Regularized parameter estimation through iterative rescaling (PETIR): an alternative to Levenberg-Marquardt's algorithm

Maillet Denis, Stéphane André, Rémy Benjamin, Degiovanni Alain

Regularized parameter estimation through iterative rescaling (PETIR): an alternative to Levenberg-Marquardt's algorithm

D Maillet[1,2], S André[1,2], B Rémy[1,2], A Degiovanni[1,2]
[1] *Université de Lorraine, LEMTA*
[2] *CNRS, LEMTA*
Laboratoire d'Energétique et de Mécanique Théorique et Appliquée, ENSEM,
2 avenue de la Forêt de Haye - TSA60604 - 54518 Vandœuvre-lès-Nancy cedex - France
denis.maillet@univ-lorraine.fr ; *stephane.andre@univ-lorraine.fr*;
benjamin.remy@univ-lorraine.fr; *alain.degiovanni@univ-lorraine.fr*

Abstract

The Gauss Newton method of least squares minimization for non-linear parameter estimation is revisited for parameters with different physical units. A normalization of each parameter with respect to its nominal value, that is at iteration number k-1, is implemented, which leads to a linear tangent model. This model uses the sensitivity matrix composed of the scaled sensitivity coefficients. It is decomposed under a singular values form and the covariance matrix of iterate number k is calculated. When the scaled standard deviation of one parameter estimates takes a too large value, inversion of the tangent model becomes ill-posed. Regularization is made by giving the smallest singular values infinite levels, which allows keeping the total number of parameters to be estimated unchanged : this regularization leads to a better conditioned problem in the following iterations until convergence of the residuals is reached. The corrresponding algorithm is tested in the case of two very ill posed-examples. This type of estimation performs very well when compared to the Lebenverg Marquardt algorithm.

MSC classification number: 62H12

Keywords: parameter estimation, dimensional analysis, Levenberg-Marquardt, regularization, singular value decomposition

Running head: Regularized PETIR

1. Introduction

Since the introduction of the least-squares method by C.F. Gauss, many developments have allowed present researchers and engineers to use minimization algorithms for the solution of non linear inverse problems. The progress in computer technology is of course part of this story. The initial Gauss Newton (GN) iterative method was designed

for minimizing the euclidian distance between experimental data and the output of a model. This minimization is achieved by adjusting parameters at each step of an iterative process. It is efficient for weakly non linear parameter estimation problems. It is especially designed to exploit the special structure of the Hessian matrix and gradient vector appearing in the least-squares formulation. Later on, the so-called Levenberg Marquardt (LM) algorithm achieved robustness by combining the steepest gradient descent method and the quadratic convergence rate of the Gauss Newton (GN) iterate (second-order series expansion of the objective function). Different versions of this type of combination exist, one of the most interesting ones being based on Singular Value Decomposition (SVD) of the Jacobian matrix, see Gill and Murray [1].

Following such ideas, the regularization is achieved by preventing the Hessian matrix to become non positive-definite through appropriate modification of its singular values. Recently, Finsterle and Kowalsky [2] came back to this idea and pushed it forward considering a truncated singular matrix. Small eigenvalues result from a lack of sensitivity to some parameters or from a high degree of correlation between some of them. They are directly responsible for the instability or singularity of the Hessian matrix and a damping is applied to them before inversion. Tonkin and Doherty [3] followed the works of Lawson and Hanson [4] and used also truncated singular value decomposition. The strategy is slightly different as the eigenvectors corresponding to vanishingly small eigenvalues are not considered in the inversion problem (the iteration increment is not computed from the projections of the solution vector onto these eigenvectors, in order to avoid noise amplification). This results in a lowering of the dimensionality of the inverse problem which leads to the concept of super parameters. In such a strategy, the regularizing parameter is chosen in order to set the condition number (the ratio between the largest and the smallest singular values, if the Eucilidian norm is used) of the regularized matrix to a lower value than the corresponding number of the original matrix [4].

What is presented here stems from these previous works, but differs because the physical dimensions of the quantities at stake in the Parameter Estimation Problem (PEP) are considered. This leads to a dimensionless form of the model and parameter structure and of a dimensionless form of the sensitivity matrix. As a consequence, a rescaling is implemented at each iteration of the non linear least squares minimization

So, we consider here a Parameter Estimation Problem based on a single output y_{mo} at time t, the independent variable, for a model that is non linear with regard to its parameters gathered in a column vector $\boldsymbol{\alpha}$ of dimensions $(n_\alpha, 1)$. This vector is composed of n_α parameters α_j (where $j = 1$ to n_α) *with physical dimensions* :

$$y_{mo} = \mu(t; \boldsymbol{\alpha}) \tag{1}$$

The output y_{mo} of this state-space model can derive from conservation equations (the heat equation and its associate conditions, for example) and on a constitutive law (Fourier law, for example). Once m discrete measurements y_i at times t_i are available, the preceding model can be put into a column vector form:

$$y_{mo} = \mu\,(t\,;\alpha) \qquad \text{with} \qquad y = y_{mo} + \varepsilon \qquad\qquad (2a, b)$$

where the dimensions of y_{mo}, y and ε are (m, 1) and ε is a noise vector. Let us note that in this case all the coefficients of these three vectors have the same physical unit (a temperature, if y_{mo} is observed through a thermococouple measurement y).

The definition of the parameters α_j (where $j = 1$ to n_α) depends on the objective of the estimation. However their number may be too large: only a prior dimensional analysis of model (1 or 2a) can show that this is not the case through a calculation of its number of degrees of freedom. This does not require the knowledge of the nominal value $\alpha = \alpha^{nom}$ of the parameter vector since it is only based on the mathematical form of function μ. As a consequence the corresponding scaled sensitivity matrix S_α^* , see J.V. Beck et K.J. Arnold [5], composed of the n_α scaled sensitivity column vectors

$$S_{\alpha\,j}^* = \alpha_j^{nom} \left.\frac{\partial\mu\,(t\,;\alpha^{nom})}{\partial\alpha_j}\right|_{t,\alpha_k \text{ for } k \neq j} \qquad \text{calculated around a nominal value } \alpha^{nom}\text{, is not}$$

mathematicaly singular ($\det(S_\alpha^*) \neq 0$), even if the corresponding estimation problem may be ill-posed, because of the presence of errors in the measurements.

So in order to make estimation of the parameters possible a parsimonious choice of the parameters to be estimated, as well as a scaling of each of them using a dimensional analysis approach, are presented in section 2. Derivation of a tangent linear model and its declination within the framework of the singular value decomposition (SVD) of the local and scaled sensitivity matrix is made in section 3. The corresponding implementation in a specific non linear least squares minimization technique based on scaled and dimensionless local parameters is dealt with in section 4 and the derived PETIR algorithm (Parameter Estimation Through Iterative Rescaling) is presented in section 5. The notion of *target parameters*, that is the parameters that are primarily looked for by the inverter in an experimental characterization process is introduced in section 6.

Regularization of the PETIR algorithm is defined in section 7: it does not relate to Truncated Singular Value Decomposition (TSVD) usually used in function estimation techniques since the inverter can not reduce the number of parameters in a parameter estimation procedure. On the contrary, the number of singular values is kept unchanged and the smallest singular values are given an infinite level, which implies a zero amplification of the corresponding components of the noise content of the right singular modes. This regularization technique is coined ILSVD (Infinite Levels Singular Value Decomposition) in this paper.

Section 8 is devoted to the implementation of the regularized PETIR-ILSVD technique to two-parameter estimation test cases, met in characterization problems, either in heat transfer or in solid rheology. These two studies are based on synthetic noisy signals corresponding each to an observation domain and to a model whose inversion is very ill-

posed. Comparison of our inversion technique with the original Levenberg Marquardt algorithm shows that it performs very well for characterization applications with a parameter vector of relatively low dimension.

2. Conditions for a physically sound estimation

2.1 Getting a parsimonious model

The original physical model (1) involves $n' = n_\alpha + 2$ quantities with a physical dimension (n_α parameters, time t and an output variable y). A dimensional analysis, see Vaschy–Buckingham π theorem [6], allows a reduction of this number to $n < n'$ and equation (1) becomes:

$$y^*_{mo} = \mu^* \left(t^* ; \alpha^* \right) \quad \text{with} \quad t^* = \beta_2 t \text{ and } y^*_{mo} = y_{mo} / \beta_1 \tag{3}$$

where all quantities with a star superscript are dimensionless and α^* is a mathematical column vector composed of $n-2$ dimensionless parameter groups that are mathematically independent. Parameter β_1 has the same dimension as y_{mo} and β_2 is a frequency (inverse of some physical characteristic time). Of course, we have now $n < n'$.

Since the dependent (y) and independent (time t) variables are measured with instruments that deliver signals with physical units, model (3) can be put under a *parsimonious* form:

$$y_{mo} = \beta_1 \, \mu^* (\beta_2 t ; \alpha^*) = \eta (t ; \beta) \quad \text{where} \quad \beta = \begin{bmatrix} \beta_1 & \beta_2 & \alpha_1^* & \cdots & \alpha_{n-2}^* \end{bmatrix}^T \tag{4}$$

In this equation, the vector function η is the structure of the model, see E. Walter and L. Pronzato [7], and column vector β is composed of two parameters β_1 and β_2 (with physical dimension) and of $n-2$ dimensionless parameter groups, the α_j^* 's.

2.2 Getting a statistically sound estimation during iterative minimization

Once the n parameters defined, iterative non linear minimization of the square of the norm of the residuals has to be implemented:

$$J (\beta) = r^T (\beta) \, r (\beta) = \left\| r (\beta) \right\|^2 \text{ with } r (\beta) = y - y_{mo} (\beta) \text{ and } y_{mo} (\beta) = \eta (t ; \beta) \tag{5}$$

In this expression, the Euclidian norm $\left\| y - y_{mo} (\beta) \right\|$ of the residual vector r is defined in the output domain. However, if the stability in the solution of the inversion

has to be studied, one notes that the corresponding norm of the estimation error $\left\| \boldsymbol{\beta} - \boldsymbol{\beta}^{exact} \right\|$ is not really defined in the parameter (input) domain, since it depends on the physical units of β_1 and β_2. So, we propose, at iteration number k in the minimization of (5), a rescaling of parameter vector $\boldsymbol{\beta}$ to make all the n components of its new form $\boldsymbol{x}$ dimensionless. At each iteration number k, this vector, whose norm is now mathematically sound, is redefined around a local value $\boldsymbol{\beta}^{nom} = \boldsymbol{\beta}^{(k)}$ considered as 'nominal':

$$x_j = \left(\beta_j - \beta_j^{nom} \right) / \beta_j^{nom} \tag{6a}$$

Let us note that in the neighbourhood of the exact value, that is for $\boldsymbol{\beta}^{nom} \approx \boldsymbol{\beta}^{exact}$, each dimensionless parameter can be related to the logarithm of its dimensional counterpart scaled by its exact value:

$$x_j = \left(\beta_j - \beta_j^{nom} \right) / \beta_j^{nom} \approx \ln \left(1 + \left(\beta_j - \beta_j^{exact} \right) / \beta_j^{exact} \right) = \ln \left(\beta_j / \beta_j^{exact} \right) \tag{6b}$$

This procedure is implemented at each iteration, with the calculation of the new scaled sensitivity matrix which is put under its Singular Value Decomposition form [3]. This factorization enables an efficient regularization of this dimensionless PEP once the smaller singular value becomes too low with respect to the standard deviation σ_ε of the noise on signal $\boldsymbol{y}$, in the case of an independent identically distributed (i.i.d.) noise $\boldsymbol{\varepsilon}$ [8]. The performances of this procedure are tested on two different physical PEPs in section 7.

3. Tangent linear model

The differential relationship between parameter vector $\boldsymbol{\beta}$ and its dimensionless normalized form $\boldsymbol{x}$ is:

$$d\boldsymbol{x} = \boldsymbol{R}_{nom}^{-1} d\boldsymbol{\beta} \approx \boldsymbol{R}^{-1} d\boldsymbol{\beta} \quad \text{with} \quad \boldsymbol{R}_{nom} = \text{diag}\left(\boldsymbol{\beta}_{nom} \right) = \begin{bmatrix} \beta_1^{nom} & 0 & \cdots & 0 \\ 0 & \beta_2^{nom} & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta_n^{nom} \end{bmatrix} \tag{7}$$

So, model (5) is also put under its local differential tangent form:

$$d\boldsymbol{y}_{mo} = \boldsymbol{S}^* \, d\boldsymbol{x} \quad \text{with} \quad \boldsymbol{S}^* = \boldsymbol{S} \, \boldsymbol{R}_{nom} \tag{8}$$

where the scaled sensitivity matrix $\boldsymbol{S}^*$ is formed of n column-vectors, the scaled sensitivity vectors $\boldsymbol{S}_j^*$:

$$S^* = \begin{bmatrix} S_1^* & S_2^* & \cdots & S_n^* \end{bmatrix} \quad \text{with} \quad S_j^* = \left. \frac{\partial \boldsymbol{\eta}\,(t\,;\boldsymbol{\beta}^{nom})}{\partial\,(\ln\beta_j)} \right|_{t\,,\,\beta_k \text{ for } k \neq j} \tag{9}$$

The 'compact' singular value decomposition of the nominal scaled sensitivity matrix S_{nom}^* is written:

$$S_{nom}^* = U\,W\,V^T \quad \text{with} \quad \dim U = (m,\,n) \quad ; \quad \dim W = \dim V = (n,\,n) \tag{10}$$

where :   - W is the (diagonal) matrix of the singular values of S_{nom}^*
  - U gathers its first n left singular vectors
  - V is the matrix of its n right singular vectors

So, tangent model (8) becomes:

$$\mathrm{d}\boldsymbol{y}_{mo} = U\,W\,V^T\,\mathrm{d}\boldsymbol{x} \tag{11}$$

The m theoretical outputs of this model are projected onto the output space whose basis is composed by the left singular vectors (dimension $n \leq m$):

$$\mathrm{d}\boldsymbol{y}_{mo} = U\,\mathrm{d}\boldsymbol{z}_{mo} \quad \Leftrightarrow \quad \mathrm{d}\boldsymbol{z}_{mo} = U^T\,\mathrm{d}\boldsymbol{y}_{mo} \tag{12}$$

Column vector $\boldsymbol{z}_{mo}$ can be called the 'diagonal output'. A change of basis is implemented next in the space of the scaled parameters (dimension n) :

$$\mathrm{d}\boldsymbol{x} = V\,\mathrm{d}\boldsymbol{p} \tag{13a}$$

Integration of this equation between nominal and current states yields:

$$\boldsymbol{x} = V\,\boldsymbol{p} \tag{13b}$$

because equation (6) shows that $\boldsymbol{x}^{nom} = \boldsymbol{0}$ and, as a consequence, $\boldsymbol{p}^{nom} = \boldsymbol{0}$.

Column vector $\boldsymbol{p}$ can be called the 'diagonal parameter vector'. Then tangent model (8) is written in these two new bases, using equations (11), (12) and (13):

$$\mathrm{d}\boldsymbol{z}_{mo} = W\,\mathrm{d}\boldsymbol{p} \quad \text{because} \quad U^T U = V^T V = I_n \tag{14}$$

where I_n designates the identity matrix of size n. A first order-approximation of differential model (14) around its nominal value yields :

$$\boldsymbol{z}_{mo}\,(\boldsymbol{x}) - \boldsymbol{z}_{mo}^{\,nom} \approx W\,(\boldsymbol{p} - \boldsymbol{p}^{nom}) \quad \text{and hence:} \quad \boldsymbol{z}_{mo}\,(\boldsymbol{x}) \approx W\,\boldsymbol{p} \tag{15a, b}$$

Integration of equations (12) yields:

$$\boldsymbol{y}_{mo}(\boldsymbol{\beta}) - \boldsymbol{y}_{mo}(\boldsymbol{\beta}^{nom}) = \boldsymbol{U} \ \boldsymbol{z}_{mo}(\boldsymbol{x}) \tag{15c}$$

Let us note that vectors $\boldsymbol{\beta}$, $\boldsymbol{x}$ and $\boldsymbol{p}$ are just dummy variables in equations (13b), (15b) and (15c).

4. Linear least squares using measurements and tangent model

Projection of the measurement (output) vector $\boldsymbol{y} = \boldsymbol{y}_{mo}(\boldsymbol{\beta}^{exact}) + \boldsymbol{\varepsilon}$ into the left singular basis, in the same way as (15c), yields:

$$\boldsymbol{y} - \boldsymbol{y}_{mo}(\boldsymbol{\beta}^{nom}) = \boldsymbol{U} \ \boldsymbol{z} \tag{16a}$$

where $\boldsymbol{z}$ is the projected measurement vector:

$$\boldsymbol{z} = \boldsymbol{z}_{mo} + \boldsymbol{U}^{T}\boldsymbol{\varepsilon} \tag{16b}$$

The residual vector is linearized at each step, that is for any value of parameter vector $\boldsymbol{\beta}$ around its nominal value $\boldsymbol{\beta}^{nom}$, taking equations (13b), (15b), (15c) and (16a) into account:

$$\boldsymbol{r}(\boldsymbol{\beta}) = \boldsymbol{y} - \boldsymbol{y}_{mo}(\boldsymbol{\beta}) \approx \boldsymbol{r}_{lin}(\boldsymbol{\beta}) \quad \text{with} \quad \boldsymbol{r}_{lin}(\boldsymbol{\beta}) \equiv \boldsymbol{U}\,(\boldsymbol{z} - \ \boldsymbol{W}\boldsymbol{V}^{T}\boldsymbol{x}) \tag{17}$$

So the least squares criterion (5) is rewritten using the linearized residuals:

$$J(\boldsymbol{\beta}) \approx \|\boldsymbol{r}_{lin}(\boldsymbol{\beta})\|^{2} = \|\boldsymbol{z} - \boldsymbol{A}\boldsymbol{x}\|^{2} \ \text{with} \ \ \boldsymbol{A} = \boldsymbol{W} \ \boldsymbol{V}^{T} \tag{18a}$$

The least-square solution of (18a), that is the estimate $\hat{\boldsymbol{x}}$ of $\boldsymbol{x}$ is explicit and can be written thanks to equation (16a):

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{A}^{T}\boldsymbol{A}\right)^{-1}\boldsymbol{A} \ \boldsymbol{z} = \boldsymbol{V} \ \boldsymbol{W}^{-1}\boldsymbol{z} = \boldsymbol{V} \ \boldsymbol{W}^{-1}\boldsymbol{U}^{T}(\boldsymbol{y} - \boldsymbol{y}_{mo}(\boldsymbol{\beta}^{nom})) \tag{18b}$$

Return into the initial parameter domain is implemented next:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{nom} + \boldsymbol{R}^{nom}\hat{\boldsymbol{x}} \tag{19}$$

The variance-covariance matrix of the scaled parameters $\hat{x}$ can be easily calculated for a i.i.d. noise ε, that is for $\text{cov}(\varepsilon) = \sigma_\varepsilon^2 \, I_m$:

$$\text{cov}(\hat{x}) = \sigma_\varepsilon^2 \, V \, W^{-2} \, V^T \tag{20}$$

The estimation is unbiased for the tangent model, that is $E(\hat{\beta}) = \beta^{exact}$, where $E(.)$ designates the expectation of a random variable.

5. Construction of the PETIR algorithm in the parametric domain of dimension n

One chooses an initial value for the PETIR (Parameter Estimation Through Iterative Renormalization) algorithm:

$$\hat{\beta}^{(0)} = \beta^{nom} \implies \hat{x}^{(0)} = 1 \tag{21}$$

Parameter vector x is redefined at each iteration using tangent model (18b) and (19). At iteration $(k - 1)$, $\hat{\beta}^{(k-1)}$ has been calculated and one writes:

$$\hat{\beta}^{(k)} = \hat{\beta}^{(k-1)} + R^{(k-1)} \hat{x}^{(k)} \qquad \text{with} \qquad R^{(k-1)} = \text{diag}(\hat{\beta}^{(k-1)}) \tag{22}$$

and $\quad \hat{x}^{(k)} = V^{(k-1)} \hat{p}^{(k)}$; $\hat{p}^{(k)} = \left(W^{(k-1)}\right)^{-1} z^{(k)}$; $z^{(k)} = \left(U^{(k-1)}\right)^T \left(y - y_{mo}(\hat{\beta}^{(k-1)})\right)$

$$\tag{23a, b, c}$$

and $\qquad\qquad S^*(\hat{\beta}^{(k-1)}) = U^{(k-1)} W^{(k-1)} \left(V^{(k-1)}\right)^T \tag{23d}$

One shows, using equations (23):

$$\hat{\beta}^{(k)} = \hat{\beta}^{(k-1)} + R^{(k-1)} V^{(k-1)} \left(W^{(k-1)}\right)^{-1} \left(U^{(k-1)}\right)^T \left(y - y_{mo}(\hat{\beta}^{(k-1)})\right) \tag{24}$$

The pseudo-inverse, at iteration $(k-1)$, of S^*, $S^{*+} = (S^{*T} S^*)^{-1} S^{*T} = V W^{-1} U^T$ is of course present in equation (24).

6. Definition of target parameter(s)

The objectives of the person in charge of the estimation of parameters of a non linear model, once noisy measurement of its output are available, can be very diverse:

- a first objective can be the rational discrimination between different candidates for a model of a constitutive law. Models of different mathematical structures can be

developed for the same experiment: this structure either depends on the choice made for the solution technique of a differential balance or can imply different numbers of parameters or degrees of freedom, according to the assumptions made related to various phenomena. The parsimony principle [9], which states that the best model is the one with the lowest number of parameters, constitutes a rational guideline for a PEP, and will be followed using the regularization technique presented further on.

- another objective can be the identification of the system using a given structure for the model. In this case all the parameters, a priori, are looked for with an equal precision.

- a last objective, which is the only one considered below, can be the indirect measurement of one parameter of a model, for example the thermal diffusity of a material in a flash experiment, see A. Degiovanni [10] and B. Hay et al. [11], the remaining ones being either 'unwished' or 'nuisance' parameters. This kind of parameter can be called a 'target parameter' and the inverter-experimenter wants to minimize the relative error of this specific parameter. So, not all the n parameters β_j of $\boldsymbol{\beta}$ are looked for, but only one of its subset. Let us assume that it is parameter number j_1 that is the target one. One can show using (20):

$$\text{cov} (\hat{\boldsymbol{x}}^{(k)}) = \sigma^2 \boldsymbol{V}^{(k-1)} \left(\boldsymbol{W}^{(k-1)}\right)^{-2} \left(\boldsymbol{V}^{(k-1)}\right)^{T} \tag{25}$$

and the relative (scaled) standard deviation of this target parameter is:

$$\frac{\sigma(\hat{\beta}_{j_1}^{(k)})}{\hat{\beta}_{j_1}^{(k)}} = \left(\left[\text{cov} (\hat{\boldsymbol{x}}^{(k)})\right]_{j_1 \ j_1}\right)^{1/2} \tag{26}$$

So, during iterative rescaling, the following criteria have to be checked at each iteration:

i)   the residuals should decrease until they reach a minimum level not lower than the measurement noise. This is the Morozov's discrepancy principle [12] : $J(\hat{\boldsymbol{\beta}}^{(k)}) \geq m \sigma^2$,

ii) $\sigma(\hat{\beta}_{j_1}^{(k)})/\hat{\beta}_{j_1}^{(k)}$ should also be calculated, in order to assess the precision of the estimation and

iii)   a stopping criterion should be given for the number of iterations : $k \leq k_{max}$.

If $\sigma(\hat{\beta}_{j_1}^{(k)})/\hat{\beta}_{j_1}^{(k)}$ is too large, a regularization of $\boldsymbol{S}^*$ has to be made:

a)   either by modifying at least one of its singular values,

b)   or by locking one of the parameters, noted β_{j_0} here, to either its current value $\hat{\beta}_{j_0}^{(k-1)}$ or to its nominal value $\beta_{j_0}^{nom}$, that is a value 'supposed to be known'.

Of course, in both cases the bias on the residuals as well as on $\hat{\beta}_{j_1}^{(k)}$, together with the standard deviation of this parameter estimate, have to be studied in the following iterations. Regularization of type **b)** is difficult to implement because there are multiple ways to define the n parameters present in $\boldsymbol{\beta}$ so, we focus here on case **a)** only.

7. PETIR Regularization by making at least one singular value infinite (ILSVD) for a non linear parameter estimation problem

Truncated singular value decomposition (TSVD) can be used for solving a linear function estimation problem, that is, for example, the solution of the following integral equation :

$$y_{mo} \ (t) = \int_0^a K \ (t, \ t') \ u \ (t') \ dt' \tag{27}$$

where function $u \ (t)$ (an input) is looked for starting from measured values of its noisy output y at m times t_i (see equation 2b). If $a = t_f$, where t_f is the final time of measurement, equation (27) is a Fredholm integral equation of the first kind, while $a = t$ corresponds to a Volterra integral equation of the first kind.

In order to write this problem in a finite-dimensional space, function $u \ (t)$ is parameterized into a finite vector $\boldsymbol{\beta}$ of arbitrary dimension $n \leq m$ and its n components correspond to a projection of y_{mo} over a set of n functions $f_j \ (t)$ (for $j = 1, 2, ..., n$) that forms a basis of the parameter space:

$$u \ (t) \approx u_{param}(t \ ; n, \ \boldsymbol{\beta}) = \sum_{i=1}^n \beta_j \ f_j \ (t) \ \text{ and } \ S_{ij} = u_j \ (t_i) \tag{28}$$

Of course, the norm of the difference between the original function $u \ (t)$ and its parameterized form $u_{param}(t, \ \boldsymbol{\beta})$ decreases with n. Substitution of (27) into (28) yields:

$$\boldsymbol{y}_{mo} = \boldsymbol{\eta}(t, \ \boldsymbol{\beta}) = \boldsymbol{S} \ \boldsymbol{\beta} \ \text{ with } \ S_{ij} = \int_0^{a_i} K \ (t_i, \ t') \ f_j \ (t') \ dt' \tag{29a, b}$$

where $a_i = t_f$ (Fredholm equation) or $a_i = t_i$ (Volterra equation). So solution of the corresponding discrete linear inverse problem consists in replacing the exact but unstable ordinary least square solution:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{S}^+ \ \boldsymbol{y} = (\boldsymbol{S}^T \ \boldsymbol{S})^{-1} \ \boldsymbol{S}^T \ \boldsymbol{y} = \boldsymbol{V} \ \boldsymbol{W}^{-1} \boldsymbol{U}^T \ \boldsymbol{y} \tag{30}$$

where S^+ is the pseudo inverse of the sensitivity matrix S (which is supoposed to be of full rank) and (U,W,V) its compact SVD deconvolution form already defined in (10) (replacing S^*_{nom} by S) by its truncated SVD (TSVD) form, see R.J. Hanson [13] and P.C. Hansen [14] :

$$\hat{\beta}_{trunc} = V_{trunc} \, W^{-1}_{trunc} \, U^T_{trunc} \, y \tag{31}$$

where only the $u < n$ first columns of (U,W,V), corresponding to the largest singular values, have been kept in $(U_{trunc}, W_{trunc}, V_{trunc})$ respectively, as well as only the u first lines of the square diagonal matrix W_{trunc}. Of course, since there are less many coefficients in $\hat{\beta}_{trunc}$ than in $\hat{\beta}$, the parameterization (28) will be made with fewer functions (u is the regularization hyperparameter) :

$$\hat{u}(t) = \hat{u}_{param}(t \; ; s, \boldsymbol{\beta}) = \sum_{i=1}^{u<n} \hat{\beta}_{trunc\, j} \; f_j(t) \tag{32}$$

It is important to note here that all the β_j parameters in the estimation have the same physical dimensions. Hence the reduction of their number does not cause any practical problem.

Such is not the case in the regularization of the non linear estimation problem of type **a** (see section 5 above) since the number n of parameters cannot be decreased and no column or line can be removed in the matrices (U,W,V) of the SVD decomposition of the local scaled sensitivity matrix given by equation (10). So, once regularization of type **a** has been applied in the iterative estimation given by equation (24), only the diagonal matrix W based on the singular values, called w_j here (with $j = 1$ to n, with S^* of full rank) is modified:

$$\hat{\beta}^{(k)}_{reg} = \hat{\beta}^{(k-1)} + R^{(k-1)} V^{(k-1)} \left(W^{(k-1)}_{reg} \right)^{-1} \left(U^{(k-1)} \right)^T \left(y - y_{mo}(\hat{\beta}^{(k-1)}) \right) \tag{33a}$$

where $\left(W^{(k-1)}_{reg} \right)^{-1}$ is a diagonal matrix whose diagonal is:

$$\mathrm{diag}\left(\left(W^{(k-1)}_{reg} \right)^{-1} \right) = \left[1/w_1^{(k-1)} \quad \cdots \quad 1/w_u^{(k-1)} \quad 0 \quad \cdots \quad 0 \right] \tag{33b}$$

This means that only the first u singular values of $W^{(k-1)}$ are kept ($u < n$), while the $(n - u)$ following ones are given an infinite level in $W^{(k-1)}_{reg}$ and consequently a zero value in its inverse $\left(W^{(k-1)}_{reg} \right)^{-1}$ given by (33b) while, contrary to the linear function estimation case (31), the matrices of the left and right singular values are kept unchanged in the regularized estimation process given by (33a).

This kind of regularization using infinite levels in the singular value decomposition (ILSVD) of the scaled sensitvity matrix can be considered as an alternative to TSVD to parameter estimation problems. Contrary to TSVD, the number of estimated parameters does not need to be reduced in ILSVD regularization. We will use the notation ILSVDu now on for designing a regularization where u singular values are kept, while the $n - u$ smaller ones have been given an infinite level.

A similar regularization path is used in the Levenberg Marquardt minimization technique, see P.E. Gill and W. Murray [1], which implies larger and larger - but not infinite - arbitrary penalty weighting on singular values in order to ensure decreasing residuals. Unfortunately, in pathological ill-posed problems such the ones studied in the next section, this algorithm fails in producing good estimates.

8. Application to two non linear physical examples

We propose to test the PETIR algorithm with ILSVD regularization by applying it to two simple models stemming from two different scientific fields. The first one is a model used in thermal characterization of a low-weight insulating material whose thermal conductivity is looked for. The second one is a phenomenological model used to describe the mechanical behavior of polymers in a tensile test.
Parameter estimation tests will be performed starting from perfectly known simulated data, with added synthetic noise, using three type of algorithms:

- the classical Gauss-Newton (GN) method (without any regularization),

- Levenberg-Marquardt (LM) algorithm [**1**] based on SVD formulation of the sensitivity matrix with random weighting on singular values (integrated in a Matlab® leasqr.m file). This algorithm is very robust and has been used by the authors during several years. It works with a non-scaled sensitivity matrix,

- the present scaled-sensitivity approach (PETIR, that is iterative rescaling) coupled with SVD formulation with regularization through infinite level values (ILSVD).

8.1 A characterization model in heat transfer

The following simple model is considered :

$$\theta\,(t, \boldsymbol{\beta}) = \beta_1\left[\exp\left(-t\,/\,\beta_2\right) - \exp\left(-\beta_3\,t\,/\,\beta_2\right)\right] \tag{34}$$

As explained in [15, page 317], this model corresponds to the 1D rear-face temperature response θ (kelvin) to a flash stimulation of a three-layer material sample composed of a low-weight insulating material (no capacity, thickness e_i and conductivity λ_i , thermal resistance $R = e_i\,/\lambda_i$ for a unit area) which is sandwiched between two identical copper layers (thickness e_c, volumetric heat capacity ρc_c , heat capacity C_c for a unit area). Heat losses (convection and linearized radiation) are present and are taken into account in the

model by an identical heat transfer coefficient h on both faces. The three parameters in the $\boldsymbol{\beta}$ vector are:

$$\beta_1 = \frac{Q}{2C_c} \;\; ; \;\; \beta_2 = \frac{C_c}{h} \;\; ; \;\; \beta_3 = 1 + \frac{2}{R\,h} \tag{35}$$

We consider the following 'exact values' of these parameters:

$$Q = 10^4 \text{ J/m}^2 \; ; \; C_c = \rho c_c e_c = (3.6 \; 10^6)\,(0.6 \; 10^{-3}) = 2160\,\text{J.m}^{-2}.\text{K}^{-1}$$
$$R_i = e_i / \lambda_i = 6.0\; 10^{-3} / 0.02 = 0.3 \text{ m}^2.\text{K/W} \;\; ; \;\; h = 10 \text{ W.m}^{-2}.\text{K}^{-1}$$

Physically speaking, β_1 (kelvin) is the adiabatic temperature reached for long times in the absence of any heat loss (case $h = 0$), and β_2 (second) is a time constant. It is quite obvious that a high value of the target parameter, thermal resistance R, will makes the two exponentials very close in model (34), making the estimation problem very ill-posed.

The exact values are:

$$\boldsymbol{\beta}^{exact} = \left[\beta_1^{exact} \quad \beta_2^{exact} \quad \beta_3^{exact} \right]^T = \left[2.3148 \quad 2160 \quad 1.6667 \right]^T$$

The purpose of this experiment is to obtain the thermal conductivity of the material through the estimation of the β_j's by an inverse technique applied on the transient measurements of θ. So, β_2 and β_3 must be estimated independently. This estimation is possible if the time interval is well chosen, at least after the time of occurrence of the maximum of the thermogram. To test the robustness of our method, we propose to work on a shorter interval (80% of the time of the maximum). Furthermore, the initial values necessary to start the iterations are chosen far enough from the exact values, respectively +40%, -80% et +80% for each parameter β_j :

$$\boldsymbol{\beta}^{init} = \left[3.2407 \quad 43.20 \quad 3.0 \right]^T$$

Figure 1 shows the theoretical $y_{mo} = \theta$ thermogram as well as the three scaled sensitivity functions S_j^* to these three parameters, for the exact values of the parameters. One can note that the three sensitivity coefficients seem to be two by two proportional.

The RSdCor (Relative Standard deviation - Correlation) matrix is a composite matrix derived from equations (25) and (26) made of the scaled standard deviations of the parameters in the main diagonal :

$$\sigma_{\hat{\beta}_j}/\beta_j^{exact} = \left(\operatorname{var}(\hat{\beta}_j)\right)^{1/2}/\beta_j^{exact} = \left(\operatorname{var}(\hat{x}_j)\right)^{1/2} = \sigma_\varepsilon \left(\left[\left(S^{*T}S\right)^{-1}\right]_{jj}\right)^{1/2}$$

(36a)

and of the correlation coefficients (scaled covariance between parameters) for the off-diagonal terms :

$$\rho_{ij} = \operatorname{cov}(\hat{\beta}_i, \hat{\beta}_j)/\sqrt{\operatorname{var}(\hat{\beta}_i)\operatorname{var}(\hat{\beta}_j)}$$

(36b)

All these coefficients are calculated with the exact values of the parameters and for the following values of the standard deviation of the noise $\sigma_\varepsilon = 0.0429\,\text{K}$. This corresponds to a quite low signal to noise ratio (SNR) $\theta_{max}/\sigma_\varepsilon$ close to 10, where $\theta_{max} = 0.4209\text{K}$ is the maximum temperature reached (it occurs past the last time of measurement and is not shown in figure 1).

The relative standard deviations for each parameter are large, respectively 627%, 192% and 329%, see figure 1, which confirms the quasi proportionality between sensitivity coefficients noticed above, and all the correlation coefficients are close to one, which means that some correlation exists between the parameters of the problem. This clearly shows that the original parameter estimation problem is ill-paused.
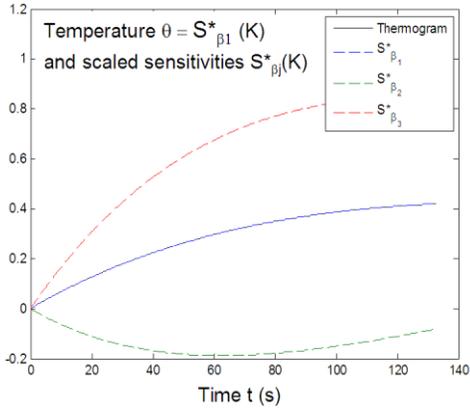
Fig. 1 - Thermogram and Sensitivity Curves / Relative Sytandard deviation - Correlation matrix

8.1.1. Estimation by PETIR-ILSVD for a noiseless signal, using reduced and non-reduced sensitivity matrices

Estimation is first performed for a thermogram without noise, with $m = 1000$ points in time for the estimation interval. Figure 2 shows the exact temperature response, the recalculated response, once convergence has been reached for the least square minimization, as well as the corresponding residuals for the PETIR-ILSVD2 estimation (a) and for the Levenberg-Marquardt (LM) estimation (b).

The corresponding relative variation of the estimation error for each of the three parameters, as a function of the iteration number k, is given for both PETIR-ILSVD2 and LM techniques in Figure 3a. It shows that the proposed method (PETIR and ILSVD with 1 infinite value, that is $u = 2$) yields lower residuals than the Levenberg-Marquardt (LM) method [1] with a lower number of iterations (8 iterations for PETIR compared to 37 for LM) and with a better accuracy since the relative errors $(\hat{\beta}_j - \beta_j^{exact})/\beta_j^{exact}$ values at convergence are: 3.5%, 1% and 1.8% for PETIR against 45%, 31% and 55% for LM.
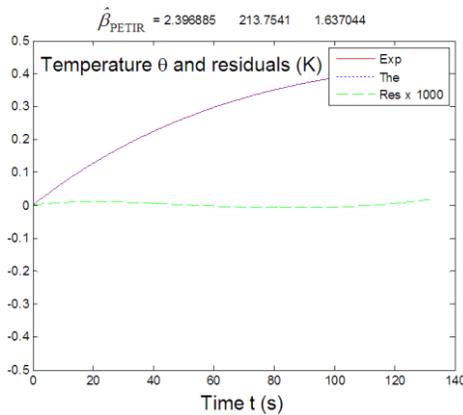
(a) PETIR-ILSVD2 (u = 2 singular values kept)   (b) LM
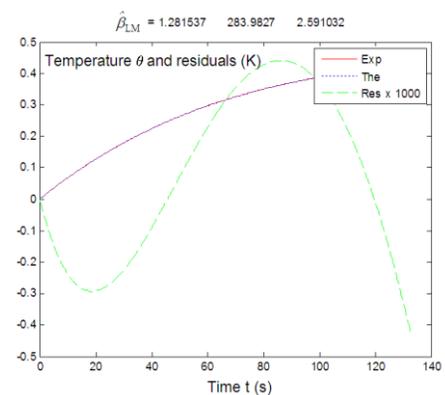
Fig. 2- Exact and reconstructed thermograms and Residuals curves (Exp = model output without any noise, The: reconstructed signal after convergence of estimates, Res = residuals)
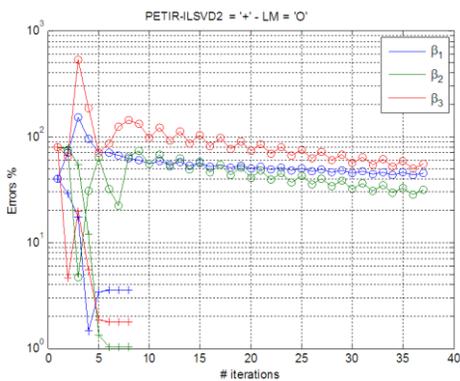
(a) Case of a scaled sensitivity matrix for   (b) Case of a non-scaled sensitiity matrix for
PETIR-ILSV2 and LM   ILSVD2 and LM

Fig. 3 - Relative error of the estimates (%) versus number of iterations, estimation from a signal without any noise.

Remark : An alternate way of using the ILSVD technique, is to apply it on the SVD decomposition of the non-scaled sensitivity matrix S. This consists in using a modified version of equations (33):

$$\hat{\boldsymbol{\beta}}_{reg-dim}^{(k)} = \hat{\boldsymbol{\beta}}^{(k-1)} + \boldsymbol{V}'^{(k-1)}\left(\boldsymbol{W}_{reg}'^{(k-1)}\right)^{-1}\left(\boldsymbol{U}'^{(k-1)}\right)^{T}\left(\boldsymbol{y} - \boldsymbol{y}_{mo}(\hat{\boldsymbol{\beta}}^{(k-1)})\right) \tag{37a}$$

$$\mathrm{diag}\left(\left(\boldsymbol{W}_{reg}'^{(k-1)}\right)^{-1}\right) = \begin{bmatrix} 1/w_1'^{(k-1)} & \cdots & 1/w_u'^{(k-1)} & 0 & \cdots & 0 \end{bmatrix} \tag{37b}$$

with:

$$S = \begin{bmatrix} S_1 & S_2 & \cdots & S_n \end{bmatrix} \quad \text{with} \quad S_j = \left.\frac{\partial \boldsymbol{\eta}(t;\boldsymbol{\beta}^{nom})}{\partial(\beta_j)}\right|_{t,\beta_k \text{ for } k \neq j} \quad \text{and} \quad S = \boldsymbol{U}'\,\boldsymbol{W}'\,\boldsymbol{V}'^{T} \tag{37c}$$

This type of inversion technique does not use rescaling of the PETIR type and is just an ILSVD regularized version of the Gauss-Newton algorithm.

As expected, if the non-scaled sensitivity curves are used for the estimation, the results are worse, see figure 3b, than for the scaled PETIR case, see figure 3a. The errors on the estimated parameters are 63%, 95% and 174% and the number of iterations until convergence (18 iterations) increases too for the ILSVD2 (but not PETIR) technique, see figure. 3b.

8.1.2. Estimation by PETIR-ILSVD for a thermogram corrupted by noise and effect of number of infinite levels

If an additive noise is added to the exact thermogram of figure 1 (i.i.d. noise of standard deviation $\sigma_\varepsilon = 0.0429\,\mathrm{K}$, that is a signal/noise ratio close to 10, see section 7.1), one gets a synthetic noisy signal. Application of the two estimation techniques to this signal yields a convergence which is achieved in 8 iterations by PETIR-ILSV2 and 38 for LM (see figures 4 c). The relative errors are equal to 3.3%, 1.3% and 1.8% for PETIR-ILSV2 and 25%, 10% and 19% for LM. The corresponding residuals are shown in figures 4a (PETIR-ILSVD2) and 4b (LM). So, this rescaling/ILSVD method is very precise here even with a high level of noise and an ill-conditioned scaled sensitivity matrix.

Finally, some tests have been performed by keeping only the largest singular value of S^* (2 infinite values, that is ILSVD1) and the corresponding variation of the estimation error of the three parameters with the iteration number are shown in figure 4d and compared to the LM case. The estimation errors with a single singular value kept ($u = 1$) are higher than in the previous case ($u = 2$), which suggests that our problem exhibits two and not one degrees of freedom.
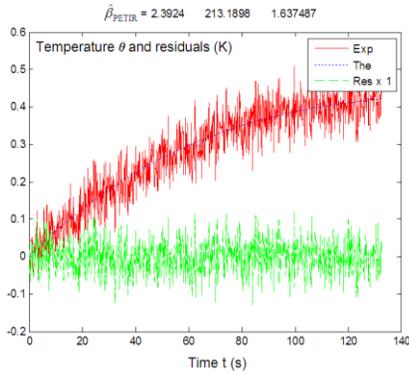
(a) Synthetic noisy signal, reconstructed signal and residuals - PETIR-ILSVD2
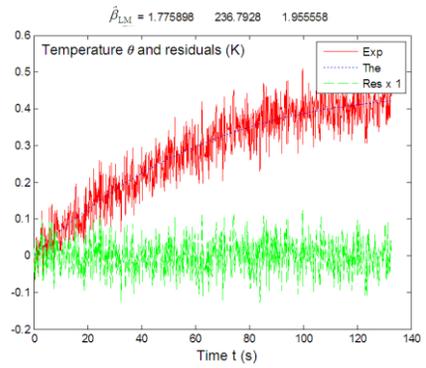
(b) Synthetic noisy signal, reconstructed signal and residuals - LM
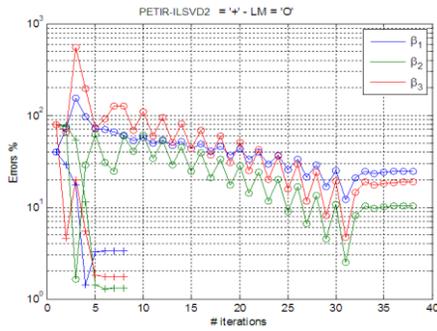
(c) Estimation errors for PETIR-ILSVD2 and LMinversions

(d) Estimation errors for PETIR-ILSVD1and LM inversions

Fig. 4 - Effect of number of singular values kept on relative estimation errors (%) and comparison between PETIR-ILSVD and Levenberg Marquardt, estimation from a noisy signal

8.2 A characterization model in solid rheology

In solid rheology, characterization of materials depends on a pertinent identification of their constitutive law. We consider here the one-directional stress (σ) − strain (ε) relationship for a tensile test. In such a test, the structure of 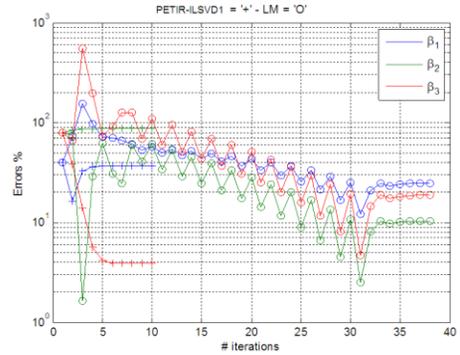the model is generally not known. As a consequence, optimal parameter estimation techniques should be able to discriminate between competing models through a correct estimation of its parameters and of the uncertainties (bias and standard deviations) associated to it.

In the case of semi-crystalline polymers, the mechanical model must take into account multiple phenomena (elasto-visco-plasticity as well as hardening or damaging effects). This is a difficult problem which has often led to derive more or less phenomenological models such as the G'sell-Jonas stress-strain relationship [16]. It has the following form for an experiment where the strain rate is kept at a constant level:

$$\sigma = k \left[1 - \exp(-W\varepsilon) \right] \left[1 + L \exp(-V\varepsilon) \right] \exp(h\varepsilon^p) \qquad (37a)$$

The exact values of the different parameters are introduced (see Table 1 further on):

$$\boldsymbol{\beta}^{exact} = \left[k^{exact} \quad W^{exact} \quad L^{exact} \quad V^{exact} \quad h^{exact} \quad p^{exact} \right]^T \qquad (37b)$$

This fitting model has no physical basis and relies on a parameter vector $\boldsymbol{\beta}$ of dimension 6, 5 parameters being dimensionless and the dimension of the last one, k, being the same as σ (expressed in Pascal). Relation (37a) is hence written as the generic starting relation (4).

Parameter estimation with relation (37a) provides a second good academic test-case to illustrate how the PETIR algorithm works. The ill-conditioning of this over-parameterized model is pathological here, since combining multiple exponentials allows to 'connect the dots' or mimic any curve, especially in the present case where a simple monotonous curve has to be fitted (see the black dotted line in figure 5).

Besides, a theoretical model having a different structure and requiring only 3 parameters has been proved to produce the exact same output, see S. André et al. [17].

As in the preceding example of section 7.1, the exact values gathered in the dimensional $\boldsymbol{\beta}^{exact}$ vector, see Table 1, have been used to generate the data using model (37) and an i.i.d. noise of standard deviation $\sigma_\varepsilon = 3\text{MPa}$ has been added to generate the synthetic stress signal $y_{mo} = \sigma$ in figure 5. This pseudo-experimental signal is plotted versus the dependent variable, here the strain $t = \varepsilon$, to use the generic notation already presented in sections 1 and 2.

Let us note that the signal to noise ratio has the same order of magnitude, equal to 10 here, as in the case of 7.1. Since the signal is monotonically increasing (no absolute maximum), the 'maximum' amplitude of the signal is replaced by the plateau value (30 MPa) of the curve of Fig.5 (in the range of 0.2-0.4 strain) in its numerator. This plasticity plateau corresponds to the yield stress where the material flows without any need for an additional force. Its denominator is the previous standard deviation σ_ε of the added noise.

Obviously, we have $\sigma = S_k^*$ here. The RSdCor (Relative Standard deviation - Correlation) matrix also calculated at this point, and valid for a non-regularized Gauss Newton least-squares minimization, clearly shows that the original parameter estimation problem is ill-posed:

$$\text{RSdcor}\,(\hat{\boldsymbol{\beta}}_{GN}) = \begin{bmatrix} 0.02\,\% & & & & & \\ -0.506 & 1.10\,\% & & & 0 & \\ 0.422 & -0.887 & 21.7\,\% & & & \\ 0.854 & -0.756 & 0.785 & 48.0\,\% & & \\ -0.986 & 0.472 & -0.379 & -0.816 & 0.06\,\% & \\ 0.968 & -0.450 & 0.354 & 0.788 & -0.996 & 0.01\,\% \end{bmatrix} \tag{38}$$
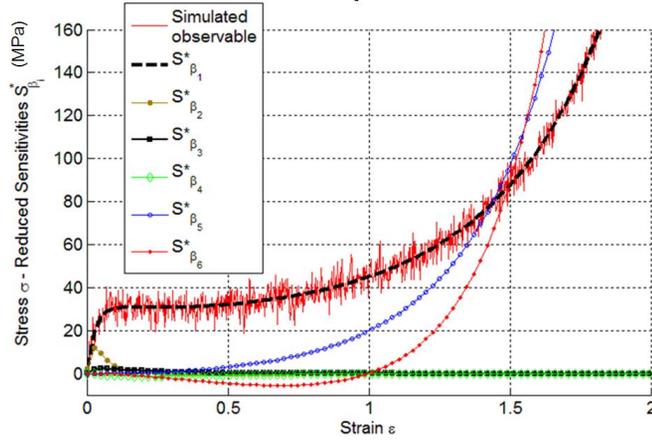
Fig. 5 - Exact and noisy output signals and scaled sensitivities for an exact value of $\boldsymbol{\beta}$

The RSdCor matrix illustrates that at the exact point, parameters 3 and 4 can be obtained only with a high relative error (21.7% and 48% respectively). This is caused by reduced sensitivities close to zero for these two parameters in the whole range of the independent variable ε, see figure 5. This PEP is strongly non linear and one can easily imagine that other nominal points would imply pathological features (non uniqueness of the solution) which would result in a severe failure for any non-regularized estimation algorithm.

The application of the algorithm will be made from initial guesses chosen as $\beta_j^{init} = \beta_j^{exact} \pm C\,\beta_j^{exact}$, where the following values are considered for C : 0.05, 0.15, 0.25 (for all parameters), the sign being randomly assigned to each of the 6 parameters.

Results of inversion of a noiseless signal are presented in Tables 1 and 2, for two different values of C. The value of the square root of the least squares criterion (5), $\sqrt{J(\hat{\boldsymbol{\beta}})}$ at convergence is given in these tables where k_{max} denotes the number of iterations required for convergence.

Table 1- Noiseless signal, C=0.05. Lines 5 to 8 give the relative error in % (with respect to exact values)

		β_1	β_2	β_3	β_4	β_5	β_6	$\sqrt{J(\hat{\boldsymbol{\beta}})}$ (k_{max})
	β_j^{exact}	K=0.29 MPa	$W=35.3$	$L=0.12$	$V=5$	$h=0.4$	$p=2.27$	
1	β_j^{init}	27.55	37.065	0.114	5.25	0.418	2.3835	
2	$\hat{\beta}_j^{GN}$	29	35.3	0.12	5	0.4	2.27	10^{-12} (6)
3	$\hat{\beta}_{reg\,j}^{LM}$	29	35.3	0.12	5	0.4	2.27	10^{-12} (9)
4	$\hat{\beta}_{dim\,j}^{ILSVD5}$	29	35.27	0.1203	5.02	0.44	2.27	0.02 (5)
5	error $\hat{\beta}_{dim\,j}^{ILSVD5}$	0 %	0.06%	0.28%	0.46%	0.01%	0 %	0.7 (6)
6	error $\hat{\beta}_{dim\,j}^{ILSVD4}$	<0.5%	1%	5.8%	5.1%	<0.5%	<0.5%	0.8 (4)
7	error $\hat{\beta}_{dim\,j}^{ILSVD3}$	<0.5%	4.7 %	5.2%	5.2%	<0.5%	<0.5%	2 (5)
8	error $\hat{\beta}_{dim\,j}^{ILSVD2}$	1.6%	5.0%	5.0%	5.0%	3.3%	1.8%	9.2 (4)

8.2.1. Some general results in the case of a non noise-corrupted signal

The case of an infinite signal to noise ratio is considered first. In this case, the estimation error have no stochastic component and can be considered as a pure bias for the estimation technique used. The following interesting features can be observed:

- When the initial parameter vector β_{init} is not too far from the exact values (C=0.05, Table1), the classical non-regularized Gauss Newton (GN) method, based on either a dimensional or a scaled-reduced sensitivity, converges to the exact parameter values with the same precision as the Levenberg-Marquardt (LM) regularized algorithm, see lines 2 and 3. If regularization is applied to the GN method using ILSVD5 applied on the non scaled (dimensional) sensity matrix (replacement of the lowest singular value by an infinite level), errors of bias type are generated on the estimated parameters, see lines 5 to 8 for columns β_2, β_3, β_4. This illustrates the effect of such a reduction. In this case where a dimensional sensitivity matrix is used, this deterioration can be very important, which will argue in favour of using a rescaled-sensitivity approach.

- When the number of infinite levels increases (up to 4 over a total of 6 parameters), the bias increases (lines 5 to 8). Parameters β_2, β_3 and β_4 are clearly those which cannot be estimated through inversion of measurements. It can also be noted that an appropriate regularization is obtained even when only $u=3$ or even 2 singular values are kept unchanged: down to three parameters kept, the estimation bias increases for

the three "bad" parameters β_2, β_3 and β_4 but the other ones remain properly estimated. When only 2 singular values are kept unchanged, a bias also begins to affect parameters β_1, β_5 and β_6. This illustrates that the present ILSVD method constitutes a tool for the determination of the exact number of degree of freedom of a model. So, aside the technical aspects associated with the regularization of the PETIR algorithm, this type of approach in parameter estimation allows to grasp some objective quantitative indicators showing the interest of the parsimony principle.

-

Table 2 - Noiseless signal, C=0.15. Lines 3 to 4 give the relative error in % (with respect to exact values) for dimensional ILSVD3 and scaled PETIR-ILSVD3 algorithms.

		β_1	β_2	β_3	β_4	β_5	β_6	$\sqrt{J(\hat{\boldsymbol{\beta}})}$ (k_{max})
	β_j^{exact}	K=0.29 MPa	$W = 35.3$	$L= 0.12$	$V = 5$	$h = 0.4$	$p = 2.27$	
1	β_j^{init}	33.35	30	0.138	4.25	0.5	1.93	
2	$\hat{\beta}_{reg\,j}^{LM}$	29	35.3	0.12	5	0.4	2.27	10^{-9} (21)
3	error $\hat{\beta}_{dim\,j}^{ILSVD3}$	14.8 %	15 %	230 %	15 %	23 %	12.5 %	55 (6)
4	error $\hat{\beta}_{scaled\,j}^{ILSVD3}$	0.5 %	13.7 %	16.1 %	15.6 %	0.7 %	0.3 %	7.8 (5)

Finally, when the initial parameter vector $\boldsymbol{\beta}^{init}$ is selected more distantly from the exact one (C=0.15, see Table2), the difference between the dimensional (ILSVD3, line 3) or scaled (PETIR-ILSVD3, line 4) versions of the regularization technique becomes apparent. The optimum value $u = 3$ for the regularization hyperparameter found previously has been kept. There is a clear superiority, in terms of estimation biases on all the pareameters, of the scaled PETIR-ILSVD3 version.

We can note that in all cases considered in Table 2, the present regularized algorithm behaves worse than the SVD-based LM algorithm. This is absolutely normal for non noisy simulations since in the LM algorithm, only an arbitrary penalty weight is applied on the singular values that do not yield a decrease of residuals at each iteration, but none of them is eliminated. In our algorithm, several (3 here) singular values are replaced by infinite levels and a bias is generated. But it can be observed that without such truncation, a pure SVD (GN)

algorithm does not converge. Here the PETIR-ILSVD3 algorithm (line 4) converges very rapidly (5 iterations instead of 21 for LM algorithm), and produces a good fit (small residuals), and good estimates of the parameters that are identifiable in this experiment and at least, a good order of magnitude for the remaining ones.

8.2.2. Comparison of estimation by PETIR-ILSVD*u* and LM algorithms for a signal corrupted by noise and effect of choice of hyperparameter *u*

We use here noised synthetic signals similar to the one presented in figure 5 (signal to noise ratio close to 10) with a starting point for the parameter vector corresponding to $C = 0.15$ for all simulations.

Table3: Statistical results (over 25 repeated estimations leading to convergence) for the 6 parameters – Average estimates, relative standard deviation, and relative error with respect to the exact values

		β_j^{exact} →	β_1	β_2	β_3	β_4	β_5	β_6
	regularized PETIR versions (lines 1-4, 6-9 and 11-14) ↓		K=0.29 MPa	$W = 35.3$	$L= 0.12$	$V = 5$	$h = 0.4$	$p = 2.27$
1	ILSVD5	$\overline{\hat{\beta}}_j$	28.4	39.7	0.1	-0.6	0.41	2.3
2	ILSVD4	$\overline{\hat{\beta}}_j$	28.7	34.2	0.1	4.3	0.45	2.2
3	ILSVD3	$\overline{\hat{\beta}}_j$	28.8	30.4	0.1	4.2	0.44	2.3
4	ILSVD2	$\overline{\hat{\beta}}_j$	27.4	30.0	0.1	4.3	0.49	2.1
5	LM	$\overline{\hat{\beta}}$	27.3	35.2	0.2	2,8	0.47	2.2
6	ILSVD5	$s_{\hat{\beta}_j} / \overline{\hat{\beta}}_j$ (%)	7.2	10.6	127.2	-2268	27.6	6.3
7	ILSVD4	$s_{\hat{\beta}_j} / \overline{\hat{\beta}}_j$ (%)	0.6	5.6	1.0	0,4	1.3	0.7
8	ILSVD3	$s_{\hat{\beta}_j} / \overline{\hat{\beta}}_j$ (%)	0.5	0.2	0.1	0.1	1.1	0.6
9	ILSVD2	$s_{\hat{\beta}_j} / \overline{\hat{\beta}}_j$ (%)	0.1	0.0	0.0	0.0	0.0	0.1
10	LM	$s_{\hat{\beta}_j} / \overline{\hat{\beta}}_j$ (%)	3.4	10.7	28.3	64.5	3.6	1.5
11	ILSVD5	$(\overline{\hat{\beta}}_j - \beta_j^{exact})/\beta_j^{exact}$ (%)	2.2	13.5	16.6	112.9	2.7	0.7
12	ILSVD4	$(\overline{\hat{\beta}}_j - \beta_j^{exact})/\beta_j^{exact}$ (%)	1.2	2.4	18.6	15.0	12.2	1.0
13	ILSVD3	$(\overline{\hat{\beta}}_j - \beta_j^{exact})/\beta_j^{exact}$ (%)	0.6	13.0	16.1	15.6	10.9	0.4
14	ILSVD2	$(\overline{\hat{\beta}}_j - \beta_j^{exact})/\beta_j^{exact}$ (%)	5.7	14.4	14.8	14.8	23.2	5.8
15	LM	$(\overline{\hat{\beta}}_j - \beta_j^{exact})/\beta_j^{exact}$ (%)	5.8	0.7	66.9	43.5	18.6	3.1

$N = 25$ simulations of the synthetic signals have been made corresponding each to a realization of a randomly generated i.i.d. noise with the same standard deviation as above. The corresponding results of this procedure of the Monte Carlo type [18] are summarized in Table 3 for different versions of PETIR-ILSVDu algorithm ($u = 2$ to 5) as well as for the LM estimations, in terms of different indicators:

- mean value of the estimations of the different parameters:

$$\bar{\bar{\beta}}_j = \frac{1}{N} \sum_{p=1}^{N} \hat{\beta}_j^p \quad \text{for} \quad j=1 \text{ to } 6 \tag{39a}$$

- relative statistical standard deviation of the estimates:

$$s_{\hat{\beta}_j} / \bar{\bar{\beta}}_j = \frac{1}{\bar{\bar{\beta}}_j} \sqrt{\frac{1}{N} \sum_{p=1}^{N} \left(\hat{\beta}_j^p\right)^2 - \left(\bar{\bar{\beta}}_j\right)^2} \quad \text{for} \quad j=1 \text{ to } 6 \tag{39b}$$

- mean relative error of the estimates:

$$(\bar{\bar{\beta}}_j - \beta_j^{exact})/\beta_j^{exact} \tag{39c}$$

Before discussing the results of Table 3, several points deserve to be noted:

- we have shown in the preceding section that the LM algorithm used is always able to produce a good fit with low residuals but here, in the presence of noise, and in such an ill-conditoned problem, it does not produce estimates as good as the best ILSVDu version of PETIR.

- whatever the inverse method used, relative errors on estimates can go from 0 to up to 100%. In other words, if model (37a) is used for mechanical characterization of a material with a 'blind' LM algorithm for estimation, the results could be meaningless.

- the GN algorithm with a dimensional sensitivity matrix never converges if noise is present in the signal. The GN re-scaled algorithm converges sometimes and when it does, it allows a good fit of the "experimental" signal: it behaves like the LM algorithm with parameters estimates which can vary within very far apart bounds.

The efficiency of the ILSVDu algorithm has been checked in Table 3, where only u singular values have been kept in the iterative equations (33a) and (33b).

For PETIR-ILSVD5, convergence is rarely achieved, the PEP remains seriously ill-conditioned. As soon as 2 singular values have been given infinite levels, convergence is systematically obtained and in fewer iterations (of the order of 5) than the LM algorithm (of the order of 8).

If $u = 3$ singular values are kept unchanged, two parameters can be measured (β_1 and β_6) with a very good precision, that is with an error lower than 1%. The other four parameters (β_2, β_3, β_4 and β_5) have the proper order of magnitude (errors of the order of 15%).

Figure 6 illustrates this for ILSVD3, the estimates having been produced through the previous 25 repeated simulations. The relative statistical standard deviation reached for each parameter is very low (see also Table 3, line 8).

It means that the regularized PETIR algorithm find always roughly the same estimates whereas it is not the case for LM where a very large dispersion is observed (see also Table 3, line 10).

When the ILSVD2 case is considered, the variance of the estimated parameter becomes very low but the properly estimated parameters (β_1 and β_6) are obtained with higher errors (ten percent or higher). This can be explained by the fact that only 2 internal degrees of freedom, represented by only two non-zero diagonal parameters p_j in equation (23b), exist whereas 6 parameters are looked for.
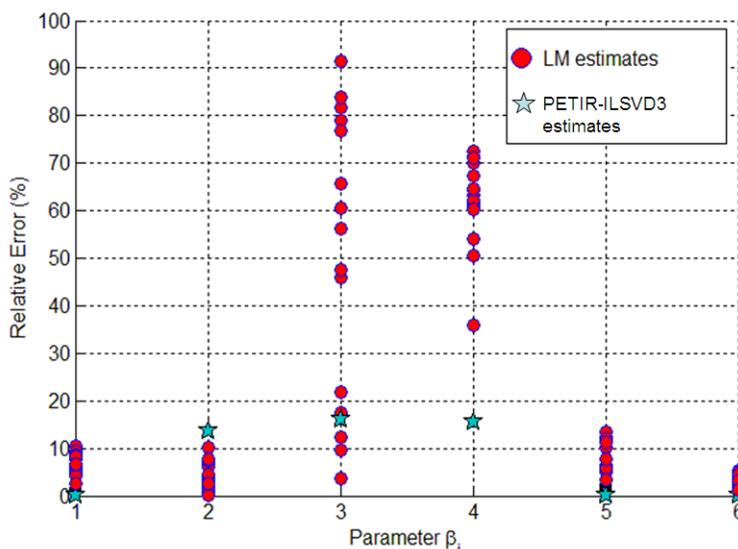

Fig. 6 - Relative errors on the 6 parameter estimates stemming from 25 different realizations of the additive noise on the same output of the model with the corresponding implementation of the LM and PETIR-ILSVD3 estimation algorithms.

Figure 7 below gives an example of estimation results for an experiment with the same signal over noise ratio and a starting point chosen with $C=0.25$, using the PETIR-ILSDV3 algorithm.). It presents the relative variation $\hat{x}_j^{(k)} = \left(\hat{\beta}_j^{(k-1)} - \hat{\beta}_j^{(k-1)}\right) / \hat{\beta}_j^{(k-1)}$ of the parameters with respect to the iteration number k, as well as the relative

decrease $\left(\sqrt{J(\boldsymbol{\beta}^{(k)})} - \sqrt{J(\boldsymbol{\beta}^{(k-1)})}\right)/\sqrt{J(\boldsymbol{\beta}^{(k-1)})}$ of the norm of the residuals. The iterative process converges very quickly (9 iterations here) and the mean of the residuals is equal to zero at convergence.
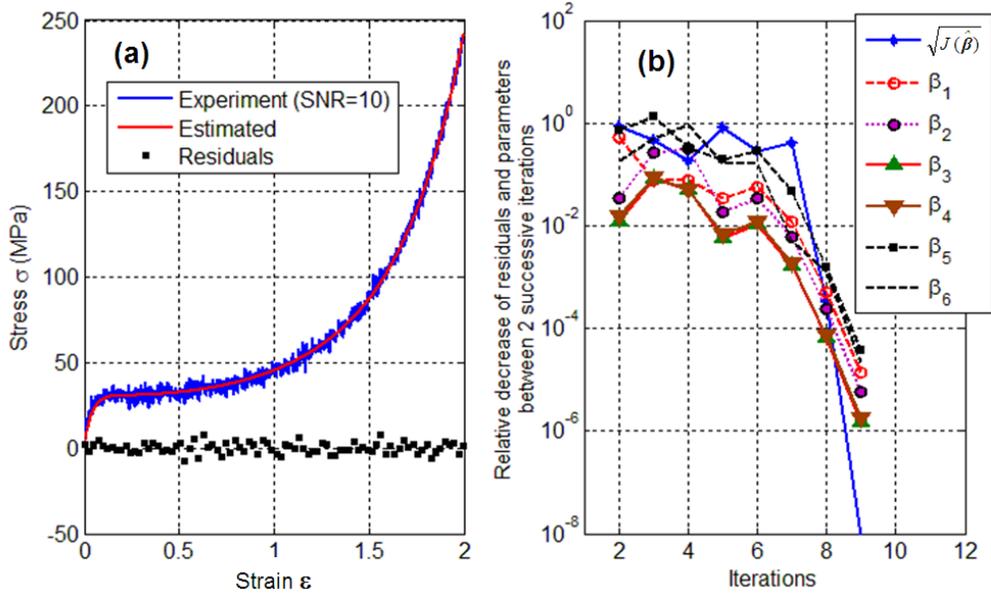

Fig. 7- PETIR-ILSVD3 applied to a signal with SNR=10
(a) Recalculated and noisy output signals and residuals
(b) Relative variation of the square root of the least-squares criterion and of the estimated parameters between successive iterations

9. Conclusion

In this paper, we proposed a non linear parameter estimation method, called PETIR (Parameter Estimation Through Iterative Rescaling) where the parameter vector is redefined by scaling at each iteration step. This local scaling allows us to get rid of the physical units of the different parameters: these units make the use of norms for the corresponding sensitivity vectors incoherent. Even if all the parameters are dimensionless, this rescaling also allows us to deal with very different order of magnitudes for them, without encountering any specific numerical difficulty.

A special kind of SVD-based regularization has been implemented: it differs from Truncated Singular Value Decomposition often used in linear inverse function estimation problems because the number of right singular vectors has to be equal to the number of parameters and cannot be decreased. The alternative proposed here is to replace the smallest singular values by infinite levels. We coined this technique Infinite Levels SVD (ILSVD).

This type of PETIR regularization has been tested on two different models used for characterization in heat transfer and in solid rheology. It can be applied to all kind of problem in engineering science. When applied to noisy signals, it produces better estimates than the current Levenberg-Marquardt algorithm based on the SVD of the dimensional sensitivity matrix. This new algorithm allows the estimation of parameters even for highly ill-conditioned scaled sensitivity matrices. Conversely, it shows that robust convergence can be achieved in very poorly conditioned cases with very good residuals, with a number of internal degrees of freedom smaller than the number of parameters of the model, since the estimates are necessarily strongly correlated: this should question the inverter-experimenter about the pertinent character of the model he uses and may lead him to adopt more parsimonious models or at least models which are not underdetermined with respect to the information contained in the data.

Future work will consist in taking the bias caused by truncation into account for optimizing the number of singular values to be kept unchanged.

REFERENCES

1. Gill P E and Murray W 1978 Algorithms for the solution of the nonlinear least-squares problem, *SIAM J. Numer. Anal*. **15** (5) 977-992.

2. Finsterle S and Kowalsky B 2011 A truncated Levenberg-Marquardt algorithm for the calibration of highly parameterized non linear models *Computers and Geosciences* **37** 731-738.

3. Tonkin M J and Doherty J 2005 A hybrid regularized inversion methodology for highly parameterized environmental models *Water Resources Research* **41** W10412.

4. Lawson C L and Hanson J R 1995 Solving Least-Squares Problems *Classics Appl.Math*. **15** 337 Soc. For Ind. And Appl. Math. Philadelphia.

5. Beck J V and Arnold K J, 1977 *Parameter Estimation in Engineering and Science* Wiley New York.

6. Buckingham E 1914 On physically similar systems. Illustrations of the use of dimensional equations *Physical Review* **4** 345-376.

7. Walter E and L. Pronzato L 1997 *Identification of Parametric Models from Experimental Data* Springer-Verlag Heidelberg.

8. Rémy B, André S, Maillet D *Non linear parameter estimation problems: tools for enhancing metrological objectives* in: Proceeding of Eurotherm Seminar 94: Metti 5 Advanced Spring School "*Thermal Measurements and Inverse Techniques*", Roscoff, Brittany, France, June 13-18, 2011, free download at http://www.sft.asso.fr/document.php?pagendx=12299&project=sft

9. Sober, E. 1988 Reconstructing the Past. Parsimony, Evolution, and Inference. MIT Press, Cambridge (Mass) London.

10. Degiovanni A 1977 Thermal Diffusivity and Flash Method, *International Journal of Thermal Sciences*, Issue **185** (16) 420-441

11. Hay B, Filtz J R,, Hameury J , Davée G, Rongione L, Enouf O. 2009 Thermal-diffusivity measurement of ceramic coatings at high temperature using "front-Face" and "rear-Face" laser flash methods *International Journal of Thermophysics* **30** (4) 1270-1282.

12. Morozov V A 1966 On the solution of functional equations by the method of regularization *Soviet Mathematics–Doklady* **7** 414–417.

13. Hanson R J 1971 A numerical method for solving Fredholm integral equations of the first kind using singular values *SIAM J. Numer. Anal.* **8** 616-622.

14. Hansen P C 1990 Truncated Singular Value Decomposition Solutions to Discrete Ill-Posed Problems with Ill-Determined Numerical Rank *SIAM J. Sci. and Stat. Comput.* **11** (3) 503–518.

15. Rémy B and André S Nonlinear estimation prolems, chapter 9, pp. 315-354, in *Thermal Measurements and Inverse Techniques*, Editors**:** Helcio R.B. Orlande; Olivier Fudym; Denis Maillet; Renato M. Cotta, Publisher: CRC Press, Taylor & Francis Group, Bocca Raton, USA, 779 pages, May 09, 2011

16. G'sell C and Jonas J J. 1979 Determination of the plastic behaviour of solid polymers at constant true strain rate *Journal of Materials Science* **14** (3), 583-591.

17. André S, Renault N, Meshaka Y, Cunat C 2012 From the thermodynamics of constitutive laws to thermomechanical experimental characterization of materials: An assessment based on inversion of thermal images *Continuum Mechanics and Thermodynamics* **24** (1) 1-20.

18. Maillet D, André S, Degiovanni A 1993 Errors in the thermal diffusivity measured by the flash method: comparison between theory and experiments (Les erreurs sur la diffusivité thermique mesurée par méthode flash: confrontation théorie-expérience) *The European Physical Journal* (Former *Journal de Physique III*) **3** 883-909.