \\ p^k \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\} \tag{21}$$

$$\begin{bmatrix} \delta_n^{k+1} \\ \hline \\ \nu_n^{k+1} \end{bmatrix} = -\frac{1}{2} Q_n^{-1} \left\{ R_n - \begin{bmatrix} 0 \\ \hline \\ 0 \\ \vdots \\ 0 \\ p^k \end{bmatrix} \right\} \tag{22}$$

$$p^{k+1} = p^k + \rho(\nu_{s,n_s}^k - \nu_{n,1}^k) \tag{23}$$

**Convergence of the algorithm** The algorithm converges then if $\rho$ is small enough. When the Hessian of the cost is a constant symmetrical matrix $A$ and the constraint is linear with a constant derivative $D$, a theoretical bound for $\rho$ is given by the ratio $\lambda_1(A)/\|D\|^2$ where $\lambda_1(A)$ is the smallest eigenvalue of $A$ (see Arrow, Hurwicz and Uzawa 1972 for details).

### *3.4 Convergence results on the identification*

We are going to show that, when $\varepsilon$ tends to 0, the plant's parameters have a prediction error in (9) which tends to 0.

We will then use the excitation assumption to show that the optimal parameters, as computed by solving (9), converge towards the plant's parameters when $\varepsilon$ tends to 0.

#### *3.4.1 Convergence of the prediction costs when using the plant's parameters*

**Notations and assumptions** We shall denote by $\delta_s^*, \nu_s^*, \delta_n^*, \nu_n^*$ the unique set of parameters in (7) and (8) such that

$$\nu_n^{*T}\Psi_{n_n}(s) = N_n(s)\ 1 - s\delta_n^{*T}\Psi_{n_n}(s) = D_n(s) \tag{24}$$
$$\nu_s^{*T}\Psi_{n_s}(s) = N_s(s)\ s^{n_s} - \delta_s^{*T}\Psi_{n_s}(s) = D_s(s) \tag{25}$$
$$T_n = \frac{N_n}{D_n}\ T_s = \frac{N_s}{D_s} \tag{26}$$

where $N_n$ and $D_n$ (resp. $N_s$ and $D_s$) are relatively prime. We will also define $\nu_s^{\#} = T_n(0)\nu_s^*$ and $\nu_n^{\#} = T_s(\infty)\nu_n^*$.

Finally, we shall assume that $u$ is of the form $u(\varepsilon, t) = u_1(t) + u_2(\varepsilon t)$, with $u_1$ and $u_2$ that satisfy (12) and (13). As we said in section 3.2.3, we can asymptotically consider that $u_1 = 0$ in $J_s$, and $u_2 = 0$ in $J_n$. This is due to the prefilters' design.

**Theorem 6 (Convergence of the slow cost)** *$J_s(\delta_s^*, \nu_s^{\#})$ tends to 0 when $\varepsilon$ tends to 0.*

**Proof:** let $U$ the Fourier transform of the input $u$. We first notice that the denominator of the cost, e.g., $\|u_s\|^2_{H_{n_s-1}}$, is equivalent to the norm of $F_l u_2$ as $\varepsilon$ tends to 0. We have

$$\begin{aligned}
J_s(\delta_s^*, \nu_s^{\#}) &= \frac{1}{\|u_s\|^2_{H^{n_s-1}}} \int_{-\infty}^{+\infty} \left|\left[T_n(\varepsilon i\omega) - T_n(0)\right] N_s(i\omega) F_l(i\omega) \varepsilon U(\varepsilon i\omega)\right|^2 d\omega \\
&\underset{\varepsilon \to 0}{\approx} \frac{1}{\|F_l u_2\|^2_{H^{n_s-1}}} \int_{-\infty}^{+\infty} \left|\left[T_n(\varepsilon i\omega) - T_n(0)\right] N_s(i\omega) F_l(i\omega) \varepsilon U(\varepsilon i\omega)\right|^2 d\omega \\
&\leq \frac{1}{\|F_l u_2\|^2_H} \| \left[T_n(\varepsilon i\omega) - T_n(0)\right] N_s(i\omega) F_l(i\omega)\|^2_{L^\infty} \|u(t/\varepsilon)\|^2_{L^2} \\
&\leq \frac{\|u(t/\varepsilon)\|^2_{L^2}}{\|F_l u_2\|^2_H} \max_{k<n_s} \| \left[T_n(\varepsilon i\omega) - T_n(0)\right] \omega^k F_l(i\omega)\|^2_{L^\infty} \|\nu_s^*\|^2_\infty \\
&\underset{\varepsilon \to 0}{\approx} \frac{\|u_2\|^2_{L^2}}{\|F_l u_2\|^2_H} \max_{k<n_s} \| \left[T_n(\varepsilon i\omega) - T_n(0)\right] \omega^k F_l(i\omega)\|^2_{L^\infty} \|\nu_s^*\|^2_\infty
\end{aligned}$$

which proves the convergence for a suitable low-pass filter, by definition of the latter property.

**Theorem 7 (Convergence of the faster cost)** *$J_n(\delta_n^*, \nu_n^{\#})$ goes to zero with $\varepsilon$.*

**Proof:** as we said before, we can neglect $u_2$ in the cost.

$$J_n(\delta_n^*, \nu_n^{\#}) \approx \frac{1}{\|F_h u\|^2_{H^{n_n-1}}} \int_{-\infty}^{+\infty} \left| \left[T_s(\frac{i\omega}{\varepsilon}) - T_s(\infty)\right] N_n(i\omega) F_h(i\omega) U_1(i\omega) \right|^2 d\omega$$

$$\leq \|(T_s(\frac{i\omega}{\varepsilon}) - T_s(i\infty))F_h(i\omega)\|_\infty^2 \|\nu_s^*\|^2 \frac{\|u_1\|_H^2}{\|F_h u_1\|_H^2}$$

which tends to zero with $\varepsilon$.

### *3.4.2 Convergence of the identified parameters*

Now that we have shown that the desired parameters are asymptotically optimal, a strong convexity argument is enough to show that the identified parameters will also converge towards the right values.

**Assumptions** We shall assume that $u$, $F_l$ and $F_h$ are such that, if $y$ is the output of $u$ through the transfer $T_\varepsilon$ of section 2.1, then $J_s$ and $J_n$ are strongly convex with constant $\mu$, with $\mu$ independent of $\varepsilon$. This can be achieved by satisfying the conditions of section 3.2.3.

**Theorem 8 (Convergence of the parameters)** *The identified parameters in (9) converge to the quadruplet $\delta_s^*, \nu_s^\#, \delta_n^*, \nu_n^\#$ as $\varepsilon$ goes to zero.*

**Proof:** remark first that the proposed quadruplet forms an admissible set of parameters. Let us denote by $\theta_0$ this quadruplet, and by $\theta(\varepsilon)$ the (unique) minimizing argument of (9). A classical result from the convexity theory (Ekeland and Temam 1976) says that

$$\begin{aligned}\mu\|\theta_0 - \theta(\varepsilon)\|^2 &\leq J(\theta_0) - J(\theta(\varepsilon)) - < J'(\theta_0), \theta_0 - \theta(\varepsilon) > \\ &= J(\theta_0) - J(\theta(\varepsilon)) \leq J(\theta_0)\end{aligned}$$

with $J = J_s + J_n$. This proves the result, thanks to the two previous theorems on the convergence of the costs.

**Comment** since the parameters of $T_\varepsilon$ can be uniquely recovered from the solution of (9), the previous theorem implies that the two time scaled method allows the asymptotically exact identification of $T_\varepsilon$ when $\varepsilon$ tends to 0.

### *3.5 About the robustness of the method*

We have just seen that the identified parameters have a finite limit when $\varepsilon$ tends to 0. We have also seen that $F_l$, $F_h$, $u_1$ and $u_2$ can be designed in order to have a $\mu$ that is independent of $\varepsilon$. Since the consistency constraint does not depend on $\varepsilon$, one can see that the solution of (9) is continuous with respect to

elementary perturbations of the identification data [6], uniformly with respect to $\varepsilon$ when $\varepsilon$ is close to 0.

Since the poles and zeroes of $T_\varepsilon$ depend smoothly on those of $\tau_n$ and $\tau_s$, we see that the reconstructed parameters of $T_\varepsilon$, whether they are expressed in linear form or pole/zero form, depend smoothly on the filtered and time rescaled identification data, uniformly in $\varepsilon$.

Because of the design of the prefilters, the latter depends itself smoothly on the original input/output data; this proves that the accuracy of the two time scaled identifiers is robust to the small values of $\varepsilon$.

## 4 A numerical illustration

### *4.1 Design of the two time scaled identifier*

Before going into the details of the numerical experiments, let us give a view of the authors' experience with the design of the two time scaled identifier.

If we omit the choice of the orders of the various systems, the identifier is defined by the choice of three design parameters: the time scales ratio $\varepsilon$, the low pass filter and the high pass filter.

#### *4.1.1 Time scaling*

In addition to the time rescaling of the slow part of the data, as described in the theory, it may be helpful to rescale the fast part of the data, in the same way that it would help in the identification of a classical system.

In the following experiments, we used for each subproblem a time step which was equal to 1/4 of the slowest time constant of each subsystem. The normalization factor of the theory was not actually used; we relied on a suitable time recaling to enhance the conditioning of the matrices.

#### *4.1.2 Filter design*

The choice of the filters, on the other hand, is essential in order to make the two time scaled method work. They should legitimate the model approximations, without losing the exciting signals. Our experience is that, when the

[6] that is, a perturbation of the matrices' elements

subsystems' time constants are not too far away, it seems advisable to allow some overlapping of the prefilters, in order to avoid losing useful data. On the contrary, when the time scales are quite distant, disconnected filters help in the identification of the subsystems.

### *4.2 Summary of the experiments*

Following the ideas of the previous sections, we wrote a simple MATLAB® program in order to compare, on a few examples, the performances of the least squares method against that of its two time scaled counterpart. [7]

The plants we considered were second order systems, of the form

$$T(s) = K\frac{s - Z}{(s - P(1))(s - P(2))} \tag{27}$$

A random input was first generated to help excite the system. Its average was kept non zero because this helped the two time scaled identifier without really affecting the classical least squares.

The output was then computed using a model matching simulation.

According to the value `digits`, a standard random noise, with a variance equal to $10^{-\texttt{digits}}$ times the empirical variance of the input, was then added to the input; a similar data corruption was performed on the output.

We used then the classical least squares method to identify the parameters. After setting a right choice of prefilters and time rescaling, we used the same corrupted data to identify the system by using the two time scaled method, as described in definition 1. [8]

The accuracies of the two methods were compared, using different time scales ratios and different levels of data corruption.

[7] If you desire a copy of the source program, please e-mail F. Chaplais at `chaplais@cas.ensmp.fr`.

[8] In this application, the continuous time differentiation is approximated here by the finite difference $(x_{t_{k+1}} - x_{t_k})$/step where step corresponds to the time scale used respectively by the slow or fast subsystem identifier.

### *4.3 Sample numerical results*

We give here the results that were obtained by using the program to identify two different systems. The duration of the signals is always equal to 3000 time steps.

#### *4.3.1 Reading the results*

The numerical results are presented under graphical form.

For each transfer, four sets of bar charts are presented. Each set corresponds a time constant, or the static gain, as labeled at the bottom of the set; it indicates the ratio between yhe identified parameter and its true value. Each bar set features eight bars. The four black bars give the results given by the classical least squares for different noise/signal ratios; the gray bars give the results obtained with the two time scaled method.

The succession of eight bars should be understood as follows: the two first bars give the result for a noise/signal ratio (as described in section 4.2) of $10^{-4}$, the following two to a ratio of $10^{-3}$; then another group of two for $10^{-2}$, and a last one for $10^{-1}$ noise/signal ratio on the input and output.

#### *4.3.2 Data and results*

The time constants of all the transfers below are expressed in time step units.

The first transfer is

$$T(s) = 2\frac{100s+1}{(200s+1)(8s+1)} \tag{28}$$

The slow and fast prefilters that we used are

$$F_l(s) = \frac{1}{(20s+1)^2} \tag{29}$$

$$F_h(s) = 1 - \frac{1}{(50s+1)^2} \tag{30}$$

The identification results are displayed in figure 1.

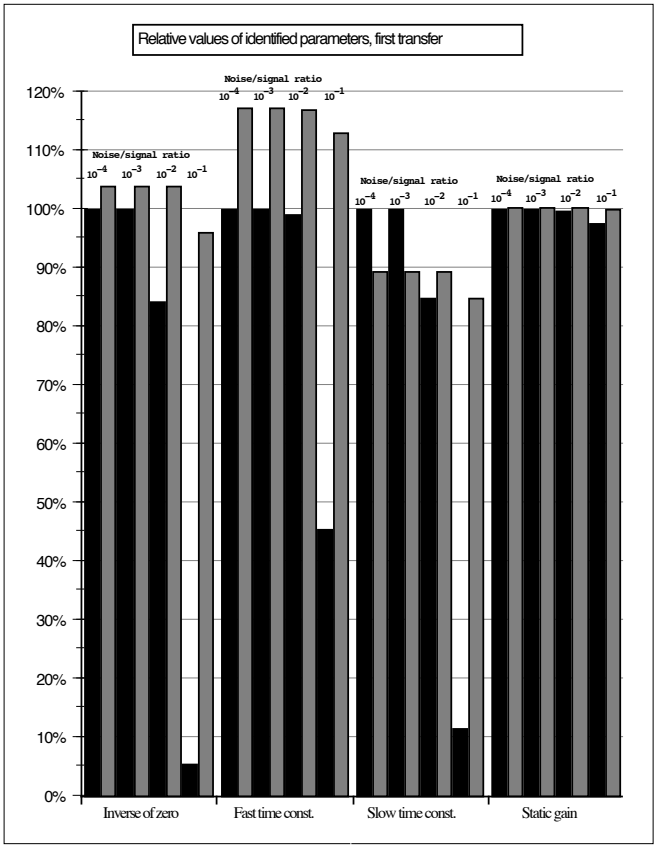


Fig. 1. Results for the first transfer.

The second transfer has more distant time scales:

$$T(s) = 2\frac{500s+1}{(1000s+1)(10s+1)} \tag{31}$$

We used the slow and fast prefilters

$$F_l(s) = \frac{1}{(50s+1)^2} \tag{32}$$

$$F_h(s) = 1 - \frac{1}{(100s+1)^2} \tag{33}$$

The results are in figure 2.

### *4.3.3 Interpretation*

The general interpretation of these results is that, when the noise/signal ratio is small, the classical method works better than the two time scaled one, for the simple reason that the second one is not model matching. When the

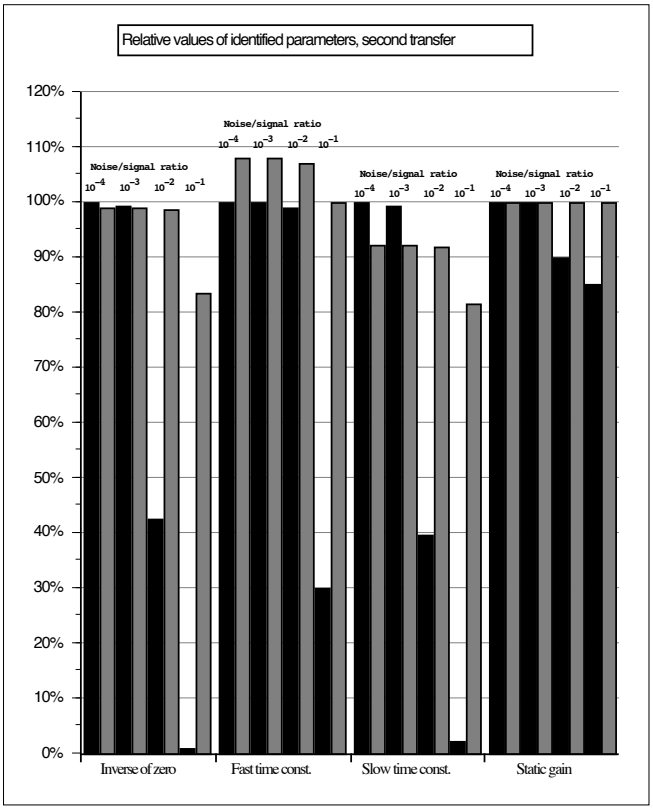


Fig. 2. Results for the second transfer.

noise/signal ratio increases (at $10^{-2}$), the classical least squares gives results that are inaccurate but within the right range, while the two time scaled method manages to keep a good accuracy. Another test, not presented here, shows that, when the noise/signal ratio gets even worse, the classical least squares method even fails to give estimates in the appropriate range, while the second method still gives a good idea of the parameters.

In short, while the least squares gives excellent results when the noise is small and poor results when the noise becomes too important, the two time scaled method is more robust; however, it seems unable to give estimates as accurate as the least squares under the best circumstances. The reason is perhaps that it is not a model matching method in our experiment.

As expected, the gap between the accuracy of the two methods increases with the gap between the time scales.

Finally, these few results confirm that the static gain is the paramete that is the most robustly identified; this is true for the two methods.

## 5 Conclusion

We have given some mathematical insight on the reasons why the classical least squares identification of two-time scaled systems fails to give accurate parameters.

We have proposed a new method for identifying these systems, based on the filtering and time rescaling of the signals, that allows a more reliable parameter identification of the systems in both time scales. The estimated parameters, even though theoretically incorrect, are proved to converge to the plant's parameters as $\varepsilon$ goes to 0, while featuring a non vanishing excitation property.

This is confirmed by a few numerical examples, which shows that the new method is more robust, at the expense of its maximum accuracy.

## References


[1] Alaoui K. (1992), *Identification des systèmes à deux échelles de temps et application au chauffage optimal des bâtiments*, Thesis, Ecole des Mines de Paris, Paris.

[2] Arrow K. J., L. Hurwicz & H. Uzawa, *Studies in Linear and Non Linear Programming*, Stanford University Press, CA, 1972.

[3] Middleton R. H. & G.C. Goodwin (1986), *Improved Finite Word Length Characteristic in Digital Control Using Delta Operators*, IEEE Trans. Aut. Control, **AC-31,1**, pp. 1015-1021.

[4] Ekeland I. & R. Temam (1976), *Convex Analysis and Variational Problems*, North-Holland, Amsterdam.

[5] Luse D. W. & H. K. Khalil (1985), *Frequency Domain Results for Systems with Slow and Fast Dynamics* IEEE Trans. Aut. Control, **AC-30,12**, pp. 1171-1178.

[6] Landau I. D. (1988), *Identification et commande des systèmes*, Hermes, Paris.

[7] Ljung L. (1987), *System Identification, Theory for the User*, Prentice-Hall, Englewood Cliffs.