A new sufficient condition for the well-posedness of non-linear least square problems arising in identification and control

Guy Chavent

UNITÉ DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
BP 105
78153 Le Chesnay Cedex
France
Tél. (1) 39 63 55 11

Rapports de Recherche

N° 1121

Programme 7
Calcul Scientifique,
Logiciels Numériques et Ingénierie Assistée

A NEW SUFFICIENT CONDITION FOR THE WELL-POSEDNESS OF NON-LINEAR LEAST SQUARE PROBLEMS ARISING IN IDENTIFICATION AND CONTROL

Guy CHAVENT

Novembre 1989

RR-1121

A NEW SUFFICIENT CONDITION FOR THE WELL-POSEDNESS OF NON-LINEAR LEAST SQUARE PROBLEMS ARISING IN IDENTIFICATION AND CONTROL

Guy CHAVENT
CEREMADE, 75775 Paris Cédex 16
INRIA, BP 105, 78153 Le Chesnay Cédex

Abstract

We show how simple 1-D geometrical calculations (but along all maximal segments of the parameter or control set!) can be used to establish the wellposedness of a non-linear least-square (NLLS) problem and the absence of local minima in the corresponding error function. These sufficient conditions, which are shown to be sharp by elementary examples, are based on the use of the recently developped "size × curvature" conditions for prooving that the output set is strictly quasiconvex. The use of this geometrical theory as a numerical or theoretical tool is discussed. Finally, application to regularized NLLS problem is shown to give new information on the choice of the regularizing parameter.

Résumé

Nous montrons comment des calculs géométriques 1-D très simples (mais à effectuer le long de tous les segments maximaux de l'ensemble des paramètres ou contrôles admissibles!) permettent de montrer qu'un problème de moindre carrés non-linéaires (MCNL) est bien posé et ne possède pas de minima locaux. Ces conditions suffisantes, que l'on montre être précises sur des exemples élémentaires, résultent de l'utilisation des toutes récentes condition de "taille × courbure" pour montrer que l'ensemble des sorties est quasiconvexe. On discute ensuite de l'utilisation possible de cette théorie géométrique comme outil numérique ou théorique. On l'applique enfin au problème MCNL régularisé, ce qui donne des informations nouvelles pour le choix du paramètre de régularisation.

Key words : Non-linear least squares, parameter estimation, identification, inverse problems.

Mots clefs : Moindres carrés non linéaires, estimation de paramètres, identification, problèmes inverses.

A NEW SUFFICIENT CONDITION FOR THE WELL-POSEDNESS OF NON-LINEAR LEAST SQUARE PROBLEMS ARISING IN IDENTIFICATION AND CONTROL

UNE NOUVELLE CONDITION SUFFISANTE POURQU'UN PROBLEME DE MOINDRE CARRES NON LINEAIRE ISSU DE L'IDENTIFICATION OU DU CONTROLE SOIT BIEN POSE

Guy CHAVENT

Contents

1 Introduction

Parameter estimation problems are often set as output-least square problems and solved numerically as such. But usually many theoretical and practical problems stay open, as existence and uniqueness of the global minimum, possible existence of local minima (which are highly undesirable from a computational point of view), stability of the global minimum with respect to the data ... The reason is that the parameter → output mapping usually exhibits no interesting mathematical properties except regularity. Hence we have tried to develop, in a serie of papers [1] [2] [3] [4] [5], some quantitative sufficient conditions for the well-posedness of the non linear least squares problem, following the intuition that this well-posedness should hold as soon as the size of the output set (on wich one projects the data) is "not too large" with respect to its curvature. We illustrate in this paper how the most precise "size × curvature" conditions developped in [5] can be used to obtain insight into well-posedness, stability and local minima of non-linear least square problems.
We refer to the introduction of [3] for a more detailed discussion of the motivations to this approach, and to reference [6] for an example of application of these technique to a plane wave inversion problem.

2 The non-linear least square problem

We consider a "parameter space" E, a "set of admissible parameters" C, a "data space" F and a "parameter → output mapping" φ satisfying

$$(2.1)\quad \begin{cases} E &= \text{Banach space, with norm } \|\|_E \\ C &= \text{Closed convex set of } E \\ F &= \text{Hilbert space, with scalar product } <,>_F \\ \varphi &: C \to F \text{ a } C^2\text{-mapping, with } \varphi, \varphi' \text{ and } \varphi'' \text{ bounded over } C \end{cases}$$

The inversion of a given "data"

$$(2.2)\qquad z \in F$$

in the least-squares sense reads then ;

$$(2.3)\qquad \text{find } \hat{x} \in C \text{ such that } J(x) = \|\varphi(x) - z\|_F^2 = \min \text{ over } C$$

which is the non-linear least-square problem we want to investigate. Of course, well-posedness of problem (2.3) does not hold under the sole hypothesis (2.1), and a classical way to enhance the behaviour of (2.3), in the absence of other information on φ, is to replace (2.3) by its Tychonov regularization

$$(2.3)_\varepsilon\quad \begin{cases} \text{find } \hat{x}_\varepsilon &\in C \cap \mathcal{E} \text{ such that} \\ J_\varepsilon(x) &= \|\varphi(x) - z\|_F^2 + \varepsilon^2 \|x - x_0\|_{\mathcal{E}}^2 = \min \text{ over } C \cap \mathcal{E} \end{cases}$$

where

$$(2.4)\quad \begin{cases} \mathcal{E} \subset E & \text{is an Hilbert subspace of } E, \text{ with scalar product } <,>_{\mathcal{E}} \\ x_0 \in \mathcal{E} & \text{is a given " a priori guess" for the unknown parameter} \\ \varepsilon > 0 & \text{is the regularization parameter} \end{cases}$$

Usually, $\mathcal{E}$ is chosen to imbed compactly in E, which automatically ensures existence of a solution to $(2.3)_\varepsilon$, but does not give any hint on uniqueness, stability of $\hat{x}_\varepsilon$ and local minima of J_ε. On the contrary, the Hilbert-type techniques used in this paper will allow to skip the compactness hypothesis (for example chosing $\mathcal{E} = E$ - provided E itself is an Hilbert) and yet proove the well-posedness of $(2.3)_\varepsilon$ for large enough ε.

Of course, $(2.3)_\varepsilon$ can be written in the same form as (2.3) by a paper choice of E, F, φ and z, namely :

$$(2.5) \quad \begin{cases} E & \text{replaced by } E_\varepsilon = \mathcal{E} \\ C & \text{replaced by } C_\varepsilon = C \cap \mathcal{E} \\ F & \text{replaced by } F_\varepsilon = F \times \mathcal{E} \text{ with the scalar product } < X, Y >_F + < x, y >_{\mathcal{E}} \\ \varphi & \text{replaced } \varphi_\varepsilon = (\varphi, \varepsilon I) \\ z & \text{replaced by } z_\varepsilon = (z, \varepsilon x_0) \end{cases}$$

Hence we shall first state our results for problem (2.3), and then specialize them to the case (2.5) of the regularized problem.

3 Pathes and Related geometrical attributes

The main step to obtain the well-posedness of (2.3) will be to proove that $\varphi(C)$ is a strictly quasiconvex set, as this will imply the existence of a neighbourhood of $\varphi(C)$ on which the "projection on $\varphi(C)$" is well behaved. Hence we need to equip $\varphi(C)$ with a collection $\mathcal{P}$ of pathes, wich we simply choose to be the image by φ of all segments of C :

$$(3.1) \qquad \mathcal{P} = \{\varphi([x, y]), x, y \in C\}$$

for which a subfamily of maximal pathes (cf [5]) is obviously given by

$$(3.2) \qquad \mathcal{P}_{\mathcal{M}} = \{\varphi([x, y]), x, y \in \partial C]\}$$

where ∂C is the relative boundary of C. A path $P = \varphi([x, y])$ of $\mathcal{P}$ is naturally parametized by $t \in [0, 1]$:

$$(3.3) \qquad P(t) = \varphi((1 - t)x + ty)$$

and we can define the velocity $V(t)$ and acceleration $A(t)$ associated to this parametization by

$$(3.4) \qquad \begin{cases} V(t) &= \varphi'(x_t)(y - x) \\ A(t) &= \varphi''(x_t)(y - x, y - x) \end{cases}$$

with

$$(3.5) \qquad x_t = (1 - t)x + ty.$$

The theory requires that pathes can be reparametized as C^2-function of the arc-length ν, which satisfies :

(3.6)
$$d\nu = \|V(t)\|dt$$

Hence at points t of a path where $V(t) \neq 0$, the velocity $v(t)$ and acceleration $a(t)$ with respect to are length ν are given by

(3.7)
$$\begin{cases} v(t) &= \dfrac{V(t)}{\|V(t)\|} \\ a(t) &= \dfrac{A(t)}{\|V(t)\|^2} - v < v, \dfrac{A(t)}{\|V(t)\|^2} > \end{cases}$$

and their norms by :

(3.8)
$$\begin{cases} \|(v(t)\| &= 1 \\ \|a(t)\| &= \dfrac{\|A(t)\|}{\|V(t)\|^2}(1- < v, \dfrac{A(t)}{\|A(t)\|} >^2)^{\frac{1}{2}} \end{cases}$$

At points to of a path where $V(to) = 0$, one may still define $v(t)$ and $a(t)$ by continuity provided that

(3.9)
$$V(to) = 0 \Rightarrow \begin{cases} \text{The right-hand sides in (3.7) are defined} \\ \text{and have a limit when } t \to to,\ t \neq to \end{cases}$$

Notice that (3.9) holds automatically as soon as $\varphi'(x)$ is injective for any x in C!

Now that $\varphi(C)$ is equipped with C^2-pathes with respect to the arc-length, we define the **radius of curvature** $\rho(t)$ of a path $P = \varphi([x, y])$ at parameter t by :

(3.10)
$$\rho(t) = \|a(t)\|^{-1} = \frac{\|V(t)\|^2}{\|A(t)\|}(1- < v, \frac{A(t)}{\|A(t)\|} >^2)^{-\frac{1}{2}} \geq \frac{\|V(t)\|^2}{\|A(t)\|}$$

Following [5], we introduce also the **global radius of anvature** $\rho_G(t, t')$ of P at t seen from t', which is given by the formula

(3.11)
$$\rho_G(t, t') \begin{cases} 0 & \text{if } Sgn(t'-t) < P' - P, v' > \ \leq 0 \\ \dfrac{Sgn(t'-t) < P' - P, v' >}{(1- < v, v' >^2)^{\frac{1}{2}}} & \text{if } Sgn(t'-t) < P' - P, v' > \ > 0 \text{ and } < v, v' > \geq 0 \\ Sgn(t'-t) < P' - P, v' > & \text{if } Sgn(t'-t) < P' - P, v' > \ > 0 \text{ and } < v, v' > \leq 0 \end{cases}$$

where $P, P', v, v' \ldots$ stand for $P(t), P(t'), v(t), v(t') \ldots$ with the convention that $\rho_G(t, t') = +\infty$ if $Sgn(t'-t) < P' - P, v >> 0$ and $< v, v' >= 1$.

Geometrically, $\rho_G(t, t')$ is the distance of $P(t)$ to the intersection of the two half spaces normal to each end of the subpath of P located between $P(t)$ and $P(t')$. It is related to the usual radius of curvature $\rho(t)$ by

(3.12)
$$\lim_{t \to t'} \rho(t, t') = \rho(t)$$

We denote also by $\delta(t,t')$ the **arc length** between the two points $P(t)$ and $P(t')$ of the path $P = \varphi([x,y])$:

$$(3.13) \qquad \delta(t,t') = \int_t^{t'} \|V(\tau)\| d\tau$$

and by $\Theta(t,t')$ the **deflection** between the path directions at the two same points ;

$$(3.14) \qquad \Theta(t,t') = \cos^{-1}(< v(t), v(t') >)(\text{in radian}),$$

which statisfies (cf [5]) :

$$(3.15) \qquad \Theta(t,t') \le \int_t^{t'} \frac{d\nu(t)}{\rho(t)} = \int_t^{t'} \frac{\|A(\tau)\|}{\|V(\tau)\|}(1- < v, \frac{A(\tau)}{\|A(\tau)\|} >^2)^{\frac{1}{2}} d\tau \le \int_t^{t'} \frac{\|A(\tau)\|}{\|V(\tau)\|} d\tau$$

These geometrical quantities are used to associate to the non-linear least square problem itself, ie to C, φ and $\mathcal{P}$, the following numbers, which shall be relevant for the study of its wellposedness :

(3.16)	$R(C,\varphi,\mathcal{P})$	$=$	$\inf_{x,y\in C} \inf_{t\in[0,1]} \rho(t)$	(smallest radius of curvature)
(3.17)	$R_G(C,\varphi,\mathcal{P})$	$=$	$\inf_{x,y\in C} \inf_{t,t'\in[0,1]} \rho_G(t,t')$	(smallest global radius of curvature)
(3.18)	$\Delta(C,\varphi,\mathcal{P})$	$=$	$\sup_{x,y\in C} \sup_{t,t'\in[0,1]} \delta(t,t')$	(largest path length)
(3.19)	$\Theta(C,\varphi,\mathcal{P})$	$=$	$\sup_{x,y\in C} \sup_{t,t'\in[0,1]} \Theta(t,t')$	(largest path-deflection)

Of course, from (3.12) we see that

$$(3.20) \qquad R_G(C,\varphi,\mathcal{P}) \le \mathcal{R}(C,\varphi,\mathcal{P}).$$

In practice, our sufficient conditions for wellposedness will require only to know lower and upper bounds to the above quantities, which we shall denote by the same letter, but without arguments :

$$(3.21) \qquad \begin{array}{rcl} R & \le & R(C,\varphi,\mathcal{P}) \\ R_G & \le & R_G(C,\varphi,\mathcal{P}) \\ \Delta & \ge & \Delta(C,\varphi,\mathcal{P}) \\ \Theta & \ge & \Theta(C,\varphi,\mathcal{P}) \end{array}$$

In view of (3.20) it is reasonable to suppose that R and R_G has been chosen such that as to satisfy :

$$(3.22) \qquad R_G \le R$$

For example,its easy to check that the following choices for R, Δ and Θ satisfy (3.21) :

$$(3.23) \qquad R = \inf_{x,y\in\partial C} \inf_{t\in[0,1]} \frac{\|\varphi'(x_t)(y-x)\|^2}{\|\varphi''(x_t)(y-x,y-x)\|}$$

$$(3.24) \qquad \Delta = \sup_{x,y\in\partial C} \int_0^1 \|\varphi'(x_t)(y-x)\| dt$$

$$(3.25) \qquad \Theta = \sup_{x,y\in\partial C} \int_0^1 \frac{\|\varphi''(x_t)(y-x,y-x)\|}{\|\varphi'(x_t)(y-x)\|} dt$$

where $x_t = (1-t)x + ty$. We recall also theorem 6.7 of [5] which allows to obtain a lower bound R to $R_G(C,\varphi,\mathcal{P})$ in terms of R, Δ and Θ :

Theorem 3.1 *Let R, Δ, Θ satisfy (3.21). Then :*

$$R_G = \begin{cases} R & \text{if } 0 \le \Theta \le \pi/2 \\ R\sin\Theta + (\Delta - R\Theta)\cos\Theta & \text{if } \pi/2 \le \Theta \le \pi \end{cases} \tag{3.26}$$

satisfies (3.21) and (3.22).

4 Wellposedness of the non-linear least square problem

We consider now the non-linear least square problem (2.3), which throughout all this paragraph is supposed to satisfy hypothesis (2.1) and (3.9) at least. When it comes to actual numerical resolution of (2.3) by an optimization algorithm, it is of utmost importance that the objective function $J(x)$ has no local minima. Hence we shall incorporate this property into our definition of wellposedness :

Definition 4.1 *(Q-well posedness) The non-linear least square problem (2.3) is said to be Q-wellposed on some open neighborhood ϑ of $\varphi(C)$ for some pseudo distance $\delta(x,y)$ on C if and only if :*

(i) *for any $z \in \vartheta$, there exists a unique solution $\hat{x}$ of (2.3)*

(ii) *the $z \rightsquigarrow \hat{x}$ mapping is Lipschitz continuons from $(\vartheta, \|\|_F)$ to $(C, \delta(x,y))$*

(iii) *for any $z \in \vartheta$, J is strictly quasiconvex, (ie has no local minima distinct from $\hat{x}$), and any minimizing sequence converges to $\hat{x}$ for $\delta(x,y)$.*

If we take $\delta(x,y) = \|x-y\|_E$ and drop condition iii), the above definition reduces to Output Least Square Identifiability (OLSI, cf [2]) introduced in the context of parameter estimation. In the general case, Q-wellposedness is both weaker (because stability is required only for $\delta(x,y)$ instead of $\|x-y\|_E$) and stronger (because it requires strict quasiconvexity of J) than OLSI.

The pseudo-distance $\delta(x,y)$ on C which appears quite natural for the NLLS problem (2.3) is the arc length, in data space F, of the path $\varphi([x,y])$, ie, with notations (3.5) :

$$\delta(x,y) = \int_0^1 \|\varphi'(x_t)(y-x)\|_F dt \tag{4.1}$$

Of course, if one makes the additional hypothesis that beside being bounded over C as stated in (2.1), $\varphi'(x)$ admits a pseudoinverse which is also bounded over C, or in other terms that

$$\begin{cases} \exists \alpha_M \ge \alpha_m > 0 & \text{such that} \\ \alpha_m \|y\|_E \le \|\varphi'(x).y\|_F \le \alpha_M \|y\|_E & \forall x \in C, \forall y \in E \end{cases} \tag{4.2}$$

then $\delta(x,y)$ becomes equivalent to the usual norm in E :

$$\alpha_m \|x-y\|_E \le \delta(x,y) \le \alpha_M \|x-y\|_E \quad \forall x,y \in C \tag{4.3}$$

It is now very easy to use the results of [5] to establish well-posedness results for (2.3). We begin with the most precise condition :

Theorem 4.1 *(R_G-sufficient condition) Let hypothesis (2.1) and (3.9) hold, and $R \geq R_G$ be lower bounds to $R(C, \varphi, \mathcal{P})$ and $R_G(C, \varphi, \mathcal{P})$ as defined in (3.16) (3.17) and (3.21).*
If :

(4.4) $$R_G > 0$$

(4.5) $$\varphi(C) \textit{closed in } F$$

Then :
i) The NLLS problem (2.3) is Q-wellposed on

(4.6) $$\vartheta = \{z \in F | d(z, \varphi(C)) < R_G\}$$

for the pseudo-distance $\delta(x, y)$ on C defined in (4.1).
ii) More precisely, if $z_j \in \vartheta, j = 0, 1$ satisfy

(4.7) $$\|z_0 - z_1\|_F + \max_{j=0,1} d(z_j, \varphi(C)) \leq d < R_G$$

for some d, then the corresponding solutions $\hat{x}_j, j = 0, 1$ of (2.3) satisfy the **stability estimate :**

(4.8) $$\delta(\hat{x}_0, \hat{x}_1) \leq (1 - d/R)^{-1} \|z_0 - z_1\|_F$$

Proof :

Hypothesis (2.1) (3.9) and (4.4) imply, using theorem 5.12 of [5], that $\varphi(C)$ is strictly quasiconvex with a neighborhood ϑ given by (4.6), which, together with the fact that $\varphi(C)$ is closed, implies existence, uniqueness and Lipschitz stability of the projection on $\varphi(C)$ all over ϑ, and absence of local minima by theorems 3.6, 3.9 and 3.5 of [5]. □

Of course, if the strong estimate (4.2) on the derivative $\varphi'(x)$-the so-called "sensitivity Matrix" in the finite dimensional case-holds, then one obtains stability for the $\|x - y\|_E$ distance on C, and $\varphi(C)$ is necessarily closed :

Corollary 4.1 *Let hypothesis (2.1) and (4.2) hold, and $R \geq R_G$ be lower bounds to $R(C, \varphi, \mathcal{P})$ and $R_G(C, \varphi, \mathcal{P})$ as defined in (3.16) (3.17) and (3.21).*
If :

(4.9) $$R_G > 0$$

Then

(4.10) $$\varphi(C) \quad \textit{is closed in } F$$

and conclusions i) and ii) of theorem (4.2) hold, with (4.8) replaced by

(4.11) $$\alpha_m \|\hat{x}_0 - \hat{x}_1\|_E \leq (1 - d/R)^{-1} \|z_0 - z_1\|_F$$

The above theorem can be seen as a generalization of the local inversion theorem or implicit function theorem to the case where the range of $\varphi'(x)$ can be strictly included in F, so that the equation $\varphi(u) = z$ can be solved only in the least-square sense.

We come now to a slightly less precise condition, based on the lower bound to $R_G(C, \varphi, \mathcal{P})$ given in therorem 3.1 :

Theorem 4.2 *(Θ-sufficient condition)*
Let hypothesis (2.1) and (3.9) hold, and R, Δ, Θ be given satisfying (3.21) (for example using formula (3.23 thru 25)), and suppose that

(4.12) $$\varphi(C) \quad \textit{is closed.}$$

If :

(4.13) $$0 \leq \Theta \leq \pi/2 \quad \textit{and } R > 0$$

or

(4.14) $$\pi/2 \leq \Theta < \pi \quad \textit{and } \Delta/R < \Theta - \tan\Theta$$

Then R_G defined by (3.26) satisfies

(4.15) $$R_G > 0$$

and conclusions i) and ii) of theorem 4.1 hold for this value of R_G.

Proof :it results immediatly from theorems 3.1 and 4.1. □

Notice (cf the properties (3.15) of the deflection Θ) that the conditions $\Theta \leq \pi/2$ or π are actually *integral size* $(d\nu)\times$ *curvature* $(1/\rho(\nu))$ *conditions* for the output set $\varphi(C)$, and that condition $\Delta/R < \Theta - \tan\Theta$ is also a *size* $(\Delta)\times$ *curvature* $(1/R)$ *condition,* but which involves the *largest size* and the *smallest curvature !*

We now specialize theorem 4.2 to the case where $\varphi'(x)$ and its pseudoinverse are uniformly bounded over C, i.e. when (4.2) holds. Using the least hypothesis of (2.1), we know that :

(4.16) $$\exists \beta > 0, \|\varphi''(x)(y,y)\|_F \leq \beta \|y\|^2 \forall x \in C, \forall y \in C$$

Then we have the

Corollary 4.2 *Let hypothesis (2.1) and (4.2) hold, and define :*

(4.17) $$\begin{aligned} R &= \alpha_m^2/\beta \\ \Theta &= (\beta/\alpha_m) diam\ C \end{aligned}$$

and :

(4.18) $$R_G = \begin{cases} R & 0 \leq \Theta \leq \pi/2 \\ R(\sin\Theta + (\alpha_M/\alpha_m - 1)\Theta\cos\Theta) & \pi/2 \leq \Theta \leq \pi \end{cases}$$

If :

(4.19) $$0 \leq \Theta \leq \pi/2$$

or

(4.20) $$\pi/2 \le \Theta < \pi \quad \textit{and } R_G > 0$$

Then the NLLS problem (2.3) is Q-vellposed on the cylindrical neighborbood (4.6) of $\varphi(C)$ for the $\|x-y\|_E$ distance on C, and the stability estimate (4.11) holds as soon as the data z_0, z_1 satisfy (4.7).

Notice that, under the hypothesis (2.1) and (4.2) of the above corollary, the Q-wellposedness of the NLLS problem will be ensured as soon as the diameter of C is small enough. Notice also that the conditions (4.19) or (4.20) on the admissible size for C are sharp, as it can be verified by considering the two following simple examples :

(4.21) $$\begin{cases} \textit{example 1} & : \text{given } X>0, \text{ choose :} \\ E=R & ,C=[0,X], \\ \varphi(x) & =(\cos x, \sin x) \end{cases}$$

(4.22) $$\begin{cases} \textit{example 2} & : \text{given } Y<0<X, \text{ choose :} \\ E=R & ,C=[Y,X], \\ \varphi(x) & =\begin{cases} (1,x) & \text{for } Y \le x \le 0 \\ (\cos x, \sin x) & \text{for } 0 \le x \le X \end{cases} \end{cases}$$

(The φ mapping of example 2 is not C^2 at $x=0$, but the whole theory still holds for path is which are only C^1 and piecewise C^2).

We conclude this paragraph with a (tentatively critical ...) discussion of the applicability of the above results to parameter estimation problems.

The fundamental prerequisite for the use of this theory is that $\varphi(C)$ is a "variety" which carries pathes $\mathcal{P}$ having radii of curvature larger than some $R>0$. (This of course requires that the "variety" $\varphi(C)$ itself has a "bounded curvature"). The condition $R_G > 0$ then can always be satisfied by reducing the size of C, and is hence less critical.

Let us first consider the case where E is finite dimensional, for example in parameter estimation in O.D.E.s, or in P.D.E.s after discretization has been performed. This case is very important practically, as it is the only one which one can actually attempt to solve on a computer ! One may then try to use the above geometrical theory to determine if the setting of the parameter estimation problem is satisfying, ie if the knowledge of C and φ allows for a unique, stable determination of the parameter, and if so, which accuracy is required on the data. This requires the estimation of lower bounds R and R_G to the smallest radius of curvature (3.16) and global radius of curvature (3.17), which by sure is not an easy task. When the dimension of E is not too large, one can try a numerical determination of R using (3.16) (3.10) and R_G using (3.17) (3.11), and use theorem 4.1 which given the most precise sufficient condition. This includes intensive computation (namely, along all segments $[x,y]$ with extremities x and y located on the (relative) boundary ∂C of C!), which may quickely become unaffordable when the number of parameters is larger than a few units... But the reward for this computational effort is a treasurable information on the wellposedness of the NLLS problem and the absence of local minima in the objective function for *any* z in ϑ, which is practically very useful both for the engineer who has set the parameter estimation problem ("do I have enough information for recovering my parameter in a unique and stable way ?") and the numerical analycist in charge of the computations ("is my optimization routine going to be stuck in local minima ?"). When this numerical approach is impossible, one may think of calculating analytically (ie with paper and pencil) a lower bound R using (3.16) (3.10), upper bounds Θ using (3.25) and Δ using (3.24) (all these quantities are expressed by simple formula involving only $\varphi'(x)(y-x)$ and $\varphi''(x)(y-x, y-x)$, and then use theorem 4.2 to get information on the wellposedness of the NLLS problem. There are yet no example where this approach has been used, but the corresponding theory is just being released now in [5] and in this paper, so we hope that some application along this line will show up in the future.

Let us turn now to the case where E is infinite dimensional, as for example in parameter estimation in P.D.E.s. We expect here that the generic (in an imprecise sense ...) situation is $R(C, \varphi, \mathcal{P}) = 0$, so that the above geometrical theory does not apply. As a support for this assertion, we refer to [2] where it was shown, for the model problem of the estimation of a diffusion coefficient function in a 1-D elliptic equation, that one can find, when the discretization is refined, a sequence of pathes on which the smallest radii of curvature tends to zero (these small radius of curvature are obtained for perturbations of the diffusion coefficient having smaller and smaller support containing one stationary point of the solution to the elliptic PDE). It would however be maybe possible to prove some wellposedness results in the somewhat academic case where the solution to the elliptic equation does not possess any stationary point. A bright singular point in this dark picture of the situation for the infinite dimensional case is given in [6], where the estimation of the shape (ie a function) and the phase (a number) of a plane wane is discussed, and analysed using the above geometrical theory. To conclude on infinite dimensional E, let us mention that the geometric theory may reveal as a useful tool for analysing how the well-posedness of the NLLS problem deteriorates when E is approximated by larger and larger finite dimensional spaces-multiscale analysis of functions should play a crucial role here (see [2] and [7] for very preliminary results).

5 Well-posedness of the regularized NLLS problem

We investigate in this paragraph the wellposedness of the regularized problem $(2.3)_\varepsilon$ under the minimum set of hypothesis (2.1) (2.4). Of course, as we have not required any compact injection from $\mathcal{E}$ into E, this minimum set of hypothesis does not ensure in general even the existence of a solution $\hat{x}_\varepsilon$, in opposition to the linear case (i.e. $\varphi \in \mathcal{L}(E, F)$) where the same hypothesis ensure the existence of a unique $\hat{x}_\varepsilon$. We shall be able in this paragraph to quantify the natural intuition that "a minimum amount" of regularization should be added in order to compensate for the non-linearity of φ, and restore a situation similar to that of the linear case.

As mentioned at the end of paragraph 2, the study of the wellposedness of the regularized problem $(2.3)_\varepsilon$ can be made very simply by applying all results of paragraph 4 to the NLLS problem (2.3) with a proper choice for E, C, F, φ and z as explaned in (2.5).

For sake of simplicity, we shall explicit this approach only for the case of corollary 4.2.

Using hypothesis (2.1) and (2.4), we know that :

$$\exists \tilde{\alpha}_M > 0, \|\varphi'(x).y\|_F \leq \tilde{\alpha}_M \|y\|_{\mathcal{E}}, \forall x \in C \cap \mathcal{E}, \forall y \in \mathcal{E} \tag{5.1}$$

$$\exists \tilde{\beta} > 0, \|\varphi''(x).(y,y)\|_F \leq \tilde{\beta} \|y\|_{\mathcal{E}}^2, \forall x \in C \cap \mathcal{E}, \forall y \in \mathcal{E} \tag{5.2}$$

(notice that if E itself happens to be an Hilbert space and the choice $\mathcal{E} = E$ is made, then $\tilde{\alpha}_M$ coincides with α_M defined in (4.2) and $\tilde{\beta}$ coincides with β defined in (4.16)).

As in corollary 4.2, we shall need the size of C in the parameter space $\mathcal{E}$:

$$\text{diam}(C, \mathcal{E}) = \sup_{x,y \in C} \|x - y\|_{\mathcal{E}} \tag{5.3}$$

but the position of the a-priori guess x_0 with respect to C with also play a role through the "radius of C seen from x_0 in $\mathcal{E}$" :

$$\text{rad}(C, x_0, \mathcal{E}) = \sup_{x \in C} \|x - x_0\|_{\mathcal{E}} \tag{5.4}$$

In order to express the results in a simple form, we introduce the following dimensionless quantities :

$$\begin{cases} \text{position index of } x_0 \text{ w.r.t. } C : \\ \eta = \text{rad}(C, x_0, \mathcal{E})/\text{diam}(C, \mathcal{E}) \end{cases} \tag{5.5}$$

$$\begin{cases} \text{adimensional regularization parameter :} \\ \bar{\varepsilon} = \varepsilon/(\tilde{\beta}\text{diam}(C, \mathcal{E})) \end{cases} \tag{5.6}$$

$$\begin{cases} \text{adimensional distance in data space :} \\ \overline{d}(z, z') = d(z, z')/(\tilde{\beta}\text{diam}(C, \mathcal{E})^2) \end{cases} \tag{5.7}$$

$$\begin{cases} \text{adimensional upper bound to sensitivity :} \\ \zeta = \tilde{\alpha}_M/(\tilde{\beta}\text{diam}(C, \mathcal{E})) \end{cases} \tag{5.8}$$

Notice that

$$1/2 \leq \eta \leq 1 \quad \text{as soon as } x_0 \in C, \tag{5.9}$$

which is the only (sensible) case which we shall consider, and that

$$\begin{cases} \eta \quad \text{close to } 1/2 \quad \Leftrightarrow x_0 \text{ is "close to center" of } C \\ \eta \quad \text{close to } 1 \quad \Leftrightarrow x_0 \text{ is "far from center" of } C \end{cases} \tag{5.10}$$

Notice also that the unit $\tilde{\beta}$ diam $(C, \mathcal{E})^2$ used in (5.7) to define the adimensional distance in data space is a measure of the maximum deviation across C of φ from its linear approximation.

We can now state the result.

Theorem 5.1 *(Wellposedness of regularized problem)*

Let $\mathcal{E}, C, \varphi, \varepsilon, x_0$ satisfy (2.1) and (2.4) only, $\eta, \zeta, \bar{\varepsilon}$ be defined by (5.5) (5.8) (5.6), and $\bar{\varepsilon}_{\min}, \overline{d}_{\max}$ by :

- *for $2/\pi \le \eta \le 1$ ("x_0 far from center of C"), and $\zeta \ge 0$:*

(5.11) $$\overline{\varepsilon}_{\min} = \eta$$

(5.12) $$\overline{d}_{\max} = \overline{\varepsilon}\{\overline{\varepsilon}^2 - \eta^2\}^{1/2}$$

- *for $1/2 \le \eta \le 2/\pi$ ("x_0 close to center of C"), and $\zeta \ge 0$:*

(5.13) $$\overline{\varepsilon}_{\min} = \begin{array}{l} \textit{largest root, in } [\eta, 2/\pi] \textit{ of the} \\ \overline{\varepsilon} \leadsto \overline{\varepsilon}\sin 1/\overline{\varepsilon} + ((1+\zeta^2/\overline{\varepsilon}^2)^{1/2} - 1)\cos 1/\overline{\varepsilon} - \eta \quad \textit{function.} \end{array}$$

(5.14) $$\overline{d}_{\max} = \begin{cases} \overline{\varepsilon}\{\overline{\varepsilon}^2 - \eta^2\}^{1/2} & \textit{for } 2/\pi \le \overline{\varepsilon} \\ \overline{\varepsilon}\{(\overline{\varepsilon}\sin 1/\overline{\varepsilon} + ((1+\zeta^2/\overline{\varepsilon}^2)^{1/2} - 1)\cos 1/\overline{\varepsilon})^2 - \eta^2\}^{1/2} \\ \textit{for } \overline{\varepsilon}_{\min} \le \overline{\varepsilon} \le 2/\pi \end{cases}$$

If :

(5.15) $$\overline{\varepsilon} > \overline{\varepsilon}\min$$

Then the regularized problem $(2.3)_\varepsilon$ is Q-wellposed on the cylindrical neighborhood of $\varphi(C)$ of adimensional size (defined in (5.7)) $\overline{d}_{\max}$ for the $\|x-y\|_{\mathcal{E}}$ distance on C.

Proof :

Rewriting problem $(2.3)_\varepsilon$ in terms of $E_\varepsilon, C_\varepsilon, F_\varepsilon, \varphi_\varepsilon$ and z_ε as indicated in (2.5), one checks easily that :

(5.16) $$\varepsilon\|y\|_{\mathcal{E}} \le \|\varphi'_\varepsilon(x)y\|_{F_\varepsilon} \le (\varepsilon^2 + \tilde{\alpha}_M^2)^{1/2}\|y\|_{\mathcal{E}} \quad \forall x \in C \cap \mathcal{E}, \forall y \in \mathcal{E}$$

(5.17) $$\|\varphi''_\varepsilon(x)(y,y)\|_{F_\varepsilon} \le \tilde{\beta}\|y\|_\varepsilon^2 \quad \forall x \in C \cap \mathcal{E}, \forall y \in \mathcal{E}$$

Following now corollary 4.2, we define :

(5.18) $$\Theta_\varepsilon = (\tilde{\beta}/\varepsilon)\text{diam } C = 1/\overline{\varepsilon}$$

(5.19) $$R_{G,\varepsilon} = \begin{cases} \overline{\varepsilon}^2\tilde{\beta}(\text{diam}C)^2 & \text{for } 0 \le 1/\overline{\varepsilon} \le \pi/2 \\ \overline{\varepsilon}\tilde{\beta}(\text{diam}C)^2\{\overline{\varepsilon}\sin 1/\overline{\varepsilon} + ((1+\zeta^2/\varepsilon^2)^{1/2} - 1)\cos 1/\overline{\varepsilon}\} \\ \text{for } \pi/2 \le 1/\overline{\varepsilon} < \pi \end{cases}$$

Corollary 4.2 implies now that, if

(5.20) $$\begin{cases} 2/\pi \le \overline{\varepsilon} \\ \text{or} \\ 1/\pi < \overline{\varepsilon} \le 2/\pi \quad \text{and } R_{G,\varepsilon} > 0 \end{cases}$$

then problem $(2.3)_\varepsilon$ is Q-wellposed on the neighborhood ϑ_ε of $\varphi_\varepsilon(C_\varepsilon)$ in F_ε of size $R_{G,\varepsilon}$. But, for any $z_\varepsilon = (z, \varepsilon x_0)$ one has :

$$d_\varepsilon(z_\varepsilon, \varphi_\varepsilon(C_\varepsilon))^2 = \inf_{x\in C\cap\mathcal{E}}\{\|\varphi(x) - z\|_F^2 + \varepsilon^2\|x - x_0\|_{\mathcal{E}}^2\}$$

i.e.

(5.21) $$d_\varepsilon(z_\varepsilon, \varphi_\varepsilon(C_\varepsilon))^2 \le d(z, \varphi(C))_F^2 + \varepsilon^2 rad(C, x_0, \mathcal{E})$$

Hence z_ε will belong surely to ϑ_ε independantly of the choice made for x_0 as soon as z belongs to the neighborhood ϑ of $\varphi(C)$ of size

$$(5.22) \qquad d_{\max} = \{R^2_{G,\varepsilon} - \varepsilon^2 rad(C, x_0, \mathcal{E})^2\}^{1/2},$$

provided of course that

$$(5.23) \qquad R_{G,\varepsilon} > \varepsilon \quad rad(C, x_0, \mathcal{E}).$$

Using the adimensional variables (5.5 thru 8), we rewrite (5.23) (which implies $R_{G,\varepsilon} > 0$!) as :

$$(5.24) \qquad \begin{cases} \overline{\varepsilon} > \eta & \text{if } 2/\pi \leq \overline{\varepsilon} \\ \overline{\varepsilon} \sin 1/\overline{\varepsilon} + ((1 + \zeta^2/\overline{\varepsilon}^2)^{1/2} - 1) \cos 1/\overline{\varepsilon} > \eta & \text{if } 1/\pi \leq \overline{\varepsilon} \leq 2/\pi \end{cases}$$

and (5.22) as :

$$(5.25) \qquad \begin{cases} \overline{d}_{\max} = \overline{\varepsilon}\{\overline{\varepsilon}^2 - \eta^2\}^{1/2} \quad \text{if } 2/\pi \leq \overline{\varepsilon} \\ \overline{d}_{\max} = \overline{\varepsilon}\{(\overline{\varepsilon} \sin 1/\overline{\varepsilon} + ((1 + \zeta^2/\overline{\varepsilon}^2)^{1/2} - 1) \cos 1/\overline{\varepsilon})^2 - \eta^2\}^{1/2} \\ \text{if } 1/\pi \leq \overline{\varepsilon} \leq 2/\pi \end{cases}$$

Hence we see that if $\overline{\varepsilon}$ satisfies (5.24), then the regularized problem $(2.3)_\varepsilon$ is Q-wellposed on the neighborhood ϑ of $\varphi(C)$ of adimensional size $\overline{d}_{\max}$ given by (5.25), which is the announced result. □

We have illustrated on figure 1 the $\eta \to \overline{\varepsilon}_{\min}$ function for various values of ζ, which makes clearly visible that the choice

$$(5.26) \qquad \overline{\varepsilon} > \max\{2/\pi, \eta\}$$

ensures the Q-wellposedness of the regularized problem independantly of the sensitivity index ζ. Figures 2 and 3 illustrate how the size $\overline{d}_{\max}$ of the neighborhood of $\varphi(C)$ depends on $\overline{\varepsilon}$ (taken larger than $\overline{\varepsilon}_{\min}$ of course !), η and ζ. These curves can be used for example to determine, given an estimation of an upper bound $\overline{d}_{\max}$ of the measurement and model error, the smallest amount of regularization $\overline{\varepsilon}$ to be used in order to restore wellposedness of the NLLS problem and suppress local minima on a neiborhood of $\varphi(C)$ large enough to contain the expected data.

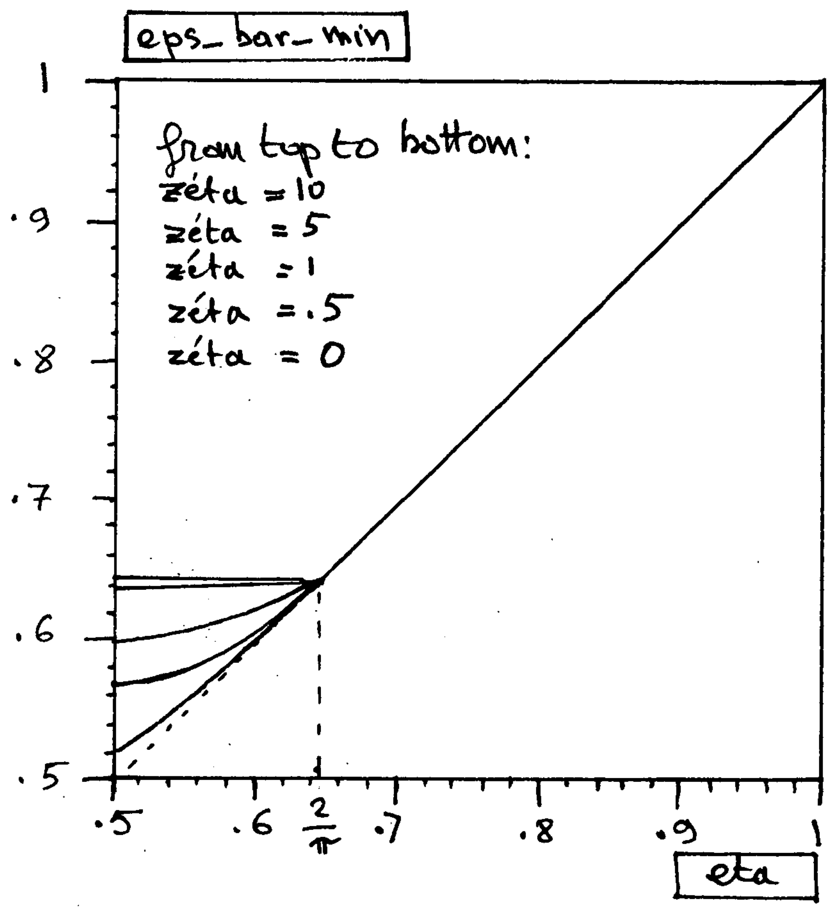
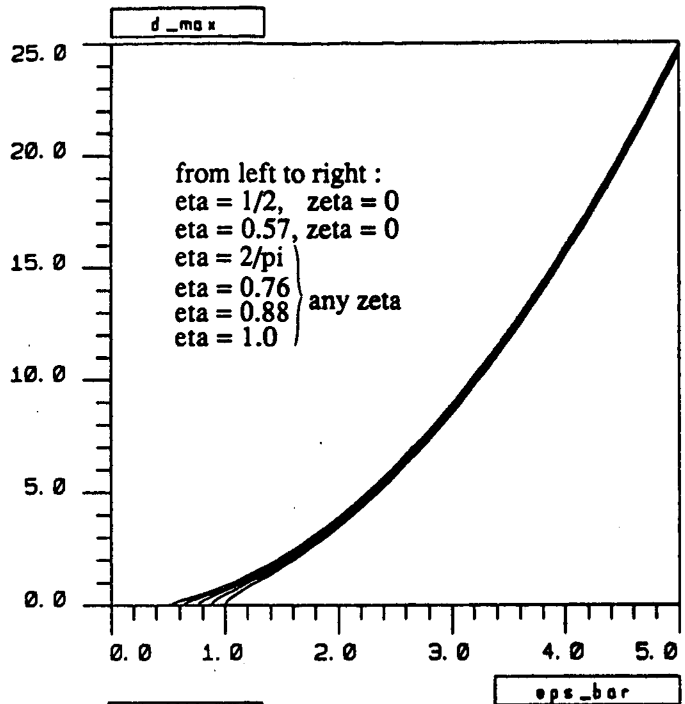


Figure 1: The minimum value $\bar{\varepsilon}$ min of the regularization parameter as function of the position index η of the a-priori guess and the sensitivity index ζ of the φ mapping



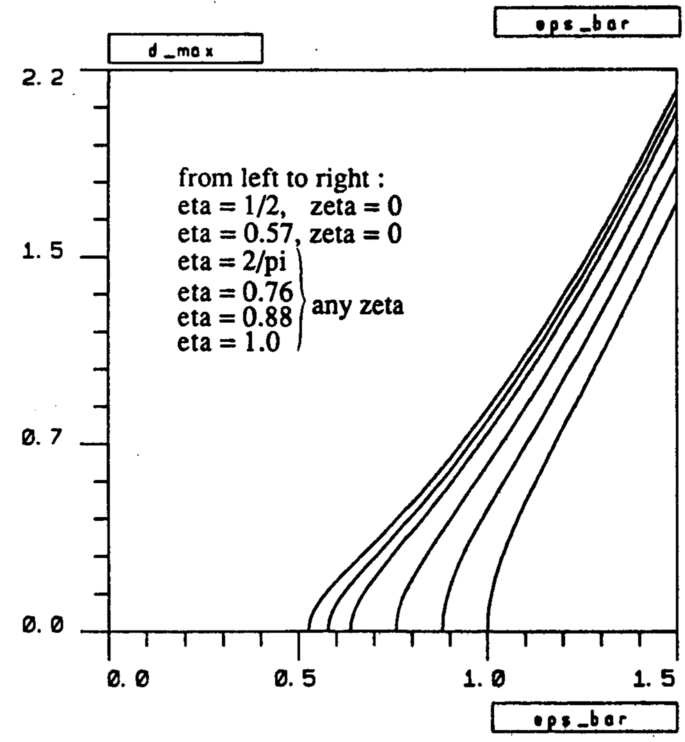


Figure 2: The $\overline{d}_{\max}$ function giving the size of the cylindrical neighborhood for the regularized problem, as function of $\overline{\varepsilon}$, for five values of the position index η of the initial guess x_0, and a zero value of the sensitivity parameters ζ for the two curves $\eta = .5$ and $\eta = .57$ corresponding to $\eta < 2/\pi$.
Top : general overview ; Bottom : close up on the $[0, 1.5]$ interval.

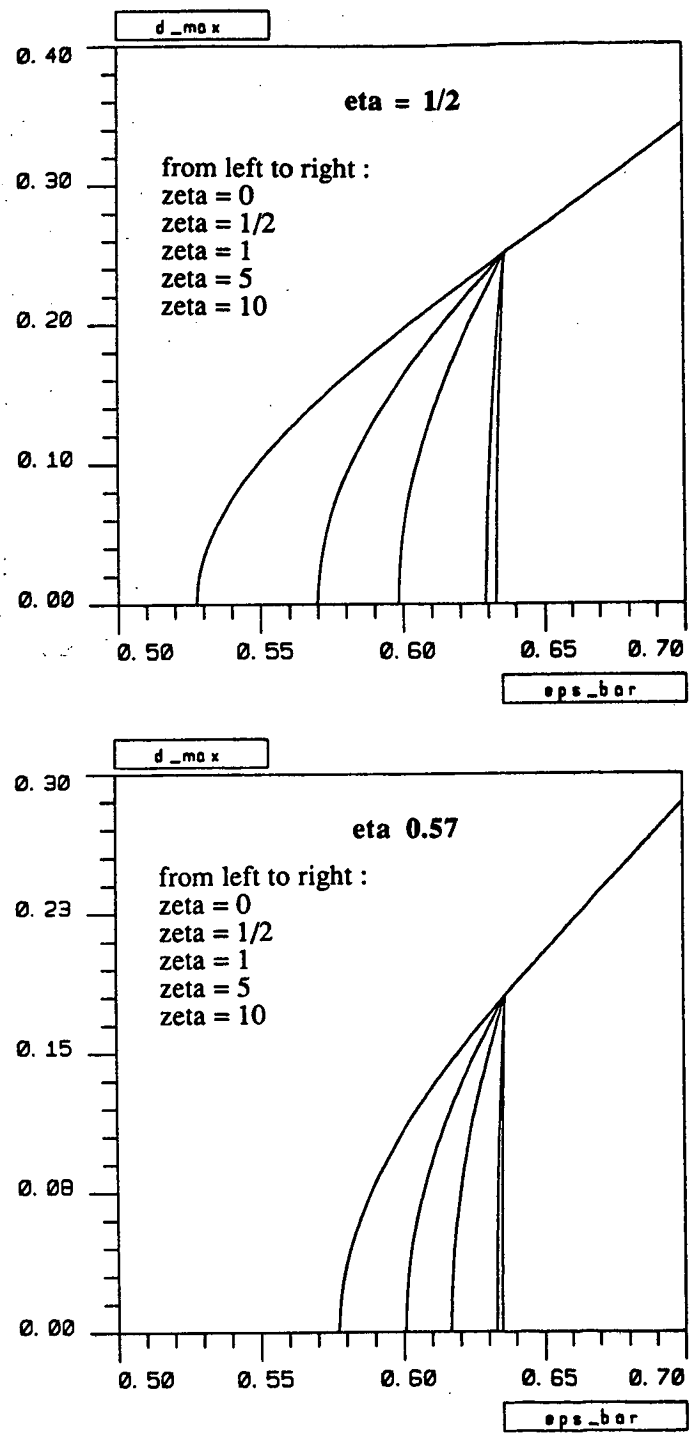


Figure 3: Influence of the sensitivity parameter ζ on $\overline{d}_{\max}$ for values of the position index η of the initial guess smaller than $2/\pi$.

References

[1] G. CHAVENT, "Local Stability of the Output Least-Square Parameter Estimation Technique", Mat. Aplic. Comp., Vol 2 n°1, pp 3-22, (1983)

[2] G. CHAVENT, "New Trends in Identification of Distributed Parameter Systems", Proceedings of the IFAC World Congress, München, Pergamon Pres, (1988)

[3] G. CHAVENT, "On the Uniqueness of Local Minima for General Abstract Non-Linear Least Square Problems", Inverse Problems 4, pp 417-433, (1988)

[4] G. CHAVENT, "Quasiconvex Sets and Size × curvature condition ; Application to non linear inversion", INRIA Report 1017, April 89 (Submitted to JAMO)

[5] G. CHAVENT, "New size × curvature conditions for strict quasiconvexity of sets", INRIA Report to appear, 1989

[6] W. W. SYMES, "The Plane Wave Detection Problem", to appear

[7] G. CHAVENT, J. LIU, "Multiscale parametrization for the estimation of a diffusion coefficient in elliptic and parabolic problems", Preprint IFAC Symposium on Control of Distributed Parameter Systems, Perpignan, France, June 26-29, 1989

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique