

Lecture 5. Nonlinear parameter estimation problems: tools for enhancing metrological objectives

B. Rémy, S. André and D. Maillet¹

¹ LEMTA, Université de Lorraine & CNRS, Vandœuvre-lès-Nancy, France

E-mails: benjamin.remy@univ-lorraine.fr
stephane.andre@univ-lorraine.fr
denis.maillet@univ-lorraine.fr

Abstract. The aim of this lecture is to present a methodology for enhancing the estimation of parameters in the case on a Non-Linear Parameter Estimation problem (NLPE). After some definitions and vocabulary precisions, useful tools to investigate NLPE problems will be introduced. Different techniques will be proposed for tracking for instance the true degree of freedom of a given estimation problem (Correlation, Rank of sensitivity matrix, SVD, ..) and enhancing the estimation of particular parameters by using either a Reduced model or a Model with some parameters fixed at their nominal values. The resulting reduced model can be unbiased or biased.

List of acronyms:

- **NLPE:** Non-Linear Parameter Estimation
- **PEP:** Parameter Estimation Problem
- **MBM:** Model-Based Metrology
- **SVD:** Singular Value Decomposition
- **OLS:** Ordinary Least Squares
- **SNR:** Signal-to-Noise Ratio

Scope

1. Introduction
2. Some definitions and vocabulary precisions
3. Useful tools to investigate NLPE problems
 - 3.1 Sensitivities
 - 3.2 Variance/Correlation matrix
 - 3.3 Ill-conditioned PEP and strategies for tracking true degrees of freedom
 - 3.3.1 Pathological example of ill-conditioning resulting from correlated parameters
 - 3.3.2 Rank of the sensitivity matrix.
 - 3.3.3 Generalization: use of SVD to track PEP degrees of freedom
 - 3.3.3.1 Parameterizing a NLPE problem around the nominal values of its parameters
 - 3.3.3.2 Reminder of the Singular Value Decomposition of a rectangular matrix
 - 3.3.3.3 Singular Value Decomposition of the scaled sensitivity matrix
 - 3.3.4 Residuals analysis and signature of the presence of a bias in the metrological process
4. Enhancing the performances of estimation
 - 4.1 Dimensional analysis or natural parameters: case of coupled conduction/radiation flash experiment
 - 4.2 Reducing the PEP to make it well-conditioned: case of thermal characterization of a deposit
 - 4.3 Note on the change of parameters
5. Conclusion

References

- Appendix 1 - Reminder of the Singular Value Decomposition of a rectangular matrix
- Appendix 2 - Singular Value Decomposition of the scaled sensitivity matrix
- Appendix 3 - Non-linear Ordinary Least Square estimator and SVD
- Appendix 4 - Variance-covariance of the Non-linear Ordinary Least Square estimator and SVD
- Appendix 5 – Residual analysis for an unbiased model using the SVD approach

1. Introduction

The Non-Linear Parameter Estimation problem (NLPE) has been the subject of numerous lectures during the past METTI schools (see [1]). This text aims first at gathering in a synthetic way the basic notions and tools that can be used practically to analyse NLPE problems in engineering and science. At the same time, it provides new insights about the tools available to:

- (i) enhance our knowledge about parameter identifiability in a given problem: which parameters can be really estimated in a given experiment and which precision can be achieved?
- (ii) track the origin of pitfalls in parameter estimation problems (PEP),
- (iii) offer new perspectives for enhancing the quality of model-based metrology (MBM) in a general way.

This lecture is composed of three different parts. The first one gives some definitions and vocabulary precisions. The second one presents some useful tools to investigate NLPE: ill-conditioned PEP will be considered and analysed and the use of SVD to track the PEP's degrees of freedom will be introduced next. The last part of this lecture consists in presenting some techniques for enhancing the performances of estimation, such as a dimensional analysis for identifying the degrees of freedom of a given problem and a reduction of the number of parameters involved in a theoretical model to make the PEP well-conditioned. As an example, the case of thermal characterization of a deposit on a substrate will be considered here.

2. Some definitions and vocabulary precisions

Performances of contemporary metrology, that is the science of measurement which includes material characterization for example, are not the result of the enhancement of the technology of measuring instruments only. They are also the consequence of the significant progresses accomplished in the field of Inverse Problems solving, especially when it is based on a very large amount of data. These are provided by new tools and by the facilities now available for numerical acquisition of experimental signals (CCD detectors allowing for 2D/3D numerical data acquisition and high frequency time resolution). Understanding the conditions for which parameters can be estimated from the model/measurements pair constitutes also a key point for reaching a high-quality estimation.

Measuring a physical quantity β_j requires a specific experiment allowing for this quantity to "express itself as much as possible" (notion of sensitivity). This experiment requires a system onto which inputs $u(t)$ are applied (stimuli) and whose outputs $y(t)$ are collected (observations). t is the explanatory variable: it corresponds to time for a purely dynamical experiment. A model M is required to mathematically express the dependence of the system's response with respect to quantity β_j and to other additional parameters β_k ($k \neq j$): $\mathbf{y}_{mo} = \boldsymbol{\eta}(\mathbf{t}; \boldsymbol{\beta}, \mathbf{u})$ where input function $u(t)$ has been parameterized, that is decomposed under a finite set of basis functions, the coefficients of this decomposition being gathered in a vector \mathbf{u} [8, page 26]. Many candidates may exist for function $\boldsymbol{\eta}$ - depending on the degree of complexity reached for modelling the physical process - which may exhibit different mathematical structure – depending for example on the type of method used to

solve the model equations. Once this model is established, the physical quantities in vector β acquire the status of model parameters. This model (called knowledge model if it is derived from physical laws and/or conservation principles) is initially established in a direct formulation. Knowing inputs $u(t)$ and the value taken by parameter β , the output(s) can be predicted.

The linear or non-linear character of the model has to be determined:

- A Linear model with respect to its Inputs (LI structure) is such as:

$$y_{mo}(t; \beta, \alpha_1 u_1(t) + \alpha_2 u_2(t)) = \alpha_1 y_{mo}(t; \beta, u_1(t)) + \alpha_2 y_{mo}(t; \beta, u_2(t)) \quad (1)$$

- A Linear model with respect to its parameters (LP structure) is such as:

$$y_{mo}(t; \alpha_1 \beta_1 + \alpha_2 \beta_2, u(t)) = \alpha_1 y_{mo}(t; \beta_1, u(t)) + \alpha_2 y_{mo}(t; \beta_2, u(t)) \quad (2)$$

In a metrological problem referred here as MBM (Model-Based Metrology), observations of the outputs will be provided by measurements. The inverse problem consists in making the direct problem work backwards with the objective of getting (extracting) β from $y_{mo}(t; \beta, u(t))$ for given inputs and observations y . This is an estimation process. The difficulty stems here from two points:

- (i) Measurements y are subjected to random perturbations (intrinsic noise ε) which in turn will generate perturbed estimated values $\hat{\beta}$ of β , even if the model is perfect: this constitutes an estimation problem.
- (ii) the mathematical model may not correspond exactly to the reality of the experiment. Measuring the value of β in such a context leads to a biased estimation, where the bias is defined as $Bias = E(\hat{\beta}) - \beta^{true}$, $E(\hat{\beta})$ being the expectation of the (stochastic) estimator $\hat{\beta}$: this gives rise to an identification problem (which model structure η to use?) associated to an estimation problem (how to estimate β for a given model structure?).

The estimation/identification process basically tends to make the model match the data (or the contrary). This is made by using some mathematical "machinery" aiming at reducing some gap (distance or norm)

$$r(\beta) = y - y_{mo}(t; \beta, u) \quad (3)$$

One of the obvious goals of NLPE (Non-Linear Parameter Estimation) studies is to assess the performed estimation through the calculation of the variances $V(\hat{\beta})$ of the estimators of the different parameters. If the probabilistic distribution law of the noise is known, this allows to give the order of magnitude of confidence bounds for the estimates. NLPE problems require the use of Non-Linear statistics for studying such properties of the estimates.

Because of the two above-mentioned drawbacks of MBM, the estimated or measured value of a parameter β_j will be considered as "good" if it is not biased (or if its relative bias is low) and if its variance is minimum. Quantifying the bias and variance is also helpful to determine which one of two rival experiments is the most appropriate for measuring the searched parameter (Optimal experiment design). In case of multiple parameters (vector $\boldsymbol{\beta}$) and NLPE problems, it is also interesting to determine which components of vector $\boldsymbol{\beta}$ are correctly estimated in a given experiment.

3. Useful tools to investigate NLPE problems

3.1. Sensitivities

The central role of the sensitivity matrix in PEP has been shown in the preceding lecture (Lecture 3). In the case of a single output signal y with m sampling points for the explanatory variable t and for a model involving n parameters, the sensitivity matrix is ($m \times n$) defined as

$$S_{ij} = \left. \frac{\partial y_{mo}(t; \boldsymbol{\beta}^{nom})}{\partial \beta_j} \right|_{t, \beta_k \text{ for } k \neq j} \quad (4)$$

As the problem is NL, the sensitivity matrix has only a local meaning. It is calculated for a given nominal parameter vector $\boldsymbol{\beta}^{nom}$.

If the model has a LP structure, this means that the sensitivity matrix is independent from $\boldsymbol{\beta}$. It can be expressed as (Lecture 3)

$$y_{mo}(t; \boldsymbol{\beta}) = \sum_{j=1}^n S_j(t) \beta_j \quad (5)$$

The sensitivity coefficient $S_j(t)$ to the j^{th} parameter β_j corresponds to the j^{th} column of matrix \mathbf{S} , once m discrete observation times have been chosen.

The primary way of getting information about the identifiability of the different parameters is to analyse and compare the sensitivity coefficients through graphical observations. This is possible only when considering reduced sensitivity coefficients S_j^* (sometimes called "scaled" sensitivity coefficients) because the parameters of a model do not have in general the same units.

$$\mathbf{S}^* = \boldsymbol{\beta} \mathbf{S} = \boldsymbol{\beta} \left. \frac{\partial \mathbf{y}_{mo}(t; \boldsymbol{\beta}^{nom})}{\partial \boldsymbol{\beta}} \right|_{t, \beta_k \text{ for } k \neq j} = \left. \frac{\partial \mathbf{y}_{mo}(t; \boldsymbol{\beta}^{nom})}{\partial (\ln \boldsymbol{\beta})} \right|_{t, \beta_k \text{ for } k \neq j} \quad (6a)$$

Or

$$\mathbf{S}^* = \mathbf{S} \mathbf{R} \quad (6b)$$

with \mathbf{R} the square diagonal matrix whose diagonal is composed of the components β_j of $\boldsymbol{\beta}$.

TOOL Nr1: A plot of all the reduced sensitivity coefficients $S_j^*(t)$ gives a first idea about the most influential parameter for a given model (largest magnitude) and about possible correlations (sensitivity coefficients following the same evolution).

Example: Measurement of thermophysical properties of a coating layer through the Flash method using thermal contrast principle (Number of parameters $n = 2$).

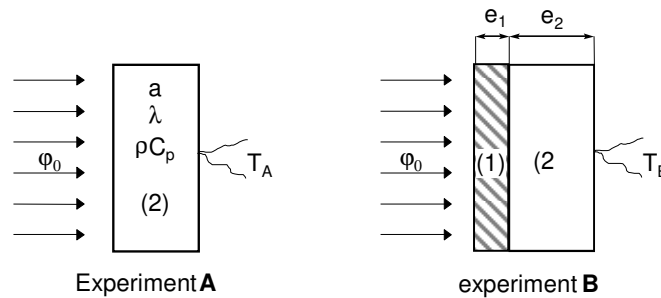


Figure 1 : Basis of the “thermal contrast” method

The thermal contrast method requires the repetition of two “flash” experiments A and B (**Figure 1**). The first one is operated on the substrate only (index (2)) whose thermophysical properties are known. The second experiment is performed on the two-layered sample (index (1)/(2)). In both cases, one records the rear face temperature evolutions. The thermograms so obtained are normalized with respect to their maximum and the difference of the scaled thermograms T_A and T_B is computed to produce the thermal contrast thermogram. This latter is a function of the thermophysical properties of the coating (1) and of the substrate (2) through two parameters:

$$K_1 = \frac{e_1}{e_2} \sqrt{\frac{a_2}{a_1}} \quad \text{and} \quad K_2 = \sqrt{\frac{\lambda_1 \rho_1 c_1}{\lambda_2 \rho_2 c_2}} \quad (7)$$

The observable (contrast curve) and the reduced sensitivity coefficients to K_1 and K_2 are plotted in **Figure 2**. They show (i) that the sensitivities have the same order of magnitude as the signal (a good thing) but unfortunately (ii) these sensitivities appear to be totally correlated, since their maxima occur at roughly the same time (a bad thing). In this case, this simple plot shows that sensitivities to K_1 and K_2 are likely proportional and therefore that the identifiability of both parameters is impossible. This example will be more thoroughly modelled and studied in section 4 of this lecture.

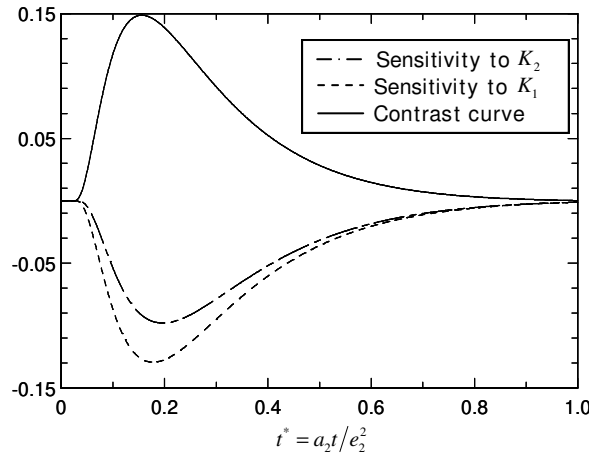


Figure 2 : Reduced sensitivity coefficients for $K_1 = 0.1$ and $K_2 = 1.36$

3.2. Variance/Correlation matrix

To go further and to investigate more deeply the PEP, the statistics of the estimator must be analysed. This can be made when (i) an estimator has been chosen (that is, a method to derive estimated values for the different parameters from the experimental signal), and (ii) the statistical properties of noise ε are known (according to experimentally founded observations).

We assume that the noise on the experimental signal is additive (this is in fact the definition of a noise), unbiased (which means that its stochastic average, its expectation is zero, for an unbiased model structure η of course) and independent (which means that the noise taken at two different times are independent) and has a constant variance σ^2 : this is sometimes called a IID. (Independent and Identically Distributed) noise, which occurs for perfect measurement with an ideal sensor. This corresponds to

$$y_i = y_{mo}(t_i; \boldsymbol{\beta}) + \varepsilon_i \quad ; \quad E(\boldsymbol{\varepsilon}) = \mathbf{0} \quad ; \quad \text{cov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_m \quad (7)$$

where \mathbf{I}_m is the identity matrix of size m (number of measurement points).

According to Beck's taxonomy (see [2] p. 134 and chapter VII), these assumptions correspond to the set "1111—11" with the following additional precisions: non stochastic independent explanatory variable (time), and no prior information for the parameters.

The OLS (Ordinary Least Squares) estimator $\hat{\boldsymbol{\beta}}_{OLS}$ minimizes the least square sum, which gives:

$$J_{OLS}(\boldsymbol{\beta}) = \mathbf{r}^T(\mathbf{t}; \boldsymbol{\beta}, \mathbf{u}) \mathbf{r}(\mathbf{t}; \boldsymbol{\beta}, \mathbf{u}) = \|\mathbf{r}(\mathbf{t}; \boldsymbol{\beta}, \mathbf{u})\|^2 = \sum_{i=1}^m (y_i - y_{mo}(t_i; \boldsymbol{\beta}, \mathbf{u}))^2 \quad (8)$$

Where:

$$\mathbf{r}(\mathbf{t}; \boldsymbol{\beta}, \mathbf{u}) = \mathbf{y} - \mathbf{y}_{mo}(\mathbf{t}; \boldsymbol{\beta}, \mathbf{u}) \quad (9)$$

are defined as the residuals.

The estimator expression is found through a minimization process, where the j^{th} equation, also called "normal equation" is:

$$\partial J_{OLS}(t, \hat{\boldsymbol{\beta}}^{OLS}) / \partial \beta_j = 0 \quad \text{for } j = 1, 2, \dots, n \quad (10a)$$

verified. If the global minimum of $J_{OLS}(\boldsymbol{\beta})$ is reached, the OLS estimator is unbiased, which means that the statistical mean of repeated estimated values $\hat{\boldsymbol{\beta}}$ is equal to the exact parameter vector $\boldsymbol{\beta}$.

Lecture 3 describes the behaviour of such an estimator for a LP model where the calculations can be fully completed to get an explicit linear OLS solution:

$$\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T \mathbf{y} \quad (11b)$$

In the case of a NL structure, the minimum is found through an iterative process using local linearity (Gauss-Newton algorithm basically, see [3]) of the form:

$$\hat{\boldsymbol{\beta}}_{OLS}^{(k+1)} = \hat{\boldsymbol{\beta}}_{OLS}^{(k)} + \left(\mathbf{S}^{(k)T} \mathbf{S}^{(k)} \right)^{-1} \mathbf{S}^{(k)T} \left(\mathbf{y} - \mathbf{y}_{mo}(\hat{\boldsymbol{\beta}}_{OLS}^{(k)}) \right) \quad (11)$$

The iterative process (12) requires computing the inverse of matrix $\mathbf{S}^T \mathbf{S}$ at each iteration k . Therefore, this latter must offer a good enough conditioning through repeated iterations. This is possible if the sensitivity coefficients are non-zero and linearly independent. Without any specialized and dedicated tool, this iterative process can be stopped when the residuals norm $\mathbf{r}^T \mathbf{r}$ is of the same order of magnitude as the measurement noise, that is when:

$$J_{OLS}(\hat{\boldsymbol{\beta}}^{(k)}) \approx m \sigma^2 \quad (12)$$

At convergence, the standard deviation of the error made for the estimated parameters can be evaluated thanks to the (symmetrical) **estimated** covariance matrix of the estimator. It characterizes the precision that can be reached on the estimated parameters (its inverse is sometimes named the precision matrix) and depends on the statistical assumptions that can be made on the data. In view of an OLS estimator, this matrix is

$$\text{cov}(\hat{\boldsymbol{\beta}}) \equiv \begin{bmatrix} \text{var}(\hat{\beta}_1) & \text{cov}(\hat{\beta}_1, \hat{\beta}_2) & \dots & \text{cov}(\hat{\beta}_1, \hat{\beta}_n) \\ \text{cov}(\hat{\beta}_1, \hat{\beta}_2) & \text{var}(\hat{\beta}_2) & & \text{cov}(\hat{\beta}_2, \hat{\beta}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(\hat{\beta}_1, \hat{\beta}_n) & \text{cov}(\hat{\beta}_2, \hat{\beta}_n) & \dots & \text{var}(\hat{\beta}_n) \end{bmatrix} \approx \sigma^2 (\mathbf{S}^T (\hat{\boldsymbol{\beta}}) \mathbf{S} (\hat{\boldsymbol{\beta}}))^{-1} \quad (13)$$

It depends on the level of the Signal-to-Noise Ratio (SNR) and brings into play the inverse of the $\mathbf{S}^T \mathbf{S}$ matrix, already pointed out as a decisive operation for an accurate estimation. Matrix $\mathbf{S}^T \mathbf{S}$, which is also called the Fisher's information matrix with assumptions (8), depends on the number m of measurement points and on their distribution along the estimation interval, which may also be optimised if necessary [2]. The diagonal coefficients are the squares of the estimated standard deviation of each parameter $\sigma_{\hat{\beta}_j}^2$. They quantify the error that one can expect through inverse estimation. This is true if the assumptions

made for the noise are consistent with the experiment. The problem being NLP, retrieving these optimum bounds through a statistical analysis may depend on the starting guesses made to initialize the estimation algorithm. This matrix can also be an indicator for detecting possible correlations between the parameters. An estimation of the correlation matrix is calculated according to:

$$\mathbf{cor}(\hat{\boldsymbol{\beta}}) \approx \begin{bmatrix} 1 & \rho_{ij} & \cdots \\ \rho_{ij} & 1 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \text{ all terms being the result of } \rho_{ij} = \frac{\text{cov}(\hat{\beta}_i, \hat{\beta}_j)}{\sqrt{\sigma_{\hat{\beta}_i}^2 \sigma_{\hat{\beta}_j}^2}} \quad (14)$$

The correlation coefficients (off-diagonal terms) correspond to a quantification of the 2 by 2 correlation existing between the two estimations of parameters β_i and β_j and, more precisely, between their errors (let us note that other forms of correlations involving more than 2 sensitivity coefficients exist, that is the multiple collinearity problem, which is detailed in section 3.3.2 further down). They vary between -1 and 1. They are global quantities (in some sense, “averaged” over the considered estimation interval, the whole $[0, t_m]$ here). Gallant [4] suggested that difficulty in computation may be encountered when the common logarithm of the ratio of the largest to smallest eigenvalues of \mathbf{cor} exceeds one-half the number of significant decimal digits used by the computer.

A more practical hybrid matrix representation \mathbf{Vcor} can be constructed. It gathers the diagonal terms of the **covariance** matrix (more precisely their square root, normalized by the value of the estimated parameter) and the off-diagonal terms of the **correlation** matrix.

$$\mathbf{Vcor}(\hat{\boldsymbol{\beta}}) \approx \begin{bmatrix} \sqrt{\text{var}(\hat{\beta}_i)} / \hat{\beta}_i & \rho_{ij} & \cdots \\ \rho_{ij} & \sqrt{\text{var}(\hat{\beta}_j)} / \hat{\beta}_j & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \quad (15)$$

TOOL Nr2: Matrix $\mathbf{Vcor}(\hat{\boldsymbol{\beta}})$ gives a quantitative point of view about the identifiability of the parameters. The main interest of this matrix lies in its diagonal coefficients, the relative standard deviation of the estimations of each parameter: these can be calculated independently from their physical units. These standard deviations of the estimated parameters are the stochastic root mean squares of the errors that are caused by the sole stochastic character of the IID noise, for an unbiased model.

The off-diagonal terms (correlation coefficients) are generally of poor interest because of their too global character. Values very close to ± 1 may explain very large variances (errors) on the parameters through a correlation effect.

NB: Another matrix, $\text{rcov}(\hat{\boldsymbol{\beta}})$ defined in equation (35) further on, is also very useful for assessing the quality of a potential inversion. Its diagonal coefficients are the squares of those of $\mathbf{Vcor}(\hat{\boldsymbol{\beta}})$, but its off-diagonal coefficients are different.

Example: Here are two Vcor matrices taken from [1]. They were obtained for the same NLPE problems and for the same given set of nominal values of the $n = 3$ parameters but considering two different observables **A** and **B** (two different locations of the temperature measurements).

$$\text{Vcor}_A(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} 0.027 & 0.994 & -0.999 \\ \square & 0.0066 & -0.989 \\ \square & \square & 0.029 \end{bmatrix} \quad \text{Observable A}$$

$$\text{Vcor}_B(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} 0.0002 & -0.38 & 0.63 \\ \square & 0.0008 & -0.93 \\ \square & \square & 0.0042 \end{bmatrix} \quad \text{Observable B}$$

In the case of observable **A**, high relative standard deviations (nearly 3%) is observed for parameters β_1 and β_3 : it can be explained by a high degree of correlation between them ($|\rho_{13} = 0.999|$). Observable **A** can clearly not be used for estimating these parameters. On the contrary, observable **B** offers good identifiability for all parameters (small relative standard deviations) and does not show any 2 by 2 correlation.

3.3. Ill-conditioned PEP and strategies for tracking true degrees of freedom

3.3.1. Pathological example of ill-conditioning resulting from correlated parameters.

The good identifiability of parameters can be related to the local convexity of the cost functional $J_{OLS}(\boldsymbol{\beta})$ in the hyper-parameter space. One obvious consequence of a correlation between parameters is that several local minima may exist and make estimation algorithms consequently fail. The discussion that follows here is taken from an example of parameter estimation in a case of coupled radiative-conductive heat transfer [5]. The thermal characterization of a semi-transparent material implies a model depending on three basic parameters at least: the thermal diffusion characteristic time $t_d = e^2 / a$, the dimensionless optical thickness τ_0 and the dimensionless Planck number N (explanations to follow in section 4.1) and so $\boldsymbol{\beta} = [t_d, \tau_0, N]^T$. The estimation of the three parameters in this NLP problem may be difficult for some range of values of parameters τ_0 and N where matrix $\text{Vcor}(\hat{\boldsymbol{\beta}})$ shows that a high degree of correlation between these two parameters exists, whereas the value of parameter t_d remains unconcerned.

A plot of the OLS criterium $J_{OLS}(\boldsymbol{\beta})$ in the 2D space (τ_0, N) for a given t_d value and a given noise σ (**Figure 3**) makes the consequence of such bad conditioning quite clear.

All level sets draw a very narrow valley oriented along a line which graphically corresponds to the relation $N \approx 2 \tau_0$. A 3D plot would show that the central line of this valley does really correspond to a descending slope and hence that no real minima can be found. The level set indicated in the figure corresponds to exactly $J_{OLS}(\boldsymbol{\beta}) = 0.07 = m \sigma^2$. Trying to make the iterative optimization algorithm works below this limit for the stopping criterion is useless. In other words, the larger the noise, the higher the stopping level-set should be.

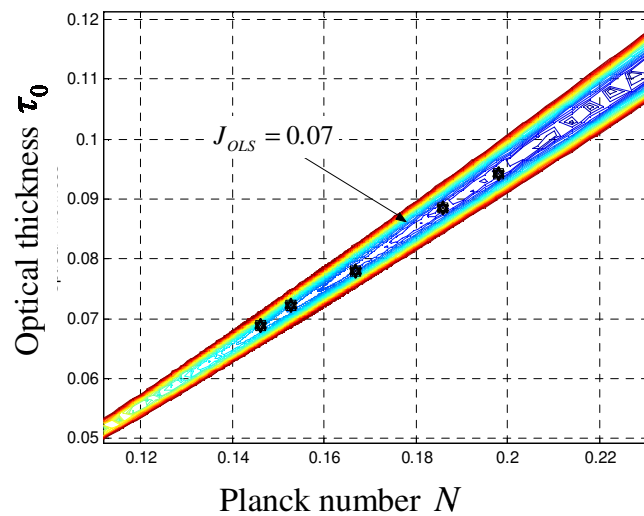


Figure 3 : Level sets for $J_{OLS}(\beta)$ in the (τ_0, N) parameter space

In the present case, this will not change the identifiability criterion. Depending on the initial guesses for the parameters, the deterministic algorithm will find different minima and different parameter estimates.

The four local minima are presented as big dots in **Figure 3** and correspond to the 3 parameters whose values are given in Table 1. Let us note that the local minimum Nr 4 Table 4) has been obtained with a stochastic algorithm (Simulated Annealing) different from a deterministic gradient based minimization algorithm used for finding the first 3 local minima. This shows that when the problem is ill-conditioned, stochastic algorithms are of little help for a correct estimation process (contrary to what is usually believed).

Such a behavior is more likely the result of a model which is not adapted to the physics involved. In the present case, it is interesting to note in Table 1 that all local minima that were found follows the relation $N(\tau_0 + 1)/\tau_0 = \text{Constant}$.

Parameter vector components	Local Minima			
	(found using either deterministic or stochastic algorithms)			
	N°1	N°2	N°3	N°4
a (10^7 m ² /s)	5.2	4.9	5.85	4.8
N	0.6	0.74	0.16	0.82
τ_0	0.38	0.5	0.07 ₆	0.56
$R_r = \frac{N_{pl}(\tau_0 + 1)}{\tau_0}$	2.18	2.22	2.26	2.28

Table 1 : Example of local minima found $\hat{\beta}$

In fact, an approximate modeling for conductive-radiative transfer in optically thin media can be shown to be more pertinent and more parsimonious. It makes naturally arise the notion of radiative resistance R_r which can be expressed as $R_r = N(\tau_0 + 1)/\tau_0$. This resistance is the appropriate parameter in this limiting behavior and prove that there is no way to identify independently τ_0 and N (Many different pairs are able to produce the same value for R_r).

TOOL Nr3: For an independent noise with known standard deviation and for a given model, it may be interesting to look at the level-set representation of the optimisation criterion in appropriate cut planes (for a given pair of parameters if $n > 3$) and compare it with the minimum achievable criterion given by $J = m\sigma^2$, where m is the number of measurements.

3.3.2. Rank of the sensitivity matrix.

We focus here on the scaled (or reduced) sensitivity matrix (see definition in equations (6a) and (6b)). This (m, n) matrix is composed of n column vectors, the reduced sensitivity coefficients \mathbf{S}_j^* :

$$\mathbf{S}^* = [\mathbf{s}_1^* \quad \mathbf{s}_2^* \quad \dots \quad \mathbf{s}_n^*] \quad \text{with} \quad \mathbf{s}_j^* = \beta_j \left. \frac{\partial \eta(\mathbf{t}; \boldsymbol{\beta}^{nom})}{\partial \beta_j} \right|_{\mathbf{t}, \beta_k \text{ for } k \neq j} \quad (16)$$

where \mathbf{t} is a column vector composed by all the m times of measurement:

$$\mathbf{t} = [t_1 \quad t_2 \quad \dots \quad t_m]^T \quad (17)$$

These n column vectors \mathbf{S}_j^* are in fact just the components of a set of n vectors $\vec{\mathbf{S}}_j^*$ in a m -dimension vector space. One can recall here that this set of vector $\Sigma = \{\vec{\mathbf{S}}_1^*, \vec{\mathbf{S}}_2^*, \dots, \vec{\mathbf{S}}_n^*\}$ is linearly independent only if m coefficients α_j exist such as:

$$\sum_{j=1}^n \alpha_j \mathbf{S}_j^* = \mathbf{0} \Rightarrow \quad \alpha_j = 0 \quad \text{for any } j \quad \text{with } 1 \leq j \leq n \quad (18)$$

This means that a linear combination of all these m vectors is equal to zero only if all its coefficients (the α_j 's here) are equal to zero. If it is not the case, system Σ is linearly dependent. Let us note that the presence of a null vector in the set of vectors Σ makes it linearly dependent: such a null vector $\vec{\mathbf{S}}_j^*$ would correspond here to a parameter that has no influence on the variation of the model output, (the very specific case of a parameter β_j rigorously equal to zero is discarded here).

So, if the set is dependent, one has to remove one vector $\vec{\mathbf{S}}_j^*$ from the original set Σ and try again to test the independence condition (19) with the $n-1$ remaining vectors. This can be

made with the n possible choices for the vector \vec{S}_j^* that is removed from set Σ . If one finds one such independent set of $n-1$ vectors, the rank of the set is $n-1$. In the opposite case, one has to test the independence with $n-2$ vectors and so on... The rank r of Σ is the larger number of vectors for an independent subset of Σ that can be formed with the n original vectors.

In order to illustrate this, we will assume that $m = n = 2$ and that the model is linear. This corresponds to two observations of a model with two parameters β_1 and β_2 . This leads to the set of two sensitivity vectors $\Sigma = \{\vec{S}_1^*, \vec{S}_2^*\}$ from which the situations shown in Figure 4 can be considered:

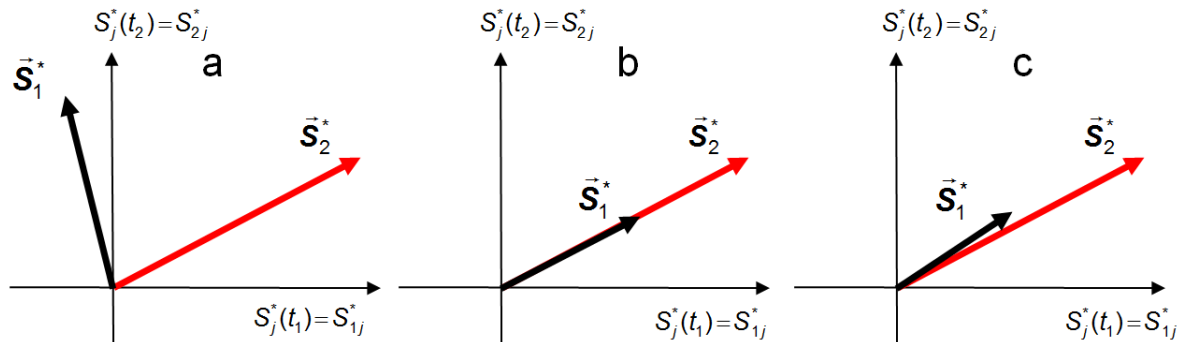


Figure 4 : *Reduced sensitivity vectors:*

a - independent sensitivities ($r = n = 2$) **b** - dependent sensitivities **c** - nearly dependent sensitivities

Case **a** corresponds to linearly independent sensitivity coefficients: the rank of Σ is equal to 2. It is also the rank of the reduced sensitivity matrix \mathbf{S}^* and hence the rank of the sensitivity matrix, since $\mathbf{S}^* = \mathbf{S} \mathbf{R}$ (where \mathbf{R} is the square diagonal matrix with two diagonal coefficients β_1 and β_2 according to equation 7). One can say that the observations of the model output provide two degrees of freedom since two parameters can be estimated.

Case **b** demonstrates a pathological nature of the sensitivity coefficients: they are proportional, with $\vec{S}_2^* = 2 \vec{S}_1^*$ (one sees that the choice $\alpha_1 = 2$ and $\alpha_2 = -1$ in (19), which allows to show that the set of vectors Σ is not independent) and estimation of both coefficients is not possible anymore. In this case, the rank of \mathbf{S}^* and hence the rank of \mathbf{S} is $r = 1$ and the determinant of the information matrix $\mathbf{S}^T \mathbf{S}$ is equal to zero. This means that the explicit value of $\hat{\beta}_{OLS}$, in the linear case (see equation 11b) and with a noise of spherical covariance matrix, which requires an inversion of the information matrix, is not possible. The same is true for the calculation of the variance-covariance matrix of $\hat{\beta}_{OLS}$: the observations of the model output provide only one degree of freedom and only one parameter can be estimated, if the value of the other one is known.

Case **c** lies in between: the two reduced sensitivity vectors are nearly proportional $\vec{S}_2^* \approx 2 \vec{S}_1^*$. Even if the mathematical rank is still equal to 2 (the previous equality is not an exact one), one guesses that the number of degrees of freedom is somewhere between one and two and a more refined statistical analysis, taking into account the noise level in the measurements, has to be implemented.

Let us note that it is possible to test the presence of two nearly proportional vectors in set Σ , in the very general case, with of course a number of parameters less or equal to the number of observations ($n \leq m$), by testing the assumption $\vec{S}_k^* - c_{kj} \vec{S}_j^* = \text{or } \approx \vec{0}$, where c_{kj} is a proportionality constant: a plot of $S_k^*(t_i)$ as a function of $S_j^*(t_i)$, for the m common values t_i of the independent variable where observations are available (parametric representation of a curve) shows whether the plots gather on the $S_k^*(t) = c_{kj} S_j^*(t)$ line or not.

As an example of this type of representation, **Figure 5** illustrates the case taken from [1] of a 1D rear face transient response of a low insulating sample (conductivity λ) sandwiched between two very thin copper layers. The knowledge model (RDM1 in [1]) assumes pure thermal resistance for the insulating layer and pure known capacities for the copper layers. The front face is stimulated by a Dirac pulse of energy Q (J.m⁻²), with a heat loss coefficient h (W.m⁻² K⁻¹) equal over its two faces: the sensitivities to the three parameters Q , λ and h seem to be qualitatively independent, but only in terms of two by two linear dependencies: this does not mean that the rank of the reduced sensitivity matrix (if only these three parameters are looked for) is equal to three, because three by three linear dependencies may be possible.

This aspect, a possible dependency between the three sensitivity coefficients, is shown in **Figure 6**, for the same experimental design: a linear combination of the form $\vec{S}_3^* - c_1 \vec{S}_1^* - c_2 \vec{S}_2^* = \text{or } \approx \vec{0}$ is looked for between the three sensitivity coefficients (for $\beta_1 = Q$, $\beta_2 = h$ and $\beta_3 = \lambda$) and a linear OLS estimation of c_1 and c_2 is made using the $S_1^*(t_i)$'s and the $S_2^*(t_i)$'s as the new independent variables and the $S_3^*(t_i)$'s as new observations. The corresponding $S_3^*(t_i)$ values are plotted as a function of the recalculated values (optimal linear combination) of the corresponding model, $\hat{c}_1 S_1^*(t) + \hat{c}_2 S_2^*(t)$: since the corresponding curve is very close to the first bisecting line, a *qualitative* 3 by 3 possible linear dependency is detected.

However, one can wonder how this dependency would impede the estimation of the three parameters: this has to be confirmed by a calculation of the covariance or Vcor matrix of the corresponding estimations, as explained in 3.2.

So, we will focus here on nonlinear parameter estimation problems where local linearization concepts as well as a Singular Value Decomposition of matrix deserve to be introduced.

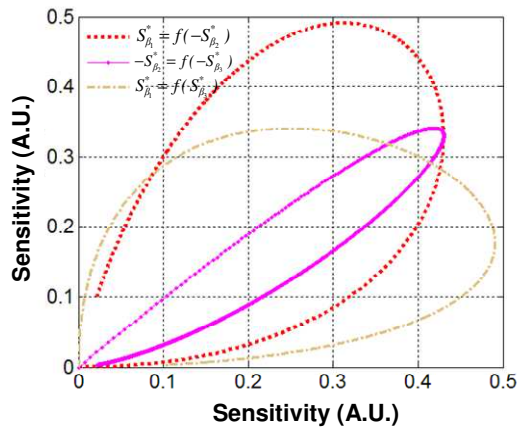


Figure 5 : Sensitivities plotted by pairs

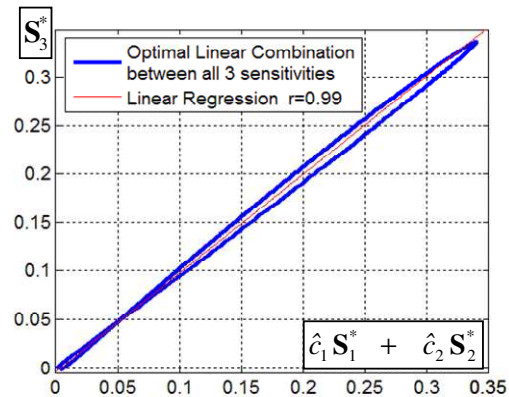


Figure 6 : Evidence of Linear Combination between all three parameters

3.3.3. Generalization: Use of SVD to track PEP degrees of freedom

It has been shown previously (see Lecture 3) that the question of identifiability of the parameters of a model relies on the condition number of the information matrix $\mathbf{S}^T \mathbf{S}$ if the physical units of the parameters are the same and of its scaled form $\mathbf{S}^{*T} \mathbf{S}^*$ if it is not the case. However, a systematic tool for tracking down hidden correlations is lacking. Such a tool will be presented now to circumvent this problem. Ultimately it will allow determining which parameters it is wise to exclude from the estimation (metrological) process, in order to get better estimates of the remaining ones.

In the next section two sequential steps will be presented.

First, in order to use all the tools available for linear estimation (see Lecture 3) on which the iterative OLS estimation (12) is based, the differential $d\mathbf{y}_{mo}$ of the model will be calculated around a reference point $\boldsymbol{\beta}^{nom}$, that is a nominal value of the parameter vector for which a sensitivity analysis has been carried out (see previous sections) and the original parameter vector $\boldsymbol{\beta}$ will be made dimensionless using the components of $\boldsymbol{\beta}^{nom}$: a reduced parameter vector \mathbf{x} with a well-defined norm will be constructed.

Second, Singular Value Decomposition (SVD) will be applied to the reduced sensitivity matrix of the "tangent" local linearized model around $\boldsymbol{\beta}^{nom}$, the ultimate goal being the determination the r parameters that can be estimated in a problem with n original parameters (with $n \geq r$), when the levels of the measurement noise and measurement magnitude are known (SNR).

The non-linear model $y_{mo}(\mathbf{t}; \boldsymbol{\beta})$ is still considered here with m available measurements.

3.3.3.1. Parameterizing a non-linear parameter estimation problem around the nominal values of its parameters

The following single-output non-linear model is considered here:

$$y_{mo} = \eta(t; \boldsymbol{\beta}) \quad (19)$$

where $\boldsymbol{\beta}$ is the column vector of the n parameters, of size $(n, 1)$, y_{mo} its (scalar) output at time t and η is a scalar function of t . If m observations of y_{mo} are available for times t_i , one can use a column vector notation:

$$\mathbf{y}_{mo} = \boldsymbol{\eta}(t; \boldsymbol{\beta}) \quad (20)$$

where \mathbf{y}_{mo} is the output vector of the model, of dimensions $(m, 1)$ and \mathbf{t} the column vector of the m times of observation. In the relation, $\boldsymbol{\eta}(\cdot)$ is a vector function whose values belong to \mathbb{R}^m .

Since the model is non-linear, it will be written under a differential form, in the neighbourhood of a reference point $\boldsymbol{\beta}^{nom}$, which corresponds to a *nominal* value, where a sensitivity study has been already implemented. This allows to use a local linearity:

$$d\mathbf{y}_{mo} = \mathbf{S}(t; \boldsymbol{\beta}^{nom}) d\boldsymbol{\beta} \quad \text{with} \quad S_{ij} = \left. \frac{\partial \eta(t_i; \boldsymbol{\beta}^{nom})}{\partial \beta_j} \right|_{t, \beta_k \text{ for } k \neq j} \quad (21)$$

Let us note that in the notation $d\mathbf{y}_{mo}$, the column vector \mathbf{t} of the measurement times has been "frozen". \mathbf{S} is the sensitivity matrix.

$$\mathbf{S} = [\mathbf{S}_1 \quad \mathbf{S}_2 \quad \dots \quad \mathbf{S}_n] \quad \text{with} \quad \mathbf{S}_j = \left. \frac{\partial \boldsymbol{\eta}(t; \boldsymbol{\beta}^{nom})}{\partial \beta_j} \right|_{t, \beta_k \text{ for } k \neq j} \quad (22)$$

In (22), the column vector $d\mathbf{y}_{mo}$ has a norm, because all its m components have the same physical units. However, such is not the case for column vector $d\boldsymbol{\beta}$, which is only a column matrix composed of n parameters whose physical dimensions are not necessarily the same: $d\beta_1$ is a very small variation in the neighbourhood of β_1^{nom} , which can be a thermal conductivity λ . $d\beta_2$ a very small variation around β_2^{nom} , which can be a volumetric heat capacity ρc and so on ...

So $d\boldsymbol{\beta}$ is not really a vector belonging to any vector space of dimension n , but a simple collection of n parameters.

In order to transform it into a real vector, a normalization of all its elements is necessary. The components of $\boldsymbol{\beta}^{nom}$ will be used for that purpose. A new dimensionless parameter \mathbf{x} is introduced.

Its components are defined by:

$$x_j = \ln(\beta_j / \beta_j^{nom}) \quad (23)$$

And its nominal value is equal to zero:

$$\mathbf{x}^{nom} = \mathbf{0} = [0 \ 0 \ \dots \ 0]^T \quad (24)$$

In the neighbourhood of β^{nom} , each component of \mathbf{x} is equal to the relative variation of the corresponding component of β around its nominal value (first order series expansion):

$$x_j = \ln(\beta_j / \beta_j^{nom}) = \ln\left(1 + \frac{\beta_j - \beta_j^{nom}}{\beta_j^{nom}}\right) \approx \frac{\beta_j - \beta_j^{nom}}{\beta_j^{nom}} \quad (25)$$

The new parameter vector \mathbf{x} is written the following way:

$$\mathbf{x} = \ln(\mathbf{R}_{nom}^{-1} \beta) \approx \mathbf{R}_{nom}^{-1} (\beta - \beta^{nom}) \quad (26)$$

with :

$$\mathbf{R}_{nom} = \begin{bmatrix} \beta_1^{nom} & 0 & \dots & 0 \\ 0 & \beta_2^{nom} & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \beta_n^{nom} \end{bmatrix} \quad (27)$$

With this definition, the differential $d\mathbf{x}$ of \mathbf{x} is the logarithmic differential of β :

$$d\mathbf{x} = [dx_1 \ dx_2 \ \dots \ dx_n]^T \quad \text{with} \quad dx_j = \frac{d\beta_j}{\beta_j^{nom}} \approx \frac{d\beta_j}{\beta_j} = d\ln(\beta_j) \quad (28)$$

Let us note that the very last equality is only valid in the neighbourhood of β^{nom} . It can also be written with a column vector notation:

$$d\mathbf{x} = \mathbf{R}_{nom}^{-1} d\beta \approx \mathbf{R}^{-1} d\beta \quad (29)$$

where \mathbf{R} is the square diagonal matrix whose diagonal is composed of the components of β , in the same way as (28) for the definition of \mathbf{R}_{nom} starting from β^{nom} .

Equation (22) is rewritten in order to make $d\mathbf{x}$ appear:

$$d\mathbf{y}_{mo} = \mathbf{S}^* d\mathbf{x} \quad \text{with} \quad \mathbf{S}^* = \mathbf{S} \mathbf{R}_{nom} \quad (30 \text{ a-b})$$

\mathbf{S}^* is the reduced sensitivity matrix calculated for $\boldsymbol{\beta}^{nom}$, see (17, 23).

So, $\mathbf{d}\mathbf{y}_{mo}$ is a column vector belonging to R^m (it can be made truly dimensionless by a division by $\|\boldsymbol{\eta}(\mathbf{t}; \boldsymbol{\beta}^{nom})\|$ but it is not necessary here) and $\mathbf{d}\mathbf{x}$ is a true column vector belonging to R^m because its norm can be defined.

Using this change of variable as well as the SVD decomposition (see *Appendix 1*) of the scaled (also called reduced) sensitivity matrix \mathbf{S}^* , one can show that equation (31a) can be used to get a first order development in the neighbourhood of $\boldsymbol{\beta}^{nom}$ (see *Appendix 2* for the demonstration):

$$\boldsymbol{\beta} \approx \mathbf{R}_{nom} \left[\mathbf{1} + \mathbf{V} \mathbf{W}^{-1} \mathbf{U}^T (\mathbf{y}_{mo} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{nom})) \right] = \boldsymbol{\beta}^{nom} + \mathbf{R}_{nom} \mathbf{V} \mathbf{W}^{-1} \mathbf{U}^T (\mathbf{y}_{mo} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{nom})) \quad (31a)$$

with the following SVD decomposition: $\mathbf{S}^*(\boldsymbol{\beta}^{nom}) = \mathbf{U} \mathbf{W} \mathbf{V}^T$ (32b)

Equation (12), that gives the Gauss-Newton algorithm can also be recast in terms of the scaled parameter \mathbf{x} :

$$\hat{\mathbf{x}} = \left(\mathbf{S}^{*T}(\boldsymbol{\beta}^{nom}) \mathbf{S}^*(\boldsymbol{\beta}^{nom}) \right)^{-1} \mathbf{S}^{*T}(\boldsymbol{\beta}^{nom}) (\mathbf{y} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{nom})) \quad (33)$$

This expression is equivalent to equation (12) where one has replaced the left-hand side $\boldsymbol{\beta}$ by its estimated value $\hat{\boldsymbol{\beta}}$ for a single iteration number k for $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{(k)}$ and $\boldsymbol{\beta}^{nom} = \hat{\boldsymbol{\beta}}^{(k-1)}$. The complete demonstration is given in *Appendix 3*.

In a similar way, the variance-covariance matrix of scaled vector $\hat{\mathbf{x}}$ can be derived from (33) and (32b), see *Appendix 4*:

$$\text{cov}(\hat{\mathbf{x}}) = \mathbf{R}_{nom}^{-1} \text{cov}(\hat{\boldsymbol{\beta}}) \left(\mathbf{R}_{nom}^{-1} \right)^T = \sigma^2 \mathbf{V} \mathbf{W}^{-2} \mathbf{V}^T \quad (34)$$

One can note that, by definition, matrix $\text{cov}(\hat{\mathbf{x}})$ is the reduced (or scaled) covariance matrix of $\hat{\boldsymbol{\beta}}$, which can be called $\text{rcov}(\hat{\boldsymbol{\beta}})$:

$\text{rcov}(\hat{\boldsymbol{\beta}}) \equiv \text{cov}(\hat{\mathbf{x}}) \equiv \begin{bmatrix} \sigma_{\hat{\beta}_1}^2 / (\beta_1^{nom})^2 & \text{cov}(\hat{\beta}_1, \hat{\beta}_2) / (\beta_1^{nom} \beta_2^{nom}) & \text{cov}(\hat{\beta}_1, \hat{\beta}_n) / (\beta_1^{nom} \beta_n^{nom}) \\ & \sigma_{\hat{\beta}_2}^2 / (\beta_2^{nom})^2 & \vdots \\ & & \ddots \\ \text{Symmetric} & & & \sigma_{\hat{\beta}_n}^2 / (\beta_n^{nom})^2 \end{bmatrix} = \sigma^2 (\mathbf{S}^{*T} \mathbf{S}^*)^{-1}$	(35)
--	------

One also shows, in Appendix 4, that the trace of $\text{cov}(\hat{\mathbf{x}})$, that is the sum of the square of the relative standard deviations of all the estimations $\hat{\beta}_j$, at convergence, is equal to the sum of the square of the inverses of the singular values of \mathbf{S}^* , with a multiplicative factor equal to the variance σ^2 of the IID noise:

$$\text{Tr}(\text{cov}(\hat{\mathbf{x}})) \equiv \sum_{j=1}^n (\sigma_{\beta_j} / \beta_j^{\text{nom}})^2 = \sigma^2 \sum_{k=1}^n \frac{1}{w_k^2} \quad (36)$$

This allows to define a criterion m_q that assesses the global precision of the estimation:

$$m_q \equiv \left(\frac{1}{n} \sum_{j=1}^n (\sigma_{\beta_j} / \beta_j^{\text{nom}})^2 \right)^{1/2} = \frac{1}{\sqrt{n}} \text{Tr}(\text{cov}(\hat{\mathbf{x}})) = \sigma \left(\frac{1}{n} \sum_{k=1}^n \frac{1}{w_k^2} \right)^{1/2} \quad (37)$$

m_q is the root mean square relative standard deviation of the different parameters. So, it can be expressed in percent. If a specific parameter is estimated with a high relative variation, this will have an effect of m_q that will get large. The advantage of this criterion is that it takes into account the level of the measurement noise, contrary to the condition number of the relative sensitivity matrix $\text{cond}(\mathbf{S}^*) \equiv w_1/w_n$ (see Lecture 3). It is quite easy to find an upper and a lower bound for it:

$$\frac{1}{\sqrt{n}} \frac{\sigma}{w_n} \leq m_q = \left(\frac{1}{n} \sum_{j=1}^n (\sigma_{\beta_j} / \beta_j^{\text{nom}})^2 \right)^{1/2} \leq \frac{\sigma}{w_n} \quad (38)$$

Other points about this criterion that allows to study the well-posedness of a non-linear parameter estimation problem are given in *Appendix 4*.

TOOL Nr4: The SVD of the normalized sensitivity matrix calculated for nominal values of parameter vector β can bring valuable information to quantify the real identifiability of the parameters, once the level of noise known.

3.3.4 *Residuals analysis and signature of the presence of a bias in the metrological process*

One way to analyse the results of an estimation process is to calculate the residuals (equation 10) at convergence, when the assumptions (8) are fulfilled (an IID noise). When the model used for the estimation is not biased, the calculation of the residual column vector $\mathbf{r}(\hat{\beta})$ whose k^{th} coefficients is the residual $r(t_k; \hat{\beta})$ at time t_k is:

$$\mathbf{r}(\hat{\boldsymbol{\beta}}) \equiv \mathbf{y} - \mathbf{y}_{mo}(\hat{\boldsymbol{\beta}}) = \mathbf{y}_{mo}(\boldsymbol{\beta}^{exact}) + \boldsymbol{\varepsilon} - \mathbf{y}_{mo}(\hat{\boldsymbol{\beta}}) \approx \boldsymbol{\varepsilon} - \mathbf{S}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{exact}) \text{ with } \mathbf{S} = \mathbf{S}(\hat{\boldsymbol{\beta}}) \quad (39)$$

One shows in *Appendix 5* that, strictly speaking, the residuals, when the model is unbiased, are correlated but, in practice, adding more measurements times for a given estimation interval tends to make them nearly uncorrelated. This is especially true for thermal characterization of materials or systems, where the number of parameters is low (2, 3, 4, ...) and the time sampling rate high enough with respect of the length of measurement (several hundredth of measurements at least for modern data acquisition systems).

So, when these previous conditions are fulfilled, "signed" residuals can be considered as the signature of some estimation based on a biased model.

This bias can stem from different causes such as:

- (i) the a priori decision that some parameters of the model are known and therefore fixed at some given value (maybe measured by another experiment). As active parameters in the PEP, they can alter the estimates of the remaining unknown parameters.
- (ii) Experimental imperfections which make the model idealized with respect to the reality of the phenomena.

The existence of a bias means that a systematic and generally unknown inconsistency exists between the model and the experimental data.

We give here an example taken from [1] and already studied in section 3.3.2 above. It concerns the simulation of a flash experiment applied to a three-layer medium: two highly capacitive and conductive coatings and a central layer made of a material with very poor conductivity (highly insulating material) and heat capacity (aerogel material). This system can be modelled through some function $T^{rear} = y_{mo}(t, \boldsymbol{\beta})$. An artificial bias $d(t)$ is introduced under the form of a linear drift superimposed to the output simulated observations. It corresponds practically to a linear deviation of the signal from the equilibrium situation before the experiment starts. So, the correct model that should be used to mimic the observed rear face measurement should be:

$$y_{mo}^{drift}(t_k, \boldsymbol{\beta}^{exact}) = y_{mo}(t_k; \boldsymbol{\beta}^{exact}) + d(t_k) \quad (40)$$

A noise respecting equation (8) is also added to the simulation of the measurements so that we have at each time t_k :

$$y_k = y_{mo}^{drift}(t_k, \boldsymbol{\beta}^{exact}) + \varepsilon_k \quad (41)$$

Of course, model $y_{mo}(t, \boldsymbol{\beta})$ is exact if no drift is present in the experiment. However, in the opposite case, it becomes biased, since it does not account for the presence of this drift.

Let us note that in this definition, the drift model is the reference one ($y_{mo}^{exact} = y_{mo}^{drift}$) and the preceding thermal model is the biased one ($\mathbf{y}_{mo}^{biased} = \mathbf{y}_{mo}$).

If this biased model is used for estimation, the minimization will be done by a minimization of the following criterion based on a biased residual vector:

$$J_{biased}(\boldsymbol{\beta}) = \mathbf{r}_{biased}^T(\boldsymbol{\beta}) \mathbf{r}_{biased}(\boldsymbol{\beta}) \quad \text{with} \quad \mathbf{r}_{biased}(\boldsymbol{\beta}) \equiv \mathbf{y} - \mathbf{y}_{mo}(\boldsymbol{\beta}) = \mathbf{y} - \mathbf{y}_{mo}^{drift}(\boldsymbol{\beta}) + \mathbf{d} \quad (42)$$

As a consequence, at convergence, the error on the estimated parameters vector will have a deterministic part and a stochastic part:

$$\mathbf{e}_{\beta} \equiv \hat{\boldsymbol{\beta}}^{biased} - \boldsymbol{\beta}^{exact} \approx \mathbf{b}_{\beta} + \mathbf{A}\boldsymbol{\varepsilon} \quad \text{with} \quad \mathbf{b}_{\beta} = \mathbf{0} \quad \text{if} \quad \mathbf{d} = \mathbf{0} \quad (43)$$

where \mathbf{A} is a matrix that corresponds to the linearization of the inverse problem with respect to the noise in the neighbourhood of the exact value of $\boldsymbol{\beta}^{exact}$ and \mathbf{b}_{β} a bias of non-zero average, that stems from the presence of the drift \mathbf{d} .

As a consequence, the residual defined in (42) can be calculated, at convergence, using (43):

$$\mathbf{r}_{biased}(\hat{\boldsymbol{\beta}}^{biased}) \equiv \mathbf{y} - \mathbf{y}_{mo}(\boldsymbol{\beta}) = \mathbf{y}_{mo}^{drift}(\boldsymbol{\beta}^{exact}) + \boldsymbol{\varepsilon} - \mathbf{y}_{mo}(\hat{\boldsymbol{\beta}}^{biased}) \quad (44)$$

or

$$\mathbf{r}_{biased} = \mathbf{y}_{mo}(\boldsymbol{\beta}^{exact}) + \mathbf{d} + \boldsymbol{\varepsilon} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{exact} + \mathbf{b}_{\beta} + \mathbf{A}\boldsymbol{\varepsilon}) \quad (45)$$

A first order development of the last term around the exact value $\boldsymbol{\beta}^{exact}$ yields:

$$\mathbf{r}_{biased}(\hat{\boldsymbol{\beta}}^{biased}) = \mathbf{y}_{mo}(\boldsymbol{\beta}^{exact}) + \mathbf{d} + \boldsymbol{\varepsilon} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{exact}) - \mathbf{S}(\boldsymbol{\beta}^{exact}) [\mathbf{b}_{\beta} + \mathbf{A}\boldsymbol{\varepsilon}] \quad (46)$$

or

$$\mathbf{r}_{biased}(\hat{\boldsymbol{\beta}}^{biased}) = \mathbf{d} + \mathbf{S}(\boldsymbol{\beta}^{exact}) \mathbf{b}_{\beta} + [\mathbf{I} - \mathbf{S}(\boldsymbol{\beta}^{exact}) \mathbf{A}] \boldsymbol{\varepsilon} \quad (47)$$

This means that the residuals are biased, because of their first deterministic component, even if its second stochastic one may be diagonal.

We return here to the estimation problem described in section 3.3.2 (flash experiment on a three layer sample for the inner insulating layer characterization): we have seen that the model used for parameter estimation was ill-conditioned: some correlation exists between the parameters (Case $n = 3$ corresponding to the correlation existing between parameters shown in **Figure 5** and **Figure 6**). **Figure 7** below shows that

- the simulated rear face noisy output of the system, with the drift and some added noise (dotted curve),
- the corresponding rear face recalculated output using the biased estimate $\hat{\boldsymbol{\beta}}$ (obtained through minimization of criterion (42)) - (blue solid line),

- the drift of the model output (function $b_y(t)$) introduced (brown solid line) . At the final time of the experiment ($t_f = 1000$ s), the magnitude of the drift represents less than 4% of the maximum level of the signal,
- the residuals curve, with the noised signal (minimization of criterion (42), grey stochastic line), and after removing the noise, that is with the same estimation process starting from a noiseless signal, that is with $\boldsymbol{\varepsilon} = \mathbf{0}$, blue solid line).

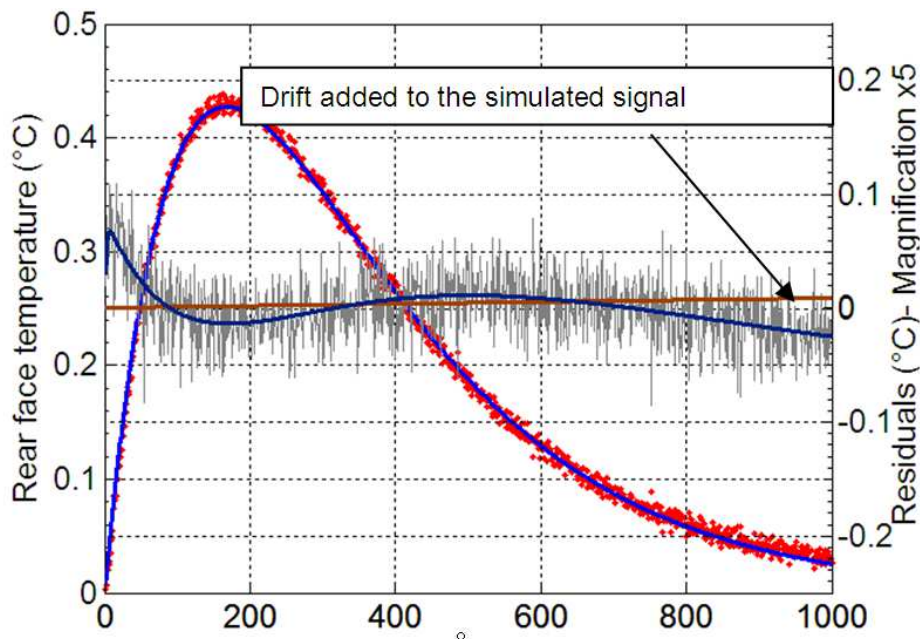


Figure 7 : Signed character of "post-estimation" residuals in the presence of a bias and using a badly conditioned PEP

The "signed" character of the residuals is obvious (oscillation around zero with a much smaller frequency than the noise). The three parameters estimated (Q , h and λ) using these biased "measurements" have averaged values (obtained by repeated Monte Carlo simulated measurements) that differ respectively by -18%, -7.5%, +19% from the exact input values. These differences are not of stochastic origin (caused by noise only) but result from the introduction of the bias.

One possibility for the experimenter who wants to check whether his estimations are biased or not, is to observe the output of the inversion process for varying identification ranges of the independent variable. For example, we can vary the identification time interval. If a bias affects the data when compared to the modeling, then the estimations will vary, depending on the selected identification interval. This is what can be observed in **Table 2** where three identifications have been performed for three different time intervals [0-70s], [0-150s], [0-300s]. In this case we have used a more refined model than the one used for **Figure 7** and thus a more badly conditioned PEP. In this table both thermal properties of the insulating material (thermal conductivity and thermal diffusivity) were estimated from the biased data.

Obviously with such a material, the small heat capacity makes a good estimation of this parameter difficult, but sadly (because of a lack of sensitivity) this also affects the estimation

of the second parameter. The thermal diffusivity and conductivity estimated from the data of **Figure 7** depend strongly on the identification intervals. The values can change within a factor of 60% or 170% in that case.

Time Interval	70 s	150 s	300 s
a (m ² /s)	$3.76 \cdot 10^{-6}$	$3.22 \cdot 10^{-6}$	$2.21 \cdot 10^{-6}$
λ (W/m/°C)	0.031	0.064	0.084

Table 2 : Influence of the existence of some bias on the parameter estimates for a badly conditioned problem

TOOL Nr5: The "post-estimation" residuals have to be analysed carefully to check the potential existence of a bias of systematic origin. Its magnitude can be compared to the standard deviation of the white noise of the sensor in order to check whether this bias may introduce too large confidence intervals for the estimates (with respect to the pure stochastic estimation of the variances of parameter estimates in the absence of any bias). Invariant estimates for different identification intervals suggest that the bias is acceptable. In the opposite case, strategies must be implemented, either to change the nature of the estimation problems (reduction of the initial goals) or to use residuals to give a fair quantitative evaluation of the confidence bounds of the estimates. Some hints on that topic will be given in the next section.

4. Enhancing the performances of estimation

Some tools have been given above: they can help the experimenter to gain insight into its metrological problem. They can lead to a conclusion of failure: the problem is ill-conditioned regarding the estimation of the interesting parameters. This means that the parameters we initially wish to measure will actually never be estimated accurately. Two strategies are possible: recognizing that the initial goal is in vain or modifying the problem through physical thinking to make it well-posed or adequately conditioned even by changing the goals themselves (number of parameters to estimate). Quoting J.V.Beck [2]: "the problem of non-identifiability can be avoided, through either the use of a different experiment or a smaller set of parameters that are identifiable".

This position emerges from the well-known parsimony "principle" (see <http://en.wikipedia.org/wiki/Parsimony>) which in the field of science could be summarized by this sentence : "trying to perfectly recover reality is indeed very easy, when one adds parameters to each other so that it connects-the-dots". There is much more to learn and to retrieve from the distance maintained between a model and the observations it is supposed to match. The resulting consequence is that any minimization algorithm is a good one because the problem is well defined. This section will now proceed to give additional tools to work out badly conditioned problems with special analysis regarding the role of known versus unknown parameters.

4.1 Dimensional analysis or natural parameters: case of coupled conduction/radiation flash experiment

Through the preceding sections, the reader should have been convinced of the importance of notions like the pertinence of a model (good representation of reality,

controlled origins of bias), the application of the parsimony principle that is to adapt one's metrological objective by making the "quality" of the available information match the degree of complexity of the model.

A reduced model, seen as a model with a reduced number of parameters, has to be considered first in the light of Dimensional Analysis. The principles of Dimensional Analysis in Engineering precisely rely on the construction of "appropriate" natural parameters (the Pi-groups) emerging from the rank determination of the dimensional matrix of all physical quantities involved in the problem with respect to a basis of "base" quantities [6].

If we consider the heat transfer problem in a semi-transparent material like glass, coupled conduction and radiation transfers must be considered. Material parameters involve classical thermophysical properties of the opaque material (thermal conductivity λ , specific heat ρc) with the additional parameters accounting for radiative transfer : the absorption (extinction coefficient) β (m^{-1}), the level of temperature of the material T_0 (in Kelvin) which rules the magnitude of radiation emission, the Stefan-Boltzmann constant σ_{SB} , the refractive index n , and the inner emissivity ε_i of the boundaries (no units - opaque coatings of the glass slab are considered here).

Let us assume that a flash experiment is planned, with an absorbed heat density Q ($J.m^{-2}$). In order to study the possibilities for a transient thermal characterization technique of such materials (which parameters can be measured with this experiment?), the model will give the rear face temperature response of the slab (thickness e) as the following function:

$$y_{mo} = T_{rear}^{flash}(t, e, Q, \rho c, \lambda, \beta, \sigma_{SB}, T_0, \varepsilon_i, n) \quad (48)$$

Practicing a "blind" Dimensional Analysis leads to the construction of a new function depending on a new set of parameters:

$$y_{mo} \equiv \frac{T_{rear}^{flash} - T_0}{T_0} = T_{rear}^{flash*} \left(t^* = \frac{at}{e^2}, \tau_0 = \beta e, N = \frac{\lambda \beta}{n^2 \sigma_{SB} T_0^3}, T_0^* = \frac{Q}{\rho c e}, \varepsilon_i \right) \quad (49)$$

which naturally produces 4 pi-groups governing heat transfer inside the sample, with a reduction of the number of initial parameters of the model from 10 to 5.

Another classical example deals with conductive and convective mechanisms of transfer which appear jointly in problems of heat transfer within boundary layers. Solving the Inverse Heat Conduction Problem in order to get a heat exchange coefficient estimation will require the introduction of the classical Reynolds, Nusselt and Prandtl numbers.

4.2 *Reducing the PEP to make it well-conditioned: case of thermal characterization of a deposit*

➤ **Model:** *Case of the contrast method*

The method of the thermal contrast already presented in **Section 3.1** consists in making two "flash" experiments in order to estimate the thermal properties of the coating layer, denoted (1) in **Figure 8** below (the same as **Figure 1**). We will now on detail the modelling already presented briefly in section 3.1, in order to be able to find out which parameters of the model can be really estimated, in this non-linear parameter estimation problem.

Let us remind that the first flash experiment is carried out on the substrate denoted (2), which allows characterization of the substrate in terms of diffusivity (the thermal capacity of the substrate is measured by another facility). The second flash experiment is performed on the two-layer material denoted (1)/(2).

In both cases, the variation of the rear-face temperature T with time, called thermogram, is measured. By taking the difference of these thermograms T_A^* and T_B^* normalized by their respective maximum, we obtain a curve called a thermal contrast curve, which is a function of the thermophysical parameters of the film (1) and of the substrate (2).

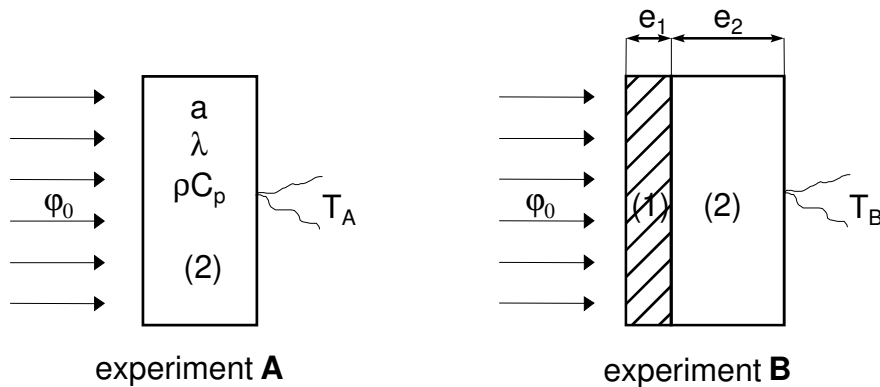


Figure 8 : *Principle of the Method*

The thermal quadrupoles method [7] is very appropriate to find the rear-face temperatures. Taking the Laplace transform of the heat equation yields a linear relationship between the different quantities of the "in" and "out" faces of each layer of the material.

Let $\theta(z, p)$ and $\phi(z, p)$ being the Laplace transforms of the temperature $T(z, t)$ and heat density $\phi(z, t)$ respectively, with z the axis normal to both faces:

$$\theta(z, p) = \mathbf{L} [T(z, t)] = \int_0^{\infty} T(z, t) \exp(-pt) dt \quad (50)$$

and

$$\phi(z, p) = \mathbf{L} [\phi(z, t)] = \int_0^{\infty} \phi(z, t) \exp(-pt) dt \quad \text{with} \quad \phi(z, t) = -\lambda \frac{\partial T}{\partial z} \quad (51)$$

The thermal quadrupoles method allows to linearly link the temperatures and the heat flux densities of a homogeneous layer (numbered i here) without any source term and with zero initial temperature, through a transfer matrix M_i , defined in the following way:

$$\begin{bmatrix} \theta_{i,in} \\ \phi_{i,in} \end{bmatrix} = \begin{bmatrix} A_i & B_i \\ C_i & D_i \end{bmatrix} \begin{bmatrix} \theta_{i,out} \\ \phi_{i,out} \end{bmatrix} \quad (52)$$

with the coefficients of the matrix being calculated as:

$$A_i = D_i = \cosh(e_i \sqrt{\rho/a_i}); B_i = \frac{1}{\lambda_i \sqrt{\rho/a_i}} \sinh(e_i \sqrt{\rho/a_i}); C_i = \lambda_i \sqrt{\rho/a_i} \sinh(e_i \sqrt{\rho/a_i})$$

The subscript (*i*) is related to the layer (*i*) : film (1) and substrate (2).

- e_i : thickness of the material
- a_i : thermal diffusivity
- λ_i : thermal conductivity
- ρC_{p_i} : specific heat

It is convenient in this 1D transient problem, to notice that time can be made dimensionless with the thermal diffusivity a_2 of the substrate and with its thickness e_2 , to make a Fourier number t^* appear, which will be associated to a reduced Laplace parameter p^* defined as:

$$t^* = \frac{a_2 t}{e_2^2}, p^* = p \frac{e_2^2}{a_2} \text{ and } s = \sqrt{p^*} \quad (53)$$

We can then define a reduced Laplace transform $\tilde{\theta}$ as:

$$\tilde{\theta}(z, p^*) = \tilde{\mathcal{L}} [T(z, t^*)] = \int_0^{\infty} T(z, t^*) \exp(-p^* t^*) dt^* = \frac{a_2}{e_2^2} \theta(z, p) \quad (54)$$

➤ **Flash Experiment on the substrate:**

The expression of the rear face response to a pulsed (Dirac) stimulation $\varphi(t) = Q_2 \delta(t)$, where Q_2 is the energy density (in J.m⁻²) absorbed by the front face, is given by the following relationship:

$$\begin{bmatrix} \theta_{2,in} \\ \phi_{2,in} = Q_2 \end{bmatrix} = \begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix} \begin{bmatrix} \theta_{2,out} \\ \phi_{2,out} = 0 \end{bmatrix} \quad (55)$$

$$\text{Hence: } \theta_{2out} = \frac{Q_2}{C_2} = \frac{Q_2}{\lambda_2 \sqrt{\frac{\rho}{a_2}} \sinh\left(\sqrt{\frac{\rho e_2^2}{a_2}}\right)} \quad (56)$$

Here subscript 'in' designates the front (stimulated) face while subscript 'out' is associated to the rear face, where temperature can be measured. This rear face is supposed to be insulated here ($\phi_{2out} = 0$ in (55)).

Setting $s = \sqrt{\rho^*}$ and normalizing the thermogram with respect to its maximum that corresponds to the adiabatic temperature: $T_{2\infty} = \frac{Q_2}{\rho_2 c_2 e_2}$ reached for long times for this adiabatic model, we obtain:

$$\theta_{2out}^* = \mathbf{L} \left(\frac{T_2}{T_{2\infty}} \right) = \frac{e_2^2}{a_2} \frac{1}{s \sinh(s)} \quad (57)$$

Using the reduced Laplace transform (57), we can write:

$$\tilde{\theta}_{2out}^* = \tilde{\mathbf{L}} \left(\frac{T_2}{T_{2\infty}} \right) = \frac{1}{s \sinh(s)} \quad (58)$$

➤ **Flash Experiment on the two-layer material:**

The expression of the rear face response of the two-layer material can also be obtained easily through the quadrupoles method:

$$\begin{bmatrix} \theta_{1/2in} \\ \phi_{1/2in} = Q_{1/2} \end{bmatrix} = \begin{bmatrix} A_{eq} & B_{eq} \\ C_{eq} & D_{eq} \end{bmatrix} \begin{bmatrix} \theta_{1/2out} \\ \phi_{1/2out} = 0 \end{bmatrix} \quad (59)$$

$$\text{where: } \begin{bmatrix} A_{eq} & B_{eq} \\ C_{eq} & D_{eq} \end{bmatrix} = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix} = \begin{bmatrix} A_1 A_2 + B_1 C_2 & A_1 B_2 + A_2 B_1 \\ A_1 C_2 + A_2 C_1 & A_1 D_2 + B_2 C_1 \end{bmatrix} \quad (60)$$

and where $Q_{1/2}$ is the energy density absorbed by the front face in this second flash experiment on the two-layer sample.

In the case of good conductive materials with small thicknesses, the Biot number which represents the ratio between the internal resistance and the external resistance is low, which justifies neglecting the heat losses in the model output (rear face temperature) above. The expression of the temperature takes the following form:

$$\theta_{1/2} = \frac{Q_{1/2}}{C_{eq}} = \frac{Q_{1/2}}{A_1 C_2 + A_2 C_1} \quad (61)$$

Note: If we switch the two layers of the material, it means inverting subscripts 1 and 2, and the expression of the rear-face temperature can be proved to remain unchanged.

$$\theta_{1/2out} = \frac{Q_{1/2}}{\lambda_1 \sqrt{\frac{p}{a_1}} \sinh\left(\sqrt{\frac{pe_1^2}{a_1}}\right) \cosh\left(\sqrt{\frac{pe_2^2}{a_2}}\right) + \lambda_2 \sqrt{\frac{p}{a_2}} \sinh\left(\sqrt{\frac{pe_2^2}{a_2}}\right) \cosh\left(\sqrt{\frac{pe_1^2}{a_1}}\right)} \quad (62)$$

If we now scale the thermogram with the adiabatic temperature of the two-layer material, that is with $T_{1/2\infty} = \frac{Q_{1/2}}{\rho_1 c_1 e_1 + \rho_2 c_2 e_2}$, the expression of the Laplace transforms of this reduced temperature $T_{1/2} / T_{1/2\infty}$ takes a simpler form:

$$\theta_{1/2out}^* = \frac{e_2^2}{a_2} \frac{1 + \frac{\rho_1 c_1 e_1}{\rho_2 c_2 e_2}}{s \left[\sqrt{\frac{\lambda_1 \rho_1 c_1}{\lambda_2 \rho_2 c_2}} \sinh\left(\frac{e_1}{e_2} \sqrt{\frac{a_2}{a_1}} s\right) \cosh(s) + \sinh(s) \cosh\left(\frac{e_1}{e_2} \sqrt{\frac{a_2}{a_1}} s\right) \right]} \quad (63)$$

As in section 3.1 two reduced parameters are introduced:

$$K_1 = \frac{e_1}{e_2} \sqrt{\frac{a_2}{a_1}} : \text{ratio of the root of characteristic times}$$

$$\text{or } K_1 = \sqrt{tc_1 / tc_2} \text{ with } tc_i = e_i^2 / a_i \quad \text{for } i=1, 2 \quad (64)$$

$$K_2 = \sqrt{\frac{\lambda_1 \rho_1 c_1}{\lambda_2 \rho_2 c_2}} : \text{ratio of the thermal effusivities}$$

$$\text{or } K_2 = \sqrt{b_1 / b_2} \text{ with } b_i = \sqrt{\lambda_i \rho c_i} \quad \text{for } i=1, 2 \quad (65)$$

We can note that K_1 is a function of the thicknesses of the substrate and coating and K_2 is an intrinsic parameter of the materials. The reduced Laplace transform of the response of the two-layer system can then be written, using (54):

$$\tilde{\theta}_{1/2out}^* = \frac{1}{s} \left[\frac{1 + K_1 K_2}{K_2 \sinh(K_1 s) \cosh(s) + \sinh(s) \cosh(K_1 s)} \right] \quad (66)$$

The heterogeneous nature of the two-layer material system appears here through the expression of the denominator that cannot be simplified: this makes the definition of an equivalent material associated to this two-layer sample impossible.

➤ **Contrast Curve:**

The contrast curve is obtained by taking the difference between the two thermograms, that is:

$$\Delta \tilde{\theta}_{out}^* = \tilde{\theta}_{1/2out}^* - \tilde{\theta}_{2out}^* = \tilde{\mathcal{L}} (T_{1/2out}^* - T_{2out}^*) = \tilde{\mathcal{L}} (\Delta T^*) \quad (67)$$

The expression of the reduced thermal contrast in the Laplace domain is:

$$\Delta \tilde{\theta}_{out}^* = \frac{1}{s} \left[\frac{1 + K_1 K_2}{K_2 \sinh(K_1 s) \cosh(s) + \sinh(s) \cosh(K_1 s)} - \frac{1}{\sinh(s)} \right] \quad (68)$$

Theoretically, K_1 and K_2 can be measured from an experimental thermal contrast curve through an "inverse" technique. The numerical inversion of the model is implemented by De Hoog's algorithm [10] whose MATLAB version (Invlap) is given in [11].

From K_1 and K_2 (or by a parameter substitution), it is also possible to calculate the thermal capacity and conductivity of the deposit by the following relations:

$$K_3 = K_1 K_2 = \frac{\rho_1 c_1 e_1}{\rho_2 c_2 e_2} \quad \text{thermal capacities ratio}$$

$$\text{or } K_3 = C_{t1} / C_{t2} \quad \text{with } C_{ti} = \rho c_i e_i \quad \text{for } i=1, 2 \quad (69)$$

and

$$K_4 = \frac{K_1}{K_2} = \frac{e_1 \lambda_2}{e_2 \lambda_1} \quad \text{thermal resistances ratio}$$

$$\text{or } K_4 = R_1 / R_2 \quad \text{with } R_i = e_i / \lambda_i \quad \text{for } i=1, 2 \quad (70)$$

Another parametrization of the same model consists in writing expression (68) as a function of K_3 and K_4 .

The expression of the theoretical model with scaled parameters clearly shows that its output is in this case only function of two parameters. This means in particular that the thermophysical properties of the deposit can theoretically be obtained only if the properties of the substrate are known and as well as the thickness of each layer. Thus, the precision of the measurement also depends on the precision of these known parameters.

In the followings, our attention will be focused on two particular cases. The first one corresponds to a conductive deposit on an insulating material. The second one corresponds to an insulating film on a conductive substrate.

In these two cases, the materials we consider have low thicknesses and are good conductors. So, the Biot number based on the properties of the substrate $Bi = he_2/\lambda_2$ is low and it is possible, as a first approximation, to neglect its influence on the measured reduced rear face contrast ΔT^* .

It can be shown that even in the presence of heat losses, there is some kind of compensation through the construction of this contrast, which is a difference, which means that the present adiabatic model is a robust one: we will see in a later section that this parameter has a low influence in the estimation of the coating properties. The thicknesses and thermophysical properties are given in **Table 3**.

	Thickness (μm)	a (m^2/s)	λ ($\text{W}/\text{m}/^\circ\text{K}$)	ρC_p ($\text{J}/\text{m}^3/^\circ\text{K}$)
Case 1 :	Aluminium coating on a Cobalt/Nickel substrate			
Film (1)	220	$9.46 \cdot 10^{-5}$	230	$2.43 \cdot 10^6$
Substrate (2)	1 100	$2.36 \cdot 10^{-5}$	84.5	$3.57 \cdot 10^6$
Case 2 :	Insulating film on a Alumina substrate			
Film (1)	247	$6.84 \cdot 10^{-7}$	2.23	$3.26 \cdot 10^6$
Substrate (2)	640	$7.47 \cdot 10^{-6}$	23	$3.08 \cdot 10^6$

Table 3: Thermophysical properties and thicknesses of the materials

The reduced thermograms for the substrate and two-layer material as well as the contrast curve are plotted for the conductive/insulating and insulating/conductive cases in **Figure 9** and **Figure 10** respectively.

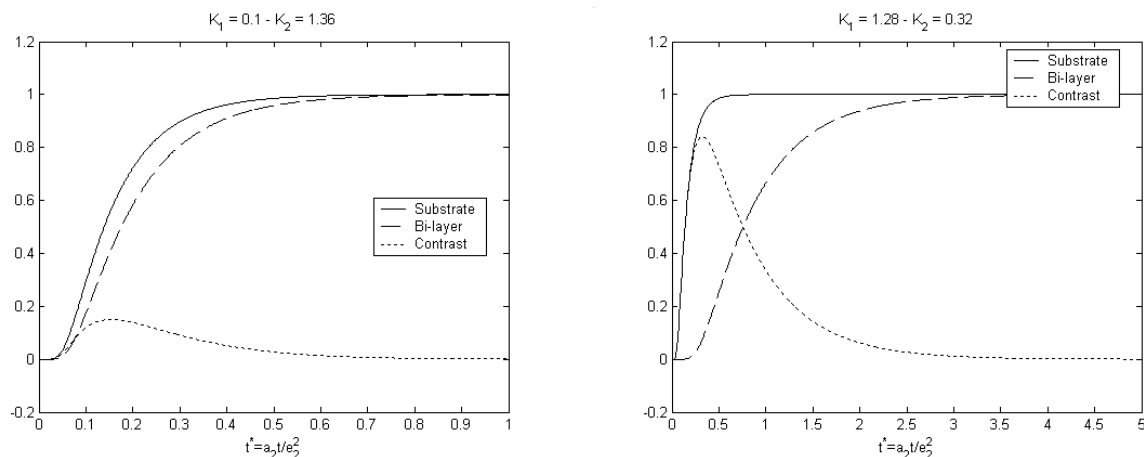


Figure 9 : Case 1 – Conductive coating / Insulating substrate

Figure 10 : Case 2 – Insulating film / Conductive substrate

➤ **Sensitivity Study**

The contrast curves and reduced sensitivities to parameters K_1 and K_2 for the two cases considered (conductive and insulating deposits) are plotted in **Figure 11** and **Figure 12**.

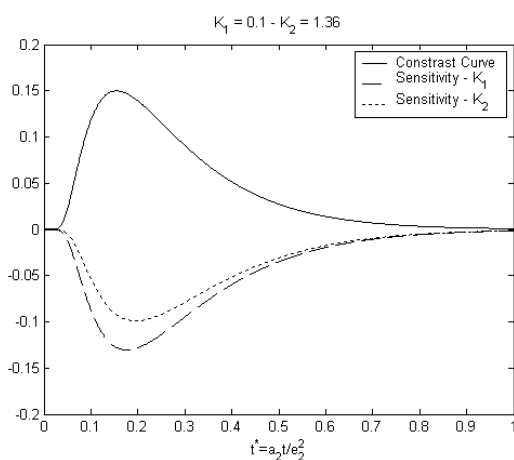


Figure 11 : Contrast curve and reduced sensitivities to K_1 and K_2 (Case 1)

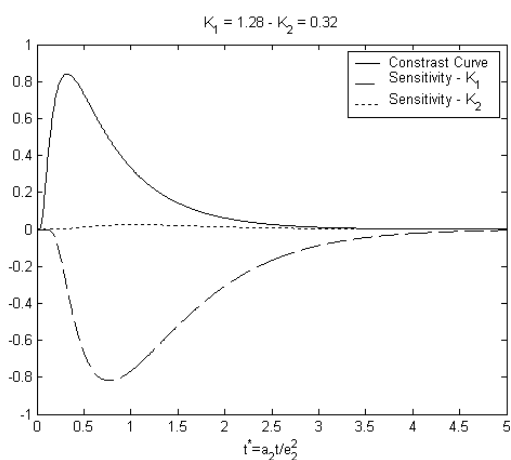


Figure 12 : Contrast curve and reduced sensitivities to K_1 and K_2 (Case 2)

These two examples are representative of most of the cases that can be met. In the first case, both sensitivities are of the same order of magnitude but seem to be strongly correlated: they exhibit a nearly constant ratio, which means that they are proportional. In the second case, one of the sensitivity is low.

➤ **Covariance and correlation matrices**

Table 4 gives the scaled covariance matrix $\text{rcov}(\hat{\mathbf{K}}) = \sigma^2(\mathbf{S}^* \mathbf{S}^*)^{-1}$ defined in (35), as well as the correlation matrix $\text{cor}(\hat{\mathbf{K}})$ defined in (15), for the two cases considered (the standard-deviation of noise σ is taken equal to unity here and 1000 points in time are used for the simulation of the thermal contrast curve).

Scaled Variance-Covariance	Scaled Variance-Covariance
$\begin{bmatrix} 28.0302 & -35.9846 \\ -35.9846 & 46.6417 \end{bmatrix}$	$\begin{bmatrix} 0.1067 & 3.1409 \\ 3.1409 & 99.1677 \end{bmatrix}$
Correlation	Correlation

$\begin{matrix} 1.0000 & -0.9952 \\ -0.9952 & 1.0000 \end{matrix}$	$\begin{matrix} 1.0000 & 0.9655 \\ 0.9655 & 1.0000 \end{matrix}$
Case 1	Case 2

Table 4 : Reduced covariance and correlation matrices K_1 and K_2 (for $\sigma = 1$)

The most interesting information is given by the reduced variance-covariance matrix $\text{rcov}(\hat{\mathbf{K}})$: it takes into account at the same times the reduced sensitivities through the inversion of the reduced information matrix $\mathbf{S}^{*T}\mathbf{S}^*$ as well as the noise through its standard deviation σ .

We calculate now the square root of the diagonal terms of matrix $\text{rcov}(\hat{\mathbf{K}})$, that is the relative standard deviations of the estimates of each parameter K_1 and K_2 , for a reduced standard deviation of the noise on each of the two T_2^* and $T_{1/2}^*$ scaled thermograms now equal to $\sigma^* = 0.01$. This corresponds to a signal over noise ratio of 100. So measurement of the (experimental) reduced thermal contrast ΔT^{*exp} is affected by a (relative) standard deviation ΔT^* equal to $\sqrt{2}\sigma^*$ (for two independent experiments, because $\text{var}(\Delta T^{*exp}) = \text{var}(T_2^{*exp}) + \text{var}(T_{1/2}^{*exp}) = 2\sigma^{*2}$), one gets (application of equation (35) with $\sqrt{2}\sigma^*$ replacing σ):

$$\text{- for case 1: } \begin{cases} \sigma_{\hat{K}_1} / K_1 = \sqrt{2}\sigma^* \sqrt{28.0302} = 0.0749 \approx 7.5\% & \text{for } K_1 = 0.1 \\ \sigma_{\hat{K}_2} / K_2 = \sqrt{2}\sigma^* \sqrt{46.6417} = 0.0966 \approx 9.5\% & \text{for } K_2 = 1.36 \end{cases} \quad (71)$$

It is interesting to calculate the singular values of the reduced sensitivity matrix \mathbf{S}^* . They are the square roots of the eigenvalues (equal to the singular values) of the reduced information matrix $\mathbf{S}^{*T}\mathbf{S}^*$ and can also be calculated through the inverse of the eigenvalues of $(\mathbf{S}^{*T}\mathbf{S}^*)^{-1}$:

$$\begin{aligned} w_1(\mathbf{S}^*) &= (w_1(\mathbf{S}^{*T}\mathbf{S}^*))^{1/2} = 1/(w_2((\mathbf{S}^{*T}\mathbf{S}^*)^{-1}))^{1/2} = 2.4347 \\ w_2(\mathbf{S}^*) &= (w_2(\mathbf{S}^{*T}\mathbf{S}^*))^{1/2} = 1/(w_1((\mathbf{S}^{*T}\mathbf{S}^*)^{-1}))^{1/2} = 0.1159 \end{aligned} \quad (72)$$

This allows to get the condition number of \mathbf{S}^* (see Lecture L3):

$$\text{cond}(\mathbf{S}^*) = w_1(\mathbf{S}^*)/w_2(\mathbf{S}^*) = 21 \quad (73)$$

We can also calculate the root mean square reduced standard deviation m_q of the estimates of both parameters K_1 and K_2 defined in (37):

$$m_q = \sigma^* \sqrt{2} \left(1/w_1^2 + 1/w_2^2\right)^{1/2} = 0.0864 \quad (74)$$

It is easy to check that this value is simply the root mean square of the relative standards deviations given in (71).

Let us note that this value (73) is close to the lower bound of m_q defined in (38), here:

$(\sigma^* \sqrt{2})/(\sqrt{2} w_2) = \sigma^* / w_2 = 0.0862$. The smallest singular value is mostly responsible for the relative errors on both parameters.

The same calculations can be made for the second case:

$$\text{- for case 2: } \boxed{\begin{array}{ll} \sigma_{\hat{K}_1} / K_1 = \sqrt{2} \sigma^* \sqrt{0.1067} = 0.0046 \approx 0.5\% & \text{for } K_1 = 1.28 \\ \sigma_{\hat{K}_2} / K_2 = \sqrt{2} \sigma^* \sqrt{99.1677} = 0.1408 \approx 14.1\% & \text{for } K_2 = 0.32 \end{array}} \quad (75)$$

$$\text{and :} \quad w_1(\mathbf{S}^*) = 11.7851 \quad w_2(\mathbf{S}^*) = 0.1004 \quad (76)$$

So, the condition number of \mathbf{S}^* is:

$$\text{cond}(\mathbf{S}^*) = w_1(\mathbf{S}^*)/w_2(\mathbf{S}^*) = 117 \quad (77)$$

which means that matrix \mathbf{S}^* is more ill-conditioned in the second case with respect to the first one.

One also gets here:

$$m_q = 0.0996 \quad \text{and lower bound for } m_q: \sigma^* / w_2 = 0.0996 \quad (78)$$

So, returning to case 1, it appears clearly that both the ratios K_1 of the characteristic times and K_2 of the effusivities can be estimated with a relative error nearly equivalent for both parameters (in the 7 to 10 % interval): this was already apparent in **Figure 11** where the reduced sensitivity curves corresponding to both parameters were very close, with a slightly higher absolute value for the sensitivity to K_1 .

For case 2, it is clearly the ratio K_1 of the characteristic times that can be reached, with a very good precision (0.5 % here): this is quite natural since the reduced sensitivity to K_2 in **Figure 12** is close to zero. So, because of the nonlinear character of this PEP problem, the accessible parameter depends on the location of the (K_1, K_2) parameter vector in the \mathbf{R}^2 plane. The question that remains is to know if is possible to measure, with higher precisions, two parameters derived from (K_1, K_2) using the experiment corresponding to case 1 for example. Let us introduce for instance the (K_3, K_4) pair instead of (K_1, K_2) in the analytical model.

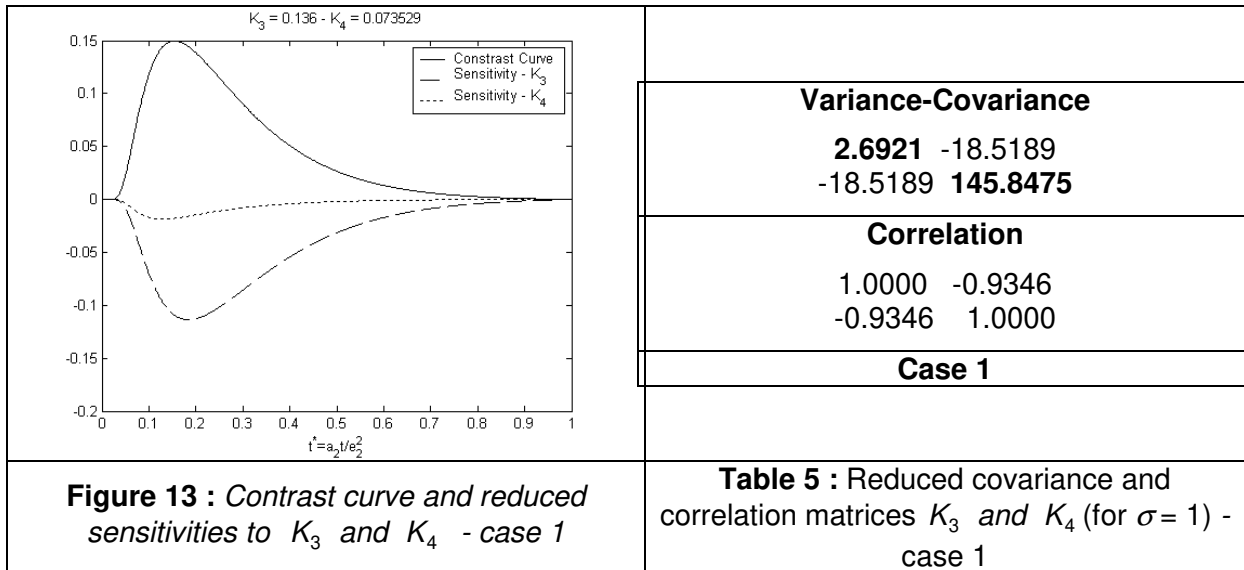


Figure 13 : Contrast curve and reduced sensitivities to K_3 and K_4 - case 1

Table 5 : Reduced covariance and correlation matrices K_3 and K_4 (for $\sigma = 1$) - case 1

The thermal contrast is naturally the same (the materials are identical).

Table 5 gives the scaled covariance matrix $\text{rcov}(\hat{\mathbf{K}})$ as well as the correlation matrix $\text{cor}(\hat{\mathbf{K}})$ for the estimator of $\mathbf{K} = [K_3 \ K_4]^T$. The relative standard deviation of both parameters becomes (for $\sigma^* = 0.01$):

$$\begin{aligned}
 &\text{- for case 1: } \left[\begin{array}{ll} \sigma_{\hat{K}_3} / K_3 = \sqrt{2} \sigma^* \sqrt{2.6921} = 0.0232 \approx 2.3\% & \text{for } K_3 = 0.136 \\ \sigma_{\hat{K}_4} / K_4 = \sqrt{2} \sigma^* \sqrt{145.8475} = 0.1708 \approx 17.1\% & \text{for } K_4 = 0.0735 \end{array} \right] \quad (79)
 \end{aligned}$$

So, when comparing (79) and (71), one clearly sees that instead of having (K_1, K_2) with quite poor precisions, the (K_3, K_4) allows to retrieve very precise values for the ratio of volumetric heat capacities K_3 . This was already apparent in **Figure 13**: the relative sensitivity to K_4 was quite low when compared to the one of K_3 , but both minima of the corresponding curves occurred at times far apart, with a degree of collinearity much weaker than in figure 11 (see also section 3.3.2 of this lecture).

This result obtained for the two cases can be explained from the expression of the contrast curve.

$$\Delta \tilde{\theta}^* = \frac{1}{s} \left[\frac{1 + K_1 K_2}{K_2 \sinh(K_1 s) \cosh(s) + \sinh(s) \cosh(K_1 s)} - \frac{1}{\sinh(s)} \right] \quad (80)$$

In the previous case (conductive coating on an insulating substrate), K_1 is close to zero. A rough approximation can be obtained by setting: $\begin{cases} \sinh(K_1 s) \approx K_1 s \\ \cosh(K_1 s) \approx 1 \end{cases}$

$$\Delta \tilde{\theta}^* = \frac{1}{s} \left[\frac{1 + K_3}{K_3 s \cosh(s) + \sinh(s)} - \frac{1}{\sinh(s)} \right] \quad (81)$$

We can see then that within this first order approximation, the model is only a function of $K_3 = K_1 K_2$. We can check the other criteria already considered for case 1 with the (K_1, K_2) parameters:

$$w_1(\mathbf{S}^*) = 1.7270 \quad w_2(\mathbf{S}^*) = 0.0821 \quad (82)$$

So, the condition number of \mathbf{S}^* is:

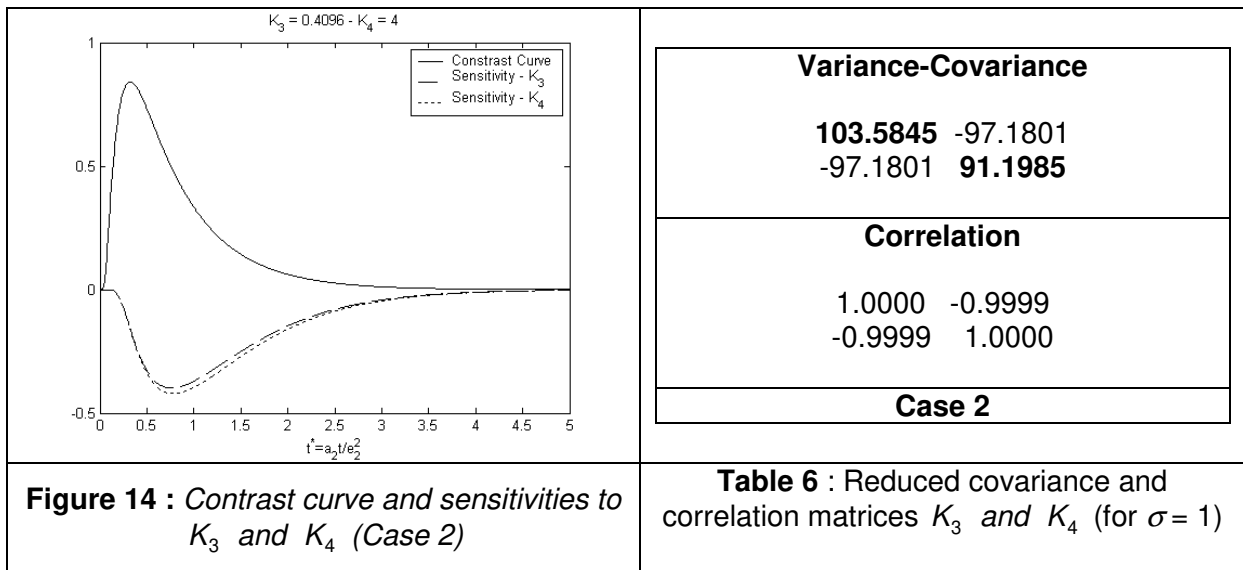
$$\text{cond}(\mathbf{S}^*) = w_1(\mathbf{S}^*)/w_2(\mathbf{S}^*) = 21 \quad (83)$$

Compared to the preceding parameterization, the reduced sensitivity matrix \mathbf{S}^* as well as its singular values have changed, but the condition number is the same, see (73).

One also gets here:

$$m_q = 0.1219 \quad \text{and lower bound for } m_q: \sigma^* / w_2 = 0.1218 \quad (84)$$

When both m_q 's are compared, see (74), one can say that the global precision of the estimation of the (K_3, K_4) parameterization is lower than the (K_1, K_2) one. However, we will see later on that this superiority of the (K_3, K_4) parameterization is only an apparent one if both thermophysical characteristics of the film are looked for.



In case 2 (insulating coating on a conductive substrate), parameters K_3 and K_4 are strongly correlated and exhibit the same sensitivity curves – see **Figure 14**. This confirms the result we observed previously, that is a thermal contrast mostly sensitive to K_1 .

$$K_3 K_4 = \frac{C_1}{C_2} \frac{R_1}{R_2} = \frac{R_1 C_1}{R_2 C_2} = \frac{tc_1}{tc_2} = K_1^2 \quad (85)$$

This can be also explained by the fact that K_1 is close to unity:

$$\sinh(K_1 s) \cosh(s) \simeq K_1 \sinh(s) \cosh(K_1 s) \quad (86)$$

This yield:

$$\Delta \tilde{\theta}_{out}^* = \frac{1}{s} \left[\frac{1}{\sinh(s) \cosh(K_1 s)} - \frac{1}{\sinh(s)} \right] \quad (87)$$

So, the thermal contrast is mainly a function of K_1 . Returning to the same calculation as in the other case, using **Table 6**, one gets:

- for case 2:	$\sigma_{\hat{K}_3} / K_3 = \sqrt{2} \sigma^* \sqrt{103.5845} = 0.1439 \approx 14.4\%$	for	$K_3 = 0.4096$	(88)
	$\sigma_{\hat{K}_4} / K_4 = \sqrt{2} \sigma^* \sqrt{91.1985} = 0.1351 \approx 13.5\%$	for	$K_4 = 4$	

The singular values of the reduced sensitivity matrix are:

$$w_1(\mathbf{S}^*) = 8.3624 \quad w_2(\mathbf{S}^*) = 0.0717 \quad (89)$$

So, the condition number of \mathbf{S}^* is:

$$\text{cond}(\mathbf{S}^*) = w_1(\mathbf{S}^*)/w_2(\mathbf{S}^*) = 117 \quad (90)$$

We observe here the same thing as for case 2: the condition number of the reduced sensitivity matrix is independent of the parameterization, see (77).

One also gets here:

$$m_q = 0.1396 \quad \text{and lower bound for } m_q: \sigma^* / w_2 = 0.1395 \quad (91)$$

When both m_q 's are compared, see (78), one can say that the global precision of the estimation of the (K_3, K_4) parameterization, which provided an excellent estimation for K_3 , is lower than the (K_1, K_2) one.

4.3 Note on the change of parameters

It has been suggested earlier that some change of parameterization would allow to overcome parameter estimation difficulties such as in the case of high correlation coefficients inducing high variances for the estimated parameters for example. We want here to come back to this discussion to give, very briefly, some precisions and our conclusions.

First, and taking experience of what has been shown previously, if a change of parameterization is made that results in the production of a new parameter of sensitivity close to zero (and thereof excluded from the model), this new parameterization will have a positive effect and will allow to properly estimate the remaining ones. Note that it is the object of Dimensional Analysis to help making such reparameterization efficient.

Second, if all the parameters of the problem have non negligible sensitivities but appear correlated, the question is: is it possible to find a new set of parameters defined from the initial one, to enhance the quality of the estimation process?

The answer is no. It can be demonstrated, see Remy [9] that the sensitivities to a new set of parameters can be derived from the sensitivities of the current set (using the Jacobian of the transformation). The same is true for the variance-covariance matrix and the explanation is obvious from the quantified SVD analysis given above (the same condition number of \mathbf{S}^* is obtained whatever set of parameterizations is used) These relationships show that:

- if two parameters appear correlated in a given set of parameters, two parameters of a new set, recombined from the previous ones, will also be correlated,
- if the sensitivity of a parameter is changed with a new parameterization (for example, it is enhanced), this will not change its variance ultimately.

For instance, if we keep the parameter K_1 and choose another second parameter instead of K_2 , we can show that the sensitivity curve to K_1 can become higher or lower: we have to remind that the partial derivative that appears in the definition (4) of a sensitivity coefficient is associated to the variation of the output of the model for a variation of a given parameter, which requires that the other ones stay fixed at given values. This means that if the definition of these other parameters is changed, such is also the case for the sensitivity coefficients. So, talking of a sensitivity coefficient to a given parameter does not mean anything if the other parameters in the parameter vector are not specified.

So, one can wonder whether it would be possible to improve the estimation of K_1 by combining this parameter with a particular parameter that can increase its sensitivity. In fact, this is not true because the standard-deviations of the estimates of the new parameters do not only depend on the sensitivities of the old parameters but also on the correlation between the estimates of the old parameters.

To show this, we are going to see through an example how the standard-deviations (square roots of variances) of the new set of parameters change when one parameter is kept as for instance parameter K_1 , that is $K_a = K_1^\alpha K_2^\beta$ with $\alpha=1; \beta=0$, while K_2 is replaced by $K_b = F_b(K_1, K_2)$:

$$\begin{aligned} K_a &= F_a(K_1) = K_1 \\ K_b &= F_b(K_1, K_2) \end{aligned} \quad (92)$$

We have:

$$\mathbf{y}_{mo} = \boldsymbol{\eta}(\mathbf{t}; \mathbf{K}) \text{ with } \mathbf{K} = \begin{bmatrix} K_1 \\ K_2 \end{bmatrix} \Rightarrow d\mathbf{y}_{mo} = \mathbf{S} d\mathbf{K} = \mathbf{S}' d\mathbf{K}' \text{ with } \mathbf{K}' = \begin{bmatrix} K_a \\ K_b \end{bmatrix} \quad (93)$$

where \mathbf{S} is the sensitivity matrix to the old (K_1, K_2) set of parameters and \mathbf{S}' the sensitivity matrix to the new (K_a, K_b) one. This requires the calculation of the Jacobian matrix \mathbf{J} of this transformation since;

$$d\mathbf{K}' = \mathbf{J} d\mathbf{K} \quad \Rightarrow \quad \mathbf{S} = \mathbf{S}' \mathbf{J} \quad \text{and} \quad \text{cov}(\hat{\mathbf{K}}') = \mathbf{J} \text{cov}(\hat{\mathbf{K}}) \mathbf{J}^T \quad (94)$$

The last equation in (94) stems from the linearization around the exact value of the \mathbf{K} parameter vector:

$$\text{cov}(\hat{\mathbf{K}}) = \text{cov}(d\hat{\mathbf{K}}) \quad (95)$$

with:

$$\mathbf{J} = \frac{D(F_a, F_b)}{D(K_1, K_2)} = \begin{bmatrix} \frac{\partial F_a}{\partial K_1} & \frac{\partial F_a}{\partial K_2} \\ \frac{\partial F_b}{\partial K_1} & \frac{\partial F_b}{\partial K_2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ F_{b,1} & F_{b,2} \end{bmatrix} \quad (96)$$

So, the sensitivity matrix to the new parameter set \mathbf{K}' is:

$$\mathbf{S}' = [\mathbf{s}_a \quad \mathbf{s}_b] = \mathbf{S} \mathbf{J}^{-1} = [\mathbf{s}_1 \quad \mathbf{s}_2] \begin{bmatrix} 1 & 0 \\ -F_{b,1}/F_{b,2} & 1/F_{b,2} \end{bmatrix} = [\mathbf{s}_1 - (F_{b,1}/F_{b,2})\mathbf{s}_2 \quad (1/F_{b,2})\mathbf{s}_2] \quad (97)$$

Here the old sensitivity column vectors \mathbf{s}_1 and \mathbf{s}_2 , as well as the new ones \mathbf{s}_a and \mathbf{s}_b , have been explicitly written in terms of the corresponding sensitivity matrices, \mathbf{S} and \mathbf{S}' respectively.

Application of (94) allows the calculation of the variances and covariance of the estimators of the new set of parameters (K_a, K_b) :

$$\text{cov}(\hat{\mathbf{K}}') = \begin{bmatrix} \text{var}(\hat{K}_a) & \text{cov}(\hat{K}_a, \hat{K}_b) \\ \text{cov}(\hat{K}_a, \hat{K}_b) & \text{var}(\hat{K}_b) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ F_{b,1} & F_{b,2} \end{bmatrix} \begin{bmatrix} \text{var}(\hat{K}_1) & \text{cov}(\hat{K}_1, \hat{K}_2) \\ \text{cov}(\hat{K}_1, \hat{K}_2) & \text{var}(\hat{K}_2) \end{bmatrix} \begin{bmatrix} 1 & F_{b,1} \\ 0 & F_{b,2} \end{bmatrix} \quad (98)$$

that is:

$$\begin{aligned} \text{var}(\hat{K}_a) &= \text{var}(\hat{K}_1) \\ \text{var}(\hat{K}_b) &= F_{b,1}^2 \text{var}(\hat{K}_1) + F_{b,2}^2 \text{var}(\hat{K}_2) + 2 F_{b,1} F_{b,2} \text{cov}(\hat{K}_1, \hat{K}_2) \\ \text{cov}(\hat{K}_a, \hat{K}_b) &= F_{b,1} \text{var}(\hat{K}_1) + F_{b,2} \text{cov}(\hat{K}_1, \hat{K}_2) \end{aligned} \quad (99)$$

We can see that even if the change of parameters modifies the sensitivity to parameter K_a , that replaces parameter K_1 in the new set of parameters, the variance of this parameter remains unchanged whatever the choice of the second parameter.

This means that the variance of a given parameter (and consequently the error on this parameter) is independent on the choice of the second parameter. Thus, identifying the parameter K_1 from the (K_1, K_2) pair is equivalent to estimating K_1 from the (K_1, K_3) or (K_1, K_4) pairs.

Similarly, we can show that estimating parameters (K_3, K_4) either through the parameterization (K_1, K_2) or directly, is strictly the same.

The conclusion is that the interest of a change of parameters is justified only when an improved estimation of a particular parameter of interest is looked for.

Whatever the parameterization, if the thicknesses of both layers are known, as well as the thermophysical properties of the substrate, we have:

$$\begin{aligned} \sigma_{\rho\hat{c}_1} / \rho c_1 = \sigma_{\hat{K}_3} / K_3 = 2.3 \% \quad \text{for case 1} \\ \sigma_{\hat{a}_1} / a_1 = \sigma_{\hat{K}_1} / K_1 = 0.5 \% \quad \text{for case 2} \end{aligned} \quad (100)$$

These relative standard deviation of the estimated thermophysical properties of the front face layer are valid for a signal to noise ratio equal to 100 for the experimental thermogram of each flash experiment (single substrate layer and two-layer sample). So, this rear face thermal contrast technique allows estimation of the capacity of the film for case 1 and of its diffusivity in case 2, for high enough signal over noise ratios.

In case of very low sensitivity to a given parameter, it is possible to fix the value of the corresponding parameter to its nominal values. So, if the number of parameters that are looked for is reduced, then the stochastic errors on the remaining parameters (reduced standard deviations) decrease. However, their estimation becomes biased and leads to a systematic error on each estimated parameter such as:

$$\mathbf{b}_{\hat{\beta}_r} = E(\hat{\beta}_r) - \beta_r = -(\mathbf{S}_r^T \mathbf{S}_r)^{-1} \mathbf{S}_r^T \mathbf{S}_c (\beta_c^{nom} - \beta_c^{exact}) \quad (101)$$

Here the initial parameter vector has been decomposed into two parts $\beta = \begin{bmatrix} \beta_r \\ \beta_c \end{bmatrix}$, where β_r gathers the parameters that are looked for and its complementary part β_c is supposed to be known, that is its value is blocked to a nominal value $\beta_c = \beta_c^{nom}$ which differs from its exact value β_c^{exact} . Equation (101), which has already been derived in the case of a linear model in lecture L3 of this series (see also [1]), corresponds here to a linearization in the neighborhood of the exact value of β .

This technique, which consists in reducing the number of parameters that are looked for, presents an interest only if the bias caused by the reduction of the number of parameters and its associated standard deviations are much lower than the initial stochastic error as illustrated in **Figure 15**.

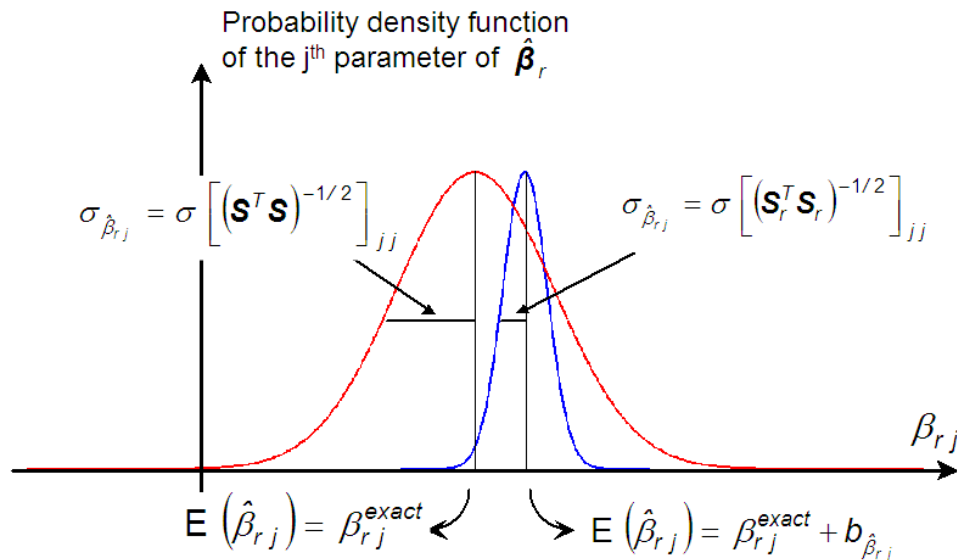


Figure 15 : Comparison between the probability density distributions of the j^{th} parameter of the parameter vector for two different estimators 1) all the parameters in β are estimated altogether (red) or 2) only the components of one of its part β_r (blue) are estimated while its complementary part β_c are blocked to its nominal value. **NB:** here one assumes that index j in β and in β_r are the same ($\beta_{r,j} = \beta_j$) and that the scale of the vertical axis is different for both distributions for practical plotting reasons (the area below both distributions should be equal to unity)

5. Conclusion

Useful tools have been introduced for the analysis of estimations (variance-covariance matrix) and the detection of the ill-conditioned character of the Parameter Estimation Problem (PEP). Different techniques have been presented for tracking the true degrees of freedom of a given PEP (matrix rank, correlations between parameters, SVD, ...). If we want to enhance the estimation of a given parameter, one solution is to use a reduced model. This reduced model can be either unbiased or biased. It is of particular interest to know if a reduced model is biased or not.

We have proposed, in the last section of the lecture, to work with a variable estimation time interval in order to evaluate the systematic error caused in the estimated parameters. We hope that the different "realistic" examples of thermal metrology presented in this lecture will help the reader to master the corresponding tools to get good estimates in a PEP.

References

- [1] Chapter 9: Nonlinear estimation techniques, B. Rémy, S. André, in *Thermal Measurements and Inverse Techniques*, 2011, Orlande H R.B., Fudym O., Maillet D. and Cotta R. M., editors: CRC Press, Series: Heat Transfer, 770p.
- [2] J.V. Beck, K..J. Arnold, *Parameter Estimation in Engineering and Science*, John Wiley & Sons, 1977.
- [3] K. Levenberg, A method for the solution of certain problems in Least Squares, *Quart. Appl. Math.* 2, 164-168, 1944.
- [4] Gallant, A.R., 1975, Nonlinear regression, *Am. Stat.*, 29:73-81.
- [5] Andre S., Serra J.J., Cella N and Silva Neto A.J., Parameter Estimation of Thermo-optical Glass Properties From Experimental Phase Lag Signals Obtained Through Periodic Heat Flux Excitation, *Proc. 4th International Conference on Inverse Problems in Engineering: Theory and Practice*, May 26-31, 2002, Angra dos Reis, Brasil.
- [6] The Physical Basis of dimensional analysis, Ain A. Sonin, MIT course manuscript, 2001 <http://me.mit.edu/people/sonin/html>
- [7] Maillet D., André S., Batsale J.C., Degiovanni A. and Moyne C., *Thermal Quadrupoles : Solving the Heat equation Through Integral Transforms*, Chichester, PA: John Wiley & Sons Ltd, 2000.
- [8] Chapter 1: Modelling in heat transfer, J.-L. Battaglia and D. Maillet, in *Thermal Measurements and Inverse Techniques*, 2011, Orlande H.R.B., Fudym O., Maillet D. and Cotta R. M., editors: CRC Press, Series: Heat Transfer, 770p.
- [9] B. Remy and A. Degiovanni, "Parameters Estimation and Measurement of Thermophysical Properties of Liquids", *International Journal of Heat & Mass Transfer* – Vol. 48, Issues 19-20, September 2005, pp 4103-4120.
- [10]. F.R. de Hoog, J.H. Knight, A.N. Stokes, An improved method for numerical inversion of Laplace transforms, *SIAM. J. Sci. Stat. Comp.* 3, 357-366, (1982)
- [11] http://www.mathworks.com/matlabcentral/answers/uploaded_files/1034/invlap.m

Appendix 1 - Reminder of the Singular Value Decomposition of a rectangular matrix

Any rectangular matrix (called \mathbf{K} here) with real coefficients and of dimensions (m, n) with $m \geq n$, can be written under the form :

$$\mathbf{K} = \mathbf{U} \mathbf{W} \mathbf{V}^T, \text{ that is } \begin{bmatrix} \mathbf{K} \end{bmatrix} = \begin{bmatrix} \mathbf{U} \end{bmatrix} \begin{bmatrix} w_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & w_n \end{bmatrix} \begin{bmatrix} \mathbf{V}^T \end{bmatrix} \quad (\text{A1a})$$

This expression is sometimes called "lean" singular decomposition or "economical" SVD and involves

- \mathbf{U} , an orthogonal matrix of dimensions (m, n) , its column vectors (the *left* singular vectors of \mathbf{K}) have a unit norm and are orthogonal by pairs: $\mathbf{U}^T \mathbf{U} = \mathbf{I}_n$, where \mathbf{I}_n is the identity matrix of dimension n . Its columns are composed of the first n eigenvectors \mathbf{U}_k , ordered according to decreasing values of the eigenvalues of matrix $\mathbf{K} \mathbf{K}^T$. Let us note that, in the general case, $\mathbf{U} \mathbf{U}^T \neq \mathbf{I}_m$.

- \mathbf{V} , a square orthogonal matrix of dimensions (n, n) , $\mathbf{V} \mathbf{V}^T = \mathbf{V}^T \mathbf{V} = \mathbf{I}_n$. Its column vectors (the *right* singular vectors of \mathbf{K}), are the n eigenvectors \mathbf{V}_k , ordered according to decreasing eigenvalues, of matrix $\mathbf{K}^T \mathbf{K}$;

- \mathbf{W} , a square diagonal matrix of dimensions (n, n) , that contains the n so-called *singular* values of matrix \mathbf{K} , ordered according to decreasing values: $w_1 \geq w_2 \geq \dots \geq w_n$. The singular values of matrix \mathbf{K} are defined as the square roots of the eigenvalues of matrix $\mathbf{K}^T \mathbf{K}$. If matrix \mathbf{K} is square and symmetric, the eigenvalues and the singular values of \mathbf{K} are the same.

Another SVD form called "Full Singular Value Decomposition" is available for matrix \mathbf{K} . In this equivalent definition, both matrices \mathbf{U} and \mathbf{W} are changed: the matrix replacing \mathbf{U} is now square (size $m \times m$) and the matrix replacing \mathbf{W} is now diagonal but non square (size $m \times n$). In the case $m \geq n$, this can be written:

$$\mathbf{K} = \mathbf{U}_0 \mathbf{W}_0 \mathbf{V}^t \quad \text{with} \quad \mathbf{U}_0 = \begin{bmatrix} \mathbf{U} & \mathbf{U}_{comp} \end{bmatrix}; \quad \mathbf{W}_0 = \begin{bmatrix} \mathbf{W} \\ \mathbf{0}_{(m-n) \times n} \end{bmatrix} \quad \text{and} \quad \dim(\mathbf{U}_{comp}) = m \times (m - n)$$

(A1b)

or:

$$\begin{bmatrix} \mathbf{K} \end{bmatrix} = \begin{bmatrix} \mathbf{U} & \mathbf{U}_{comp} \end{bmatrix} \begin{bmatrix} w_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & w_n \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \mathbf{V}^t \end{bmatrix} \quad (\text{A1c})$$

Matrix \mathbf{U}_{comp} is composed of the $(m - n)$ left singular column vectors do not present in \mathbf{U} . So, the concatenated matrix \mathbf{U}_0 verifies now:

$$\mathbf{U}_0^t \mathbf{U}_0 = \mathbf{U}_0 \mathbf{U}_0^t = \mathbf{U} \mathbf{U}^t + \mathbf{U}_{comp} \mathbf{U}_{comp}^t = \mathbf{I}_m \quad (\text{A1d})$$

This singular value decomposition (A1b) can be implemented for any matrix \mathbf{K} , with real value coefficients, for $m \geq n$.

Appendix 2 - Singular Value Decomposition of the scaled sensitivity matrix

This singular value decomposition can be implemented for any matrix \mathbf{K} .

A double change of basis, in the measurement domain and in the parameter domain, using the matrices of the left \mathbf{U} and right \mathbf{V} , in the SVD of \mathbf{S}^* written for $\mathbf{K} = \mathbf{S}^*$ yields:

$$\mathbf{S}^* = \mathbf{U} \mathbf{W} \mathbf{V}^T \quad (\text{A2})$$

Matrix \mathbf{V} is used as a (square) change of matrix basis and it transforms the differential of the reduced parameter vector $d\mathbf{x}$, see (29) into a new differential vector $d\mathbf{p}$, where \mathbf{p} can be called the diagonal parameter vector, of dimensions $(n, 1)$.

Matrix \mathbf{U} allows to change the differential observation vector $d\mathbf{y}_{mo}$ of dimensions $(m, 1)$ into a differential vector $d\mathbf{z}_{mo}$ of smaller length, where \mathbf{z}_{mo} can be called the diagonal observation vector, of dimensions $(n, 1)$.

$$d\mathbf{y}_{mo} = \mathbf{U} d\mathbf{z}_{mo} \quad \text{and} \quad d\mathbf{x} = \mathbf{V} d\mathbf{p} \quad (\text{A3a, b})$$

Let us note here that the reduction of the length of the observation vector (m observations for $d\mathbf{y}_{mo}$ and only n components in $d\mathbf{z}_{mo}$ stems from the fact that the $(m-n)$ singular eigenvectors \mathbf{U}_k not present in matrix \mathbf{U} corresponds to null singular values w_k (for $k > n$).

Use of equations (A1) to (A3), together with the property $\mathbf{U}^T \mathbf{U} = \mathbf{V}^T \mathbf{V} = \mathbf{I}_n$, allows to get the equivalent of the differential model (31a) in the double transformed space:

$$d\mathbf{z}_{mo} = \mathbf{W} d\mathbf{p} \quad (\text{A4})$$

This equation corresponds to a diagonalization of the model in R^n , and one gets then, component by component:

$$dp_k = \frac{1}{w_k} dz_{mo,k} \quad \text{for } k = 1, 2, \dots, n \quad (\text{A5})$$

Combining (A3a, b) and (A4) yields:

$$d\mathbf{x} = \mathbf{V} \mathbf{W}^{-1} \mathbf{U}^T d\mathbf{y}_{mo} = \mathbf{S}^{*+} d\mathbf{y}_{mo} \quad (\text{A6})$$

where $\mathbf{S}^{*+} = \mathbf{V} \mathbf{W}^{-1} \mathbf{U}^T$ is the pseudo-inverse, or Moore-Penrose inverse, of the scaled sensitivity matrix \mathbf{S}^* .

Combination of the preceding equations leads to a relationship between $d\boldsymbol{\beta}$ and $d\mathbf{y}_{mo}$:

$$d\boldsymbol{\beta} = \mathbf{R}_{nom} \mathbf{V} \mathbf{W}^{-1} \mathbf{U}^T d\mathbf{y}_{mo} \quad (\text{A7})$$

and an integration can be implemented to give the relationship between the diagonal and original sets of parameters in a column vector form:

$$\mathbf{p} = \mathbf{V}^T \mathbf{x} = \mathbf{V}^T \ln(\mathbf{R}_{nom}^{-1} \boldsymbol{\beta}) \approx \mathbf{V}^T \mathbf{R}_{nom}^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}^{nom}) \quad \text{because } \mathbf{p}^{nom} = \mathbf{V}^T \mathbf{x}^{nom} = \mathbf{0} \quad (\text{A8})$$

The transformed observation vector can be expressed:

$$\mathbf{z}_{mo} = \mathbf{U}^T (\mathbf{y}_{mo} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{nom})) = \mathbf{W} \mathbf{p} \quad \text{because } \mathbf{z}_{mo}^{nom} = \mathbf{W} \mathbf{p}^{nom} = \mathbf{0} \quad (\text{A9})$$

Combining (A8) and (A9) yields:

$$\mathbf{p} = \mathbf{V}^T \ln(\mathbf{R}_{nom}^{-1} \boldsymbol{\beta}) = \mathbf{W}^{-1} \mathbf{U}^T (\mathbf{y}_{mo} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{nom})) \Rightarrow \boldsymbol{\beta} = \mathbf{R}_{nom} \exp(\mathbf{V} \mathbf{W}^{-1} \mathbf{U}^T (\mathbf{y}_{mo} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{nom}))) \quad (\text{A10})$$

An approximation of this expression in the neighbourhood of $\boldsymbol{\beta}^{nom}$ is available:

$$\boldsymbol{\beta} \approx \mathbf{R}_{nom} \left[\mathbf{1} + \mathbf{V} \mathbf{W}^{-1} \mathbf{U}^T (\mathbf{y}_{mo} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{nom})) \right] = \boldsymbol{\beta}^{nom} + \mathbf{R}_{nom} \mathbf{V} \mathbf{W}^{-1} \mathbf{U}^T (\mathbf{y}_{mo} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{nom})) \quad (\text{A11})$$

where $\mathbf{1}$ is the column vector of length n whose coefficients are equal to unity.

Appendix 3 – Non-linear Ordinary Least Square estimator and SVD

It is interesting to compare diagonal equation (A5) that shows the interest of an inversion in the left and right singular spaces with the OLS estimator (12) of parameter $\boldsymbol{\beta}$. So, if the first

order approximation in the neighbourhood of $\boldsymbol{\beta}^{nom}$ is considered, the difference between measurements and model outputs can be expressed with the *residual* vector defined in (10), and \mathbf{r}_{lin} the linearized form of this difference vector:

$$\mathbf{r}(\boldsymbol{\beta}) = \mathbf{y} - \mathbf{y}_{mo}(\boldsymbol{\beta}) \approx \mathbf{r}_{lin}(\boldsymbol{\beta}) = \mathbf{y} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{nom}) - \mathbf{S}(\boldsymbol{\beta}^{nom})(\boldsymbol{\beta} - \boldsymbol{\beta}^{nom}) \quad (\text{A12})$$

The least squares sum J_{OLS} can be written as a quadratic form J^\diamond , using the fact that $J_{OLS} = J_{OLS}^T$ (scalar):

$$\begin{aligned} J(\boldsymbol{\beta}) = \mathbf{r}^T(\boldsymbol{\beta}) \mathbf{r}(\boldsymbol{\beta}) \approx J^\diamond(\boldsymbol{\beta}) &= (\mathbf{y} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{nom}))^T (\mathbf{y} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{nom})) \\ &+ (\boldsymbol{\beta} - \boldsymbol{\beta}^{nom})^T \mathbf{S}^T(\boldsymbol{\beta}^{nom}) \mathbf{S}(\boldsymbol{\beta}^{nom}) (\boldsymbol{\beta} - \boldsymbol{\beta}^{nom}) - 2(\boldsymbol{\beta} - \boldsymbol{\beta}^{nom})^T \mathbf{S}^T(\boldsymbol{\beta}^{nom}) (\mathbf{y} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{nom})) \end{aligned} \quad (\text{A13})$$

When the minimum is reached, one gets:

$$\frac{dJ^\diamond}{d\boldsymbol{\beta}} = 0 \quad \Rightarrow \quad \mathbf{S}^T(\boldsymbol{\beta}^{nom}) \mathbf{S}(\boldsymbol{\beta}^{nom}) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{nom}) = \mathbf{S}^T(\boldsymbol{\beta}^{nom}) (\mathbf{y} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{nom})) \quad (\text{A14})$$

which leads to an approximation of the OLS estimator:

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{nom} = (\mathbf{S}^T(\boldsymbol{\beta}^{nom}) \mathbf{S}(\boldsymbol{\beta}^{nom}))^{-1} \mathbf{S}^T(\boldsymbol{\beta}^{nom}) (\mathbf{y} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{nom})) \quad (\text{A15})$$

This is exactly the same equation as the iterative algorithm (12), with $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{OLS}^{(k+1)}$ and $\boldsymbol{\beta}^{nom} = \hat{\boldsymbol{\beta}}_{OLS}^{(k)}$. One shows, using (31b) and (A2):

$$(\mathbf{S}^T(\boldsymbol{\beta}^{nom}) \mathbf{S}(\boldsymbol{\beta}^{nom}))^{-1} \mathbf{S}^T(\boldsymbol{\beta}^{nom}) = \mathbf{R}_{nom} \mathbf{V} \mathbf{W}^{-1} \mathbf{U}^T \quad (\text{A16})$$

The least square estimator (A15), with the diagonal parameter \mathbf{p} and the experimental diagonal signal \mathbf{z} in their new bases, can be written thanks to (A16):

$$\hat{\mathbf{p}} = \mathbf{W}^{-1} \mathbf{z} \quad \text{with} \quad \mathbf{z} = \mathbf{U}^T (\mathbf{y} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{nom})) \quad (\text{A17a, b})$$

Equation (A17a) is diagonal. Use of (A15) and (A16) provides a new expression for the OLS estimator of $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = \mathbf{R}^{nom} (\mathbf{1} + \mathbf{V} \mathbf{W}^{-1} \mathbf{U}^T (\mathbf{y} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{nom}))) \quad (\text{A18})$$

This expression is the same as relationship (A1) that links $\boldsymbol{\beta}$ and $\mathbf{y}_{mo}(\boldsymbol{\beta})$: these corresponding two values are simply replaced by the linearized OLS estimator $\hat{\boldsymbol{\beta}}$ and by measurements \mathbf{y} respectively.

The linearized OLS estimator of the reduced parameter vector stems directly from (A19):

$$\hat{\mathbf{x}} = \left(\mathbf{S}^{*T} (\boldsymbol{\beta}^{nom}) \mathbf{S}^* (\boldsymbol{\beta}^{nom}) \right)^{-1} \mathbf{S}^{*T} (\boldsymbol{\beta}^{nom}) (\mathbf{y} - \mathbf{y}_{mo} (\boldsymbol{\beta}^{nom})) \quad (\text{A20})$$

Appendix 4 – Variance-covariance of the Non-linear Ordinary Least Square estimator and SVD

With the noise properties defined in (8), the variance-covariance of the linearized OLS estimator $\hat{\boldsymbol{\beta}}$ given by equation (A15), can be written thanks to (31b) and (A2):

$$\begin{aligned} \text{cov}(\hat{\boldsymbol{\beta}}) &= \sigma^2 \left(\mathbf{S}^T (\boldsymbol{\beta}^{nom}) \mathbf{S} (\boldsymbol{\beta}^{nom}) \right)^{-1} = \sigma^2 \left(\mathbf{R}_{nom}^{-1} \mathbf{S}^{*T} \mathbf{S}^* \mathbf{R}_{nom}^{-1} \right)^{-1} \\ &= \sigma^2 \mathbf{R}_{nom} \left(\mathbf{S}^{*T} \mathbf{S}^* \right)^{-1} \mathbf{R}_{nom} = \sigma^2 \mathbf{R}_{nom} \mathbf{V} \mathbf{W}^{-2} \mathbf{V}^T \mathbf{R}_{nom} \end{aligned} \quad (\text{A21})$$

This expression is valid if the difference between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^{nom}$ is small: it is always the case near convergence of algorithm (12) where $\boldsymbol{\beta}^{nom}$ can be redefined as $\boldsymbol{\beta}^{nom} = \hat{\boldsymbol{\beta}}_{OLS}^{(k)}$ and with $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{OLS}^{(k+1)}$.

The expression of the variance-covariance matrix of $\hat{\mathbf{x}} = \mathbf{R}_{nom}^{-1} \hat{\boldsymbol{\beta}}$ becomes:

$$\text{cov}(\hat{\mathbf{x}}) = \mathbf{R}_{nom}^{-1} \text{cov}(\hat{\boldsymbol{\beta}}) \left(\mathbf{R}_{nom}^{-1} \right)^T = \sigma^2 \mathbf{V} \mathbf{W}^{-2} \mathbf{V}^T \quad (\text{A22a})$$

The first relationship in equation (A22a) allows to calculate the reduced covariance matrix of $\hat{\boldsymbol{\beta}}$, $\text{rcov}(\hat{\boldsymbol{\beta}})$, whose diagonal coefficients are the reduced variances of the estimators of each parameter, using the nominal values of the parameters as scaling factors:

$$\text{rcov}(\hat{\boldsymbol{\beta}}) = \text{cov}(\hat{\mathbf{x}}) = \begin{bmatrix} \sigma_{\hat{\beta}_1}^2 / (\beta_1^{nom})^2 & \text{cov}(\hat{\beta}_1, \hat{\beta}_2) / (\beta_1^{nom} \beta_2^{nom}) & & \text{cov}(\hat{\beta}_1, \hat{\beta}_n) / (\beta_1^{nom} \beta_n^{nom}) \\ & \sigma_{\hat{\beta}_2}^2 / (\beta_2^{nom})^2 & & \\ & & \ddots & \\ & & & \sigma_{\hat{\beta}_n}^2 / (\beta_n^{nom})^2 \end{bmatrix} = \sigma^2 (\mathbf{S}^{*T} \mathbf{S}^*)^{-1} \quad (\text{A22b})$$

Symmetric

where $\sigma_{\hat{\beta}_j}$ is the standard deviation of $\hat{\beta}_j$. The square roots of the diagonal terms of this matrix, $\sigma_{\hat{\beta}_j} / \beta_j^{nom}$, can be considered as a measure of the relative error made for each parameter and caused by presence of noise in the measurements \mathbf{y} .

It is very interesting to calculate the trace of this matrix, which is equal to the sum of the variances of the different components of $\hat{\mathbf{x}}$:

$$\begin{aligned} \text{Tr}(\text{cov}(\hat{\mathbf{x}})) &\equiv \sum_{j=1}^n \sigma_{x_j}^2 = \sum_{j=1}^n (\sigma_{\beta_j} / \beta_j^{\text{nom}})^2 \\ \Rightarrow \sum_{j=1}^n (\sigma_{\beta_j} / \beta_j^{\text{nom}})^2 &= \sigma^2 \text{Tr}(\mathbf{V} \mathbf{W}^{-2} \mathbf{V}^T) = \sum_{k=1}^n \frac{\sigma^2}{w_k^2} \sum_{i=1}^n v_{ik}^2 \end{aligned} \quad (\text{A23})$$

where σ_{x_j} is the standard deviation of the estimate of reduced parameter x_j and σ_{β_j} the corresponding one for β_j . Since the right singular vectors have a unit norm ($\|\mathbf{v}_k\|^2 = \sum_{i=1}^n v_{ik}^2 = 1$), this last equation becomes:

$$\text{Tr}(\text{cov}(\hat{\mathbf{x}})) = \sum_{j=1}^n (\sigma_{\beta_j} / \beta_j^{\text{nom}})^2 = \sigma^2 \sum_{k=1}^n \frac{1}{w_k^2} \quad (\text{A24})$$

In order to get a good estimation (in percent) of all the parameters of the model, the quadratic mean of the relative standard deviations of their estimates m_q should be smaller than a given level $m_{q\text{max}}$ (NB: subscript q corresponds here to the quadratic mean of the normalized standard deviations):

$$m_q = \left(\frac{1}{n} \sum_{j=1}^n (\sigma_{\beta_j} / \beta_j^{\text{nom}})^2 \right)^{1/2} = \sigma \left(\frac{1}{n} \sum_{k=1}^n \frac{1}{w_k^2} \right)^{1/2} \leq m_{q\text{max}} \quad (\text{A25})$$

One of the objectives of the "inverter" (the person in charge of the inversion) is to get a relative error m_q , expressed in term of quadratic mean, lower than an upper threshold $m_{q\text{max}}$ equal to a few percent. This means that as soon as the number n of parameters that have to be estimated becomes large, the singular values w_k of the corresponding reduced sensitivity matrix decrease, which increases the error. This increase of the error is proportional to the standard deviation of the noise. This standard deviation has the same unit as the output of the signal and the same is true for the singular values which do not depend on the structure of the model (function $\boldsymbol{\eta}$) only, but also on the intensity of the stimulation (in a problem where the output is related to a field: temperature, concentration, ...) and on the choice of the "times" of observation \mathbf{t} .

Both a lower and an upper level can also be constructed for the criterion of global relative error m_q defined in (A25), using the smaller singular value w_n :

$$\boxed{\frac{1}{\sqrt{n}} \frac{\sigma}{w_n} \leq m_q = \left(\frac{1}{n} \sum_{j=1}^n (\sigma_{\beta_j} / \beta_j^{\text{nom}})^2 \right)^{1/2} \leq \frac{\sigma}{w_n}} \quad (\text{A26})$$

This clearly shows that a too large value for the ratio σ / w_n , between the standard deviation of the measurement noise and the smaller singular value of the reduced sensitivity matrix

$\mathbf{S}^* (\boldsymbol{\beta}^{nom})$, can make the estimation of the whole set of parameters « explode ». In that case, one of the β_j parameters (the parameter "supposed to be known", β_{sk}) has to be removed from the original set of parameters to be estimated. This will lead to a new parameter vector $\boldsymbol{\beta}'$ to be estimated, of smaller dimensions ($n-1, 1$), with a better (smaller) associated m_q criterion (lower average dispersion) but with the apparition of a bias on its $n-1$ estimates, because of the biased value of the removed parameter β_{sk} that will be fixed to its nominal value that is different from its exact value (see Lecture 3).

Appendix 5 – Residual analysis for an unbiased model using the SVD approach

If the model used for estimation is unbiased, the residual vector, at convergence, is defined by:

$$\mathbf{r}(\hat{\boldsymbol{\beta}}) \equiv \mathbf{y} - \mathbf{y}_{mo}(\hat{\boldsymbol{\beta}}) = \mathbf{y}_{mo}(\boldsymbol{\beta}^{exact}) + \boldsymbol{\varepsilon} - \mathbf{y}_{mo}(\hat{\boldsymbol{\beta}}) \approx \boldsymbol{\varepsilon} - \mathbf{S}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{exact}) \text{ with } \mathbf{S} = \mathbf{S}(\hat{\boldsymbol{\beta}}) \quad (\text{A27})$$

The last approximation in equation (A27) is based on a first order development of the model with respect to parameter $\boldsymbol{\beta}$, assuming that $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^{exact}$ are close. So,

$$\mathbf{r}(\hat{\boldsymbol{\beta}}) \approx \boldsymbol{\varepsilon} - \mathbf{S}(\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T (\mathbf{y} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{exact})) = \boldsymbol{\varepsilon} - \mathbf{S}(\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T (\mathbf{y}_{mo}(\boldsymbol{\beta}^{exact}) + \boldsymbol{\varepsilon} - \mathbf{y}_{mo}(\boldsymbol{\beta}^{exact})) \quad (\text{A28})$$

The second term in equation (A28) is also a first order development that stems from the Gauss-Newton algorithm (12) used for minimizing $J_{OLS}(\boldsymbol{\beta})$ defined in (9) in an iterative way.

After simplification, equation (A28) can be rewritten using the scaled sensitivity matrix \mathbf{S}^* :

$$\mathbf{r}(\hat{\boldsymbol{\beta}}) \approx (\mathbf{I}_m - \mathbf{S}(\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T) \boldsymbol{\varepsilon} = (\mathbf{I}_m - \mathbf{S}^* (\mathbf{S}^{*T} \mathbf{S}^*)^{-1} \mathbf{S}^{*T}) \boldsymbol{\varepsilon} \text{ with } \mathbf{S}^* = \mathbf{S}(\hat{\boldsymbol{\beta}}) \text{diag}(\hat{\boldsymbol{\beta}}) \quad (\text{A29})$$

The lean SVD form (32b) (in the main body of this paper) of the scaled sensitivity matrix (see also Appendix 1) can be used then:

$$\mathbf{S}^* (\hat{\boldsymbol{\beta}}) = \mathbf{U} \mathbf{W} \mathbf{V}^T \quad (\text{A30})$$

This yield, using the orthogonality property of the right singular matrix \mathbf{V} :

$$\mathbf{r}(\hat{\boldsymbol{\beta}}) \approx (\mathbf{I}_m - \mathbf{U} \mathbf{U}^T) \boldsymbol{\varepsilon} \quad (\text{A31})$$

So, under the IID noise assumption, for an unbiased model, one can show that the expectation of the residual vector is equal to-zero:

$$\mathbf{E}(\mathbf{r}(\hat{\boldsymbol{\beta}})) \approx (\mathbf{I}_m - \mathbf{U} \mathbf{U}^T) \mathbf{E}(\boldsymbol{\varepsilon}) = \mathbf{0} \quad (\text{A32})$$

This means that if the model used for describing the experiment is appropriate, the residuals curve is centred on the $y = 0$ axis.

In order to get “unsigned” residuals, the variance-covariance matrix of the residuals should be diagonal. If the model is unbiased, this matrix is:

$$\text{cov}(\mathbf{r}(\hat{\boldsymbol{\beta}})) \approx (\mathbf{I}_m - \mathbf{U}\mathbf{U}^T)^T \text{cov}(\boldsymbol{\varepsilon}) (\mathbf{I}_m - \mathbf{U}\mathbf{U}^T) = \sigma^2 (\mathbf{I}_m - \mathbf{U}\mathbf{U}^T) = \sigma^2 \mathbf{U}_{comp} \mathbf{U}_{comp}^t \quad (\text{A33})$$

Here \mathbf{U}_{comp} is the complementary left singular vectors matrix composed of the $(m - n)$ left singular vectors, that appear in the full SVD decomposition of $\mathbf{S}^*(\hat{\boldsymbol{\beta}})$ given by equation (A1b) in *Appendix 1*:

$$\mathbf{S}^* = \mathbf{U}_0 \mathbf{W}_0 \mathbf{V} \quad \text{with} \quad \mathbf{U}_0 = [\mathbf{U} \quad \mathbf{U}_{comp}] \quad \text{where} \quad \mathbf{U}_0^T \mathbf{U}_0 = \mathbf{I}_m \quad \text{and} \quad \mathbf{W}_0 = \begin{bmatrix} \mathbf{W} \\ \mathbf{0}_{(m-n) \times n} \end{bmatrix} \quad (\text{A34})$$

In case of a square non-linear least square problems, there are as many measurements as parameters to be estimated ($m = n$) and $\mathbf{I}_m = \mathbf{U}\mathbf{U}^T$. So, in this case, the residuals (A27) are deterministic and equal to zero (\mathbf{U}_{comp} is an ‘empty’ matrix with 0 column in that degenerated case). As soon as the number m of measurements gets higher than the number n of parameters, matrix $\mathbf{U}_{comp} \mathbf{U}_{comp}^t$ becomes non-diagonal, especially if the difference $(m - n)$ is small and the residuals are correlated. However, when this difference increases, that is when the number of measurements is a lot higher than the number of parameters, the ratio n / m goes to zero and \mathbf{U}_{comp} becomes very close to \mathbf{U}_0 , which means that

$$\text{cov}(\mathbf{r}(\hat{\boldsymbol{\beta}})) \approx \sigma^2 \mathbf{U}_{comp} \mathbf{U}_{comp}^t \xrightarrow{\text{as } n/m \rightarrow 0} \sigma^2 \mathbf{U}_0 \mathbf{U}_0^T = \sigma^2 \mathbf{I}_m \quad (\text{A35})$$

This means that, strictly speaking, the residuals are correlated, even for an unbiased model but, in practice, adding more many measurement times to a given estimation interval tends to make them nearly uncorrelated. This is especially true for thermal characterization of materials or system, where the number of parameters is low (2, 3, 4, ...) and the time sampling rate high enough with respect of the length of measurement (several hundred measurements at least for modern data acquisition systems) where the asymptotic level given by (A35) is reached.