NONLINEAR SYSTEM IDENTIFICATION IN PRESENCE OF NUISANCE PARAMETERS

N. Ramdani*, T. Poinot**

* CERTES, Université Paris XII - Val de Marne,
IUT de Créteil, Ave Général de Gaulle, 94010 Créteil Cedex
Tel : +33 145 171 836, fax : +33 145 176 551
e-mail : ramdani@univ-paris12.fr

** LAII-ESIP, 40 Ave du Recteur Pineau, 86 022 Poitiers Cedex
Tel: +33 549 453 512, fax +33 549 454 034
e-mail : Thierry.Poinot@esip.univ-poitiers.fr

Keywords: system identification, physical parameters, bayesian inference, nuisance parameters, prior knowledge.

Abstract

This paper deals with the general problem of the estimation of physical parameters in the presence of nuisance parameters for which an explicit characterisation of the uncertainty is available. Within a bayesian context, the maximum *a posteriori* estimation is used. The asymptotic confidence ellipsoid is then determined. The projection of the latter onto the parameter axes is used to obtain the confidence interval. The method is illustrated on an example with actual data from a measurement device in the context of thermo-physical parameters identification.

1 Introduction

This paper is dedicated to the identification of physical parameters when the experimental set-up is subject to nuisance parameters with explicit characterisation of uncertainties.

In the context of physical systems identification, such as thermal processes for instance, the systems studied often encompass, in addition to the unknown quantities to be identified, some extra nuisance parameters usually assumed known but in fact subject to some *a priori* known uncertainty. In the following, it is assumed that there is no intrinsic variability of these nuisance parameters. They are supposed constant but with their actual true values remaining not well known. For geometrical parameters, direct measurements are often possible : a random normal error is usually associated to the parameters. When the measurement is not possible, values are taken from the literature and their prior dispersion law must then be established.

Reducing the influence of nuisance parameters is a well-known topic of parameter identification : optimal experiments are often designed in order to diminish the effect of the nuisance parameters uncertainty and improve the confidence in the identification of the unknown parameters [6].

In this paper, a bayesian inference is used to identify both the unknown quantities and the nuisance parameters while accounting for the prior knowledge of the error in nuisance parameters. In section 2, the most important topics of the bayesian inference are presented. An illustration of the methodology on a thermal experimental set-up is given in section 3.

2 Bayesian inference

Assume a class $\mathsf{M}(\mathbf{p})$ of models is built prior to any measurement on the basis of the sole physical knowledge. Assume there exists a true parameter vector $\mathbf{p}^*$ such as actual data satisfies

$$y = M\left(\mathbf{p}^*\right) + \varepsilon \tag{1}$$

where the error noise ε is assumed additive.

In the bayesian approach, the prior knowledge on the additive errors acting on the model output ε and on the unknown parameters $\mathbf{p}$ is translated to prior probability laws $\pi(\varepsilon)$ and $\pi(\mathbf{p})$. Then, the Bayes rule makes it possible to determine the posterior probability law of the unknown parameters [6]. Since

$$\pi\left(y \mid \mathbf{p}\right)\pi\left(\mathbf{p}\right) = \pi\left(\mathbf{p} \mid y\right)\pi\left(y\right) \tag{2}$$

the posterior probability law for $\mathbf{p}$ when taking the data y into account is given by

$$\pi\left(\mathbf{p} \mid y\right) = \frac{\pi\left(y \mid \mathbf{p}\right)\pi\left(\mathbf{p}\right)}{\pi\left(y\right)} \tag{3}$$

The key issue in Bayesian approaches is then how to express the prior probability laws. The law $\pi\left(y \mid \mathbf{p}\right)$ stands for the likelihood of the measurements y and its expression comes from the information on the noise. The prior law $\pi\left(\mathbf{p}\right)$ is either known or devised thanks to maximum entropy approaches [3].

A point estimator can then be obtained for the unknown parameters, by choosing the value of the *maximum a posteriori*, i.e. which maximizes the posterior law as follows

$$\hat{\mathbf{p}} = \arg\max \pi\left(\mathbf{p} \mid y\right) \tag{4}$$

which is equivalent to write

$$\hat{\mathbf{p}} = \arg\max - J_{MAP}\left(\mathbf{p}\right) \tag{5}$$

where the *maximum a posteriori* criterion is defined by

$$J_{MAP}\left(\mathbf{p}\right) = \ln \pi\left(y \mid \mathbf{p}\right) + \ln \pi\left(\mathbf{p}\right) \tag{6}$$

The maximum a posteriori has the same asymptotic properties than the maximum likelihood estimator. The asymptotic distribution of the estimator $\hat{\mathbf{p}}$ is normal, as follows

$$\hat{\mathbf{p}} \in \mathrm{N}\left(\mathbf{p}*, \mathbf{F}_b^{-1}\left(\mathbf{p}*\right)\right) \tag{7}$$

where $\mathbf{F}_b$ is the bayesian Information matrix.

2.1 Normal prior laws

If the additive noise is assumed normal, centred and with a constant variance σ^2_{noise}, and if the prior parameter law is assumed normal, as follows

$$\mathbf{p} \in N\left(\mathbf{p}_0, \mathbf{\Omega}\right) \tag{8}$$

then, the *maximum a posteriori* criterion writes as follows

$$J_{MAP}\left(\mathbf{p}\right) = \frac{1}{\sigma^2_{noise}} \sum_{k=1}^{n_t} \left(y_k - M\left(\mathbf{p}\right)_k\right)^2 + \left(\mathbf{p} - \mathbf{p}_0\right)^t \mathbf{\Omega}^{-1} \left(\mathbf{p} - \mathbf{p}_0\right) \tag{9}$$

When the noise variance σ^2_{noise} is not precisely know, the Morozov's discrepancy principle [4] can be used: the noise variance is acting as a regularization parameter, it is chosen such that it is equal to the posterior residual variance.

Finally, the bayesian Information matrix is given by

$$\mathbf{F}_b\left(\mathbf{p}*\right) = \mathbf{F}\left(\mathbf{p}*\right) + \mathbf{\Omega}^{-1} \tag{10}$$

where $\mathbf{F}$ is the Fisher Information matrix. In practice, the true parameter vector being unknown, the *maximum a posteriori* estimate is used instead. Finally, a 100(1-η)% confidence ellipsoid for the parameter vector is given by the following set

$$\mathbb{S}_{(1-\eta)} = \left\{ \mathbf{p} \in \mathbb{R}^{n_p} \mid \frac{\left(\mathbf{p} - \hat{\mathbf{p}}\right)^t \mathbf{F}_b\left(\hat{\mathbf{p}}\right)\left(\mathbf{p} - \hat{\mathbf{p}}\right)}{n_p f_{(\eta)}\left(n_p, n_t - n_p\right)} \leq 1 \right\} \tag{11}$$

where the threshold η is chosen as say $\eta = 0{,}05$, f_η is an outcome of a Fisher-Snedecor law, n_p is the number of parameters and n_t, the number of samples [6].

In order to derive a confidence interval for each component of the parameter vector, the authors suggest to project onto the parameter axes the confidence ellipsoid defined by equation (11). The so-derived intervals will then account for any correlation between the identified parameters and thus, constitute outer enclosures of the confidence intervals.

Now, one must note that an ellipsoid defined by

$$\mathbb{E} = \left\{ \mathbf{p} \in \mathbb{R}^{n_p} \mid \left(\mathbf{p} - \mathbf{q}\right)^t \mathbf{P}^{-1} \left(\mathbf{p} - \mathbf{q}\right) \leq 1, \mathbf{P} > 0 \right\} \tag{12}$$

where $\mathbf{q}$ is the centre of $\mathbb{E}$ and $\mathbf{P}$ a definite positive matrix, is also defined by [1]

$$\mathbb{E} = \left\{ \mathbf{p} \in \mathbb{R}^{n_p} \mid \mathbf{q} + \mathbf{V}\mathbf{v}, \mathbf{v} \in B^{n_p} \right\} \tag{13}$$

where B^{n_p} is the unity ball of $\mathbb{R}^{n_p}$ and the matrix $\mathbf{V}$ satisfies

$$\mathbf{P} = \mathbf{V}\mathbf{V}^t \tag{14}$$

The projection $\Pi_i(\mathbb{E})$ of $\mathbb{E}$ onto the axis of the i^{th} component of the parameter vector is the interval defined by

$$\Pi_i\left(\mathbb{E}\right) = \left[\underline{p}_i, \overline{p}_i\right] = \left[q_i - \tilde{v}_i, q_i + \tilde{v}_i\right] \tag{15}$$

Denote $\mathbf{V}_i$, the i^{th} row of the $\mathbf{V}$ matrix, then the quantity $\tilde{v}_i$ is given by non-linear constrained optimization, as follows

$$\tilde{v}_i = \max_{\|\mathbf{v}\|=1} \left(\mathbf{V}_i \mathbf{v}\right) \tag{16}$$

Equations (15)-(16) will be used to derive a confidence interval for the maximum *a posteriori* estimate defined by (5).

3 The experimental procedure

The experimental procedure under analysis hereafter is devoted to the measurement of the thermal properties of materials : the thermal diffusivity and the thermal conductivity of a sample are measured simultaneously by using a so-called *periodic* method, using multi-harmonic heating signals.

3.1 The set-up

The experimental set-up is shown on figure 1. The sample under study is fixed within a metallic rack, with a glue of very large conductivity. The front side of the rack, made of brass, is also fixed to a heating device. The rear side, made of copper, is in contact with air at ambient temperature. To reduce lateral heat losses, radiative shields are used. The heating sequence is composed of 5 sinusoids, which frequencies are the following: {0,781 ; 1,562 ; 3,125 ; 6,25 and 12,5 *mHz* } [5].

The temperatures of rear and front sides are measured at 100*Hz* sample rate with the thermocouples put as indicated in figure 1. Data are then low-pass filtered and under-sampled at 0,4*Hz* sample rate. The temperature spectra, taken as the Fourier transform of the time-history signals, are then used to estimate the experimental frequency response, as follows

$$H_S\left(j\omega\right) = \frac{T_{rear}\left(j\omega\right)}{T_{front}\left(j\omega\right)} \tag{17}$$

The duration of each experiment is equal to 85 min and the experiment is repeated 30 times.

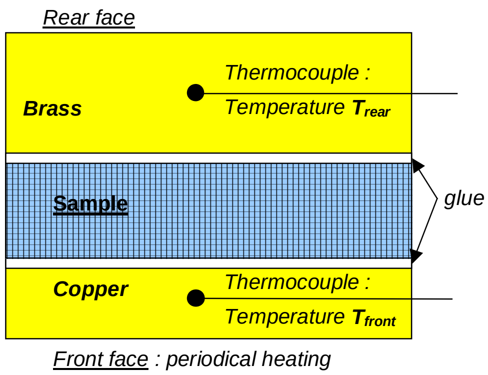

Figure 1: The set-up.

3.2 The model

The system under study is modelled with one-dimensional quadrupoles (two-port transfer functions). The quadrupole method is well known and extensively used in thermal sciences [7].

For each layer of homogeneous material (see figure 2), the law of energy conservation writes as follows

$$\frac{\partial^2 T(x,t)}{\partial x^2} = \frac{1}{a}\cdot\frac{\partial T(x,t)}{\partial t} \tag{18}$$

(see Table 1 for the symbol nomenclature).

Define the following variables as boundary conditions :

$$T(x=0,t) = T_0(t) \tag{19}$$

$$T(x=e,t) = T_1(t) \tag{20}$$

$$-\lambda\cdot\left.\frac{\partial T(x,t)}{\partial x}\right|_{x=0} = \varphi_0(t) \tag{21}$$

$$-\lambda\cdot\left.\frac{\partial T(x,t)}{\partial x}\right|_{x=e} = \varphi_1(t) \tag{22}$$

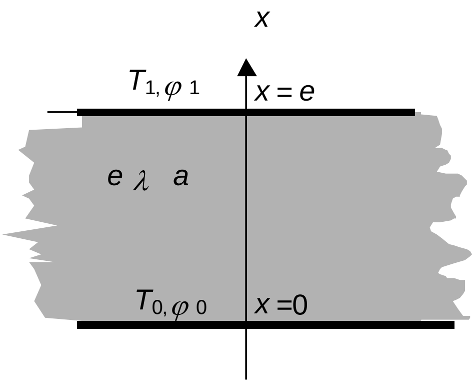

Figure 2: Homogeneous material.

	Thermal models
$a=\lambda/\rho c_p$	thermal diffusivity (Ws/kgK)
c_p	specific heat (Ws/kgK)
e	thickness (m)
h	surface heat exchange coefficient (W/m^2K)
R	thermal resistance e/λ (K/W)
s	Laplace variable
t	time (s)
T_0	front face temperature (K)
T_1	rear face temperature (K)
x	one-dimensionnal coordinate (m)
$\mathbf{Z}(.)$	thermal "quadrupole"
φ_0	front face feat flux (W/m^2)
φ_1	rear face heat flux (W/m^2)
λ	thermal conductivity (W/mK)
ρ	density (kg/m^3)
	Bayesian inference
B^{n_p}	unity ball of $\mathbb{R}^{n_p}$
$\mathbf{F}(.)$	Fisher information matrix
f_η	outcome of a Fisher-Snedecor law
$H_S(.)$	experimental frequency response
n_p	number of parameters
n_t	number of data samples
$\mathbf{P}$	ellipsoid shape and orientation
$\mathbf{p}$	parameter vector
$\mathbf{p}_0$	parameter vector prior mean
$\mathbf{q}$	ellipsoid centre
$\mathbb{S}_{(1-\eta)}$	100(1-η)% confidence ellipsoid
y_k	sample k of actual measured data
$\Pi_i(\mathbb{E})$	projection of $\mathbb{E}$ onto the i^{th} axis
σ^2_{noise}	noise prior variance
σ_i	nuisance parameters prior standard-error
ζ_i%	confidence ratio
$\mathbf{\Omega}$	parameter vector prior variance

Table 1: Symbol nomenclature.

The Laplace transform of equation (18) is as follows :

$$\frac{\partial^2 T(x,s)}{\partial x^2} = \frac{s}{a}\cdot T(x,s) \tag{23}$$

The boundary conditions becomes

$$T(x=0,s) = T_0(s) \tag{24}$$

$$T(x=e,s) = T_1(s) \tag{25}$$

$$-\lambda\cdot\left.\frac{\partial T(x,s)}{\partial x}\right|_{x=0} = \varphi_0(s) \tag{26}$$

$$-\lambda\cdot\left.\frac{\partial T(x,s)}{\partial x}\right|_{x=e} = \varphi_1(s) \tag{27}$$

Material	*Parameters*	*Scale*	*Nominal value*	*Standard -error*
Brass sheet	Diffusivity, *a*	10^{-6} m^2.s^{-1}	34,0	0,50
	Conductivity, λ	W.m^{-1}.K^{-1}	105,0	2,50
	Brass thermocouple – PVC interface distance	10^{-3} m	3,0	0,50
Copper sheet	Diffusivity, *a*	10^{-6} m^2.s^{-1}	114,0	0,50
	Conductivity, λ	W.m^{-1}.K^{-1}	400,0	1,00
	Copper thermocouple – PVC interface distance	10^{-3} m	5,0	1,00
Glue	Thermal resistance	10^{-6} K.m^2.W^{-1}	35,0	7,50
Convection	Surface heat exchange coefficient	W.m^{-2}.K^{-1}	7,5	1,25

Table 2: The nuisance parameters.

The solution of equation (23) is given by

$$\forall x, \quad 0 < x < e,$$
$$T(x,s) = A(s)\cdot\cosh\left(\sqrt{s/a}\cdot x\right) + B(s)\cdot\sinh\left(\sqrt{s/a}\cdot x\right) \qquad (28)$$

Finally, using equations (24)-(27), one obtains the following relationship between the rear face 'temperature – heat flux' couple and the front face one :

$$\begin{bmatrix} T_0(s) \\ \varphi_0(s) \end{bmatrix} = \mathbf{Z}(s)\cdot\begin{bmatrix} T_1(s) \\ \varphi_1(s) \end{bmatrix} \qquad (29)$$

where the "quadrupole" **Z**(*s*) is defined by

$$\mathbf{Z}(s) = \begin{bmatrix} \cosh\left(\sqrt{\frac{s}{a}}\cdot e\right) & \frac{1}{\lambda\cdot\sqrt{\frac{s}{a}}}\cdot\sinh\left(\sqrt{\frac{s}{a}}\cdot e\right) \\ \lambda\cdot\sqrt{\frac{s}{a}}\cdot\sinh\left(\sqrt{\frac{s}{a}}\cdot e\right) & \cosh\left(\sqrt{\frac{s}{a}}\cdot e\right) \end{bmatrix} \qquad (30)$$

For the particular case of the glue layer, which is supposed with no inertia, the relationship uses the resistance only and becomes

$$\mathbf{Z}(s) = \begin{bmatrix} 1 & R \\ 0 & 1 \end{bmatrix} \qquad (31)$$

The model transfer function is then given by

$$H(s,\mathbf{p}) = \frac{T_{rear}(s)}{T_{front}(s)} \qquad (32)$$

where the front temperature is given by (the *s* symbol being removed, for convenience)

$$\begin{bmatrix} T_{front} \\ \varphi_{front} \end{bmatrix} = \mathbf{Z}_{Copper}\cdot\mathbf{Z}_{Glue}\cdot\mathbf{Z}_{Sample}\cdot\mathbf{Z}_{Glue}\cdot\mathbf{Z}_{Brass}\cdot\begin{bmatrix} T_0 \\ hT_0 \end{bmatrix} \qquad (33)$$

and the rear temperature is given by

$$\begin{bmatrix} T_{rear} \\ \varphi_{rear} \end{bmatrix} = \mathbf{Z}_{Brass\_half}\begin{bmatrix} T_0 \\ hT_0 \end{bmatrix} \qquad (34)$$

Although the use of a constant heat surface exchange coefficient is common for modelling heat transfers at ambient temperature, the value actually depends on several parameters such as the sign of the flux, the flow velocity, or the surface tilt angle. However, one does not know how to accurately quantify this parameter when the surface temperature varies as a sinusoidal function.

3.3 The model nuisance parameters

The thermo-physical parameters of brass and copper are taken from the literature. The measurement of the thicknesses of the brass and copper sheets are done with high accuracy. However, the widths of the thermocouples used lead to large uncertainty in the evaluation of the thermocouple – interface distances. The assumed values and uncertainties for the nuisance parameters are given in table 2.

3.4 The experimental results

Prior to any identification procedure, one must check whether this estimation is feasible. Several tools can then be used to assess the parameters identifiability. In particular, the analysis of the model derivatives can give some indications, at least locally.

For the studied system, the analysis of the derivatives shows that the whole parameter vector, including both the unknown and nuisance parameters, is not fully identifiable with regard to the information contained in the data. Consequently, the extra information provided by the prior distributions assumed for the nuisance parameters works as a device for bypassing this loss of identifiability.

Furthermore, the projection formulas as defined by (15)-(16) will serve as a tool for propagating the uncertainty in the nuisance parameters into the values and the confidence

Parameters	*Scale*	*Method*	*Identified value*	*Confidence ratio*
Diffusivity, *a*	10^{-6} $m^2.s^{-1}$	*Maximum likelihood*	*0,125*	*8,6%*
		Maximum *a posteriori*	**0,118**	**14,8%**
Conductivity, λ	$W.m^{-1}.K^{-1}$	*Maximum likelihood*	*0,170*	*2,5%*
		Maximum *a posteriori*	**0,172**	**4,5%**

Table 3: The identified sample thermo-physical parameters

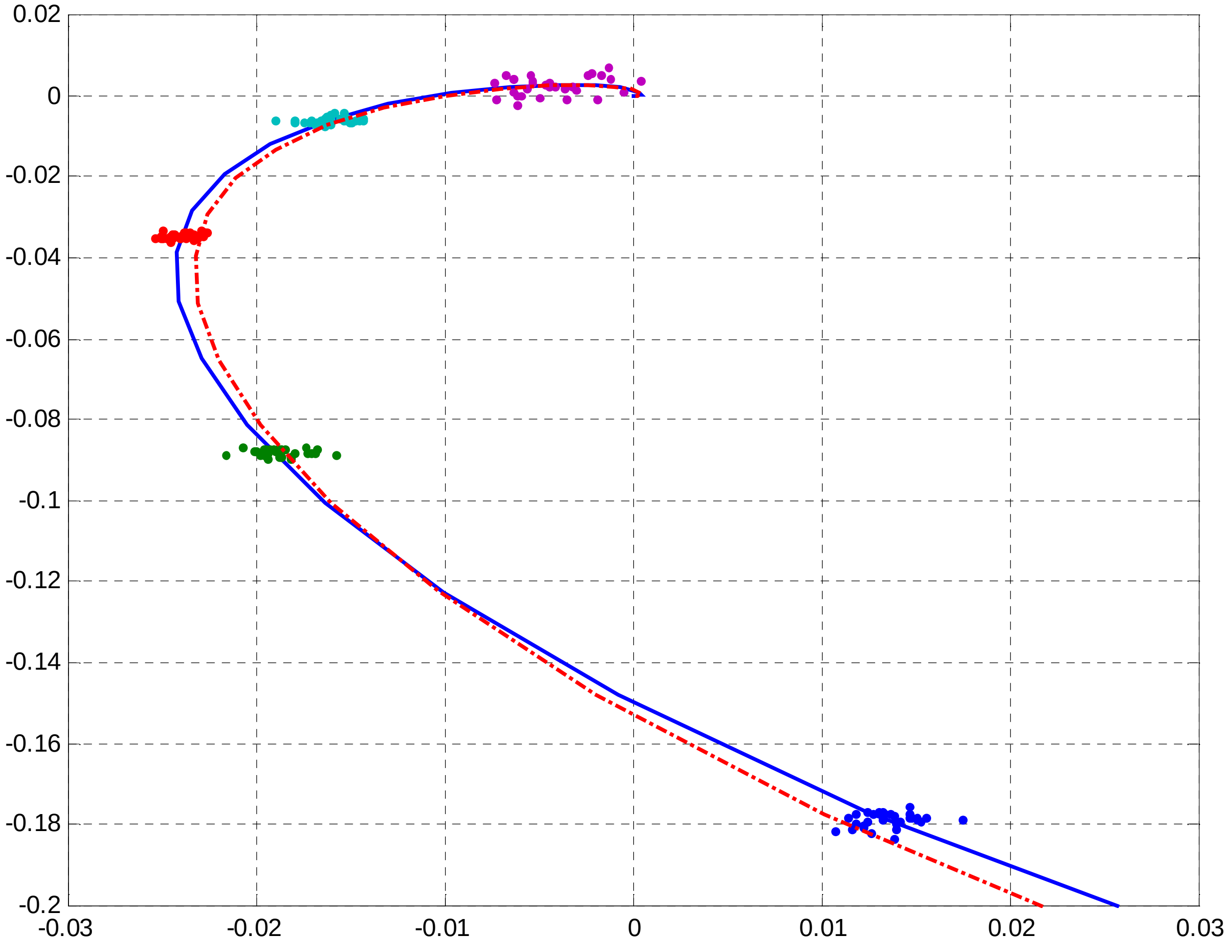

Figure 3: Experimental and models Nyquist plot
(dots : experimental data, continuous line : *maximum a posteriori*, dash-dot : *maximum likelihood*).

intervals derived for the unknown parameters. Hence, for emphasizing the effect of this uncertainty, the confidence intervals obtained with the *maximum a posteriori* estimation are compared to the ones derived when the nuisance parameters are assumed perfectly known.

According to table 2, the *maximum a posteriori* criterion (4) re-writes then as follows:

$$J_{MAP}(\mathbf{p}) = J_{MV}(\mathbf{p}) + \sum_{i=3}^{10}\left[\frac{p_i - p_{0i}}{\sigma_i}\right]^2 \qquad (35)$$

where the *maximum likelihood* criterion corresponds to

$$J_{ML}(\mathbf{p}) = \frac{1}{\sigma_{noise}^2} \sum_{l=1}^{l=30} \sum_{i=1}^{i=5} \left[\begin{array}{l} \left(H_{S,l}\left(j\omega_i\right)_{real} - H\left(j\omega_i,\mathbf{p}\right)_{real} \right)^2 \\ + \left(H_{S,l}\left(j\omega_i\right)_{imag} - H\left(j\omega_i,\mathbf{p}\right)_{imag} \right)^2 \end{array} \right] \quad (36)$$

Note that no prior information is actually used for the unknown parameters in equation (35).

Within the *maximum a posteriori* estimation, the prior noise variance is derived through an iterative procedure: the value of the prior noise variance is chosen such that the prior noise variance matches the variance of the model residuals. It is taken as $\sigma_{noise}^2 = 2{,}08\,10^{-6}$.

The non-linear optimizations are performed with the Levenberg-Marquardt method [6].

The trajectory plots of both models (*maximum a posteriori* and *maximum likelihood*) are given along with experimental data in figure 3. For the *maximum likelihood*, the posterior residuals variance equals $\sigma^2 = 3{,}12\ 10^{-6}$.

The derived parameters for the two cases are given in table 3. The confidence ratio is defined according to (15)–(16), by

$$\zeta_i\% = 100 \frac{\overline{c}_i - \underline{c}_i}{\overline{c}_i + \underline{c}_i} \quad (37)$$

The conclusions are :

- The derived figures for the unknown parameters are in agreement with the rough prior values encountered in the literature [2].
- When the nuisance parameters are assumed perfectly known, the derived uncertainty is fairly large already.
- When the nuisance parameters are taken uncertain, accounting for this uncertainty induces a significant change of the estimated values and provokes an inflation of the size of the confidence intervals. Indeed, when the uncertainty in the nuisance parameters is accounted for, the uncertainty in the identified parameters increases. In addition, the derived posterior noise variances indicate that the residuals derived for the *maximum a posteriori* are smaller than the ones derived for the *maximum likelihood*. This result is also clearly visible on the trajectory plots of figure 3. The *maximum a posteriori* derives then a much better fit to the actual data .

4 Conclusion and Perspectives

Bayesian inference makes it possible to account for the uncertainty in any model nuisance parameter. When this uncertainty is taken into account, the identified model performs a better fit to the actual data but provokes an increase of the size of the unknown parameters confidence intervals, taken as the projection of the confidence ellipsoid onto the unknown parameters axes.

This work will be continued in two ways : In a first step, the estimation of the frequency response will be improved along with a better evaluation of the measurement noise. In a second step, more realistic prior laws will be investigated for the nuisance parameters.

References

[1] C. Durieu, E. Walter, B. Polyak. "Multi-input multi-output ellipsoidal state bounding", *JOTA*, **111**, pp. 273–303, (2001).

[2] A. F. Mills. "Heat Transfert", *Prentice Hall, New Jersey*, (1999).

[3] A. Mohammad-Djafari, J.-F. Giovannelli, G. Demoment, J. Idier. "Regularization, maximum entropy and probabilistic methods in mass spectrometry data processing problems", *International Journal of Mass Spectrometry*, **215**, pp. 175–193, (2002).

[4] W. B. Muniz, F. M. Ramos, H. F. De Campos Velho. "Entropy- and Tikhonov-Based Regularization Techniques Applied to the Backwards Heat Equation", *Computers and Mathematics with Applications*, **40**, pp. 1071-1084, (2000).

[5] E. Tang-Kwor. "Contribution au développement de méthodes périodiques de mesure de propriétés thermophysiques des matériaux opaques", *PhD Dissertation, Univ.Paris XII Val de Marne*, (1998).

[6] E. Walter, L. Pronzato. "Identification of Parametric Models from Experimental Data", *Springer*, (1997).

[7] H. Wang, A. Degiovanni, C. Moyne. "Periodic thermal contact: a quadrupole model and an experiment", *Int. J. Therm. Sci.*, **41**, pp.125–135, (2002).